Journal of Multivariate Analysis 101 (2010) 409–418

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Cokriging for spatial functional data

David Nerini [a,*], Pascal Monestiez [b], Claude Manté [a]

[a] Laboratoire de Microbiologie, Géochimie et Ecologie Marines, UMR 6117 CNRS, Centre d'Océanologie de Marseille, Case 901, Campus de Luminy, 13288 MARSEILLE Cedex, France

[b] INRA,Unité Biostatistique et Processus Spatiaux, Domaine Saint Paul, Site Agroparc, 84914 AVIGNON, France

ARTICLE INFO

Article history:
Available online 25 March 2009

AMS 2000 subject classifications:
91B72
91B74
62P12
47B32

Keywords:
Functional data analysis
RKHS
Functional linear model
Coregionalization
Cokriging
Legendre polynomials

ABSTRACT

This work proposes to generalize the method of kriging when data are spatially sampled curves. A spatial functional linear model is constructed including spatial dependencies between curves. Under some regularity conditions of the curves, an ordinary kriging system is established in the infinite dimensional case. From a practical point-of-view, the decomposition of the curves into a functional basis boils down the problem of kriging in infinite dimension to a standard cokriging on basis coefficients. The methodological developments are illustrated with temperature profiles sampled with dives of elephant seals in the Antarctic Ocean. The projection of sampled profiles into a Legendre polynomial basis is performed with a regularization procedure based on spline smoothing which uses the variance of the sampling devices in order to estimate coefficients by quadrature.

1. Introduction

One of the hottest challenges for Functional Data Analysis [1] is the development of statistical methods adapted for spatially connected curves. Even if theoretical works are scarce in this area [2–4], many applications raise the need to include spatial relations between functional variables into statistical analysis especially in environmental science [5]. In oceanography, surveys provide vertical profiles of temperature, salinity or other variables that are spatially dependent and sampled along the depth. The recent emergence of autonomous underwater vehicles (AUV) carrying a variety of sensors allows a high-frequency sampling of the ocean [6]. If these AUVs are used to take vertical profiles of data, giving scientists a clearer understanding of the temperature, salinity, and turbidity of specific areas of the oceans, they also provide great amount of high-dimensional spatially correlated data. Most of the time, the analysis of such data involves geostatistical methods in order to map the area, for instance [7]. At best, the vertical dimension is included as a third spatial dimension and analysis is achieved with standard kriging. However, the average depth of the ocean is small compared to the horizontal dimensions. Yet, at the same scale, the vertical structures observed in the ocean (thermoclynes, salinity gradients) are much more important than those observed horizontally. The 3D kriging is often problematic due to strong and complex anisotropy and to non-stationarity along the vertical dimension. An alternative way is to discretize the curves and to modelize them using multivariate geostatistics [8]. The latter approach also suffers from several drawbacks: vertical data are often not available at the same depths, data analysis does not include the functional form of the variable along the profiles and computation is rapidly limited when profiles are recorded on a fine grid.

* Corresponding author.
E-mail address: david.nerini@univmed.fr (D. Nerini).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2009.03.005

Thus, the aim of this work is to propose a method which extends the coregionalization approach taking into account the functional nature of the data. We establish the problem of kriging in the infinite dimensional case under some regularity conditions of the functions. It is shown that, in the framework of a functional linear model with spatial dependency, the estimation of the regressor reduces to a multivariate cokriging problem for suitable choice of the functional space. The method is illustrated with data analysis of temperature profiles in the Antarctic Ocean, where marine mammals are used as samplers. Practical solutions are proposed in order to obtain a smoothed version of profiles whose shape includes the knowledge of the variance of the measuring devices.

2. Spatial linear model for functional data

Let us consider a collection of curves $E = \{y_i, i = 1, \ldots, n\}$ sampled at n random spatial locations $\mathbf{x}_i$ over a domain $\mathcal{D}$. Each $y_i(t)$ is an unique observation of $Y_i(t)$, a random function at location $\mathbf{x}_i$ where argument t varies in a compact interval τ of $\mathbb{R}$. The function Y_i takes values in a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ of functions on τ where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ denotes its inner product and $\|\cdot\|_{\mathcal{H}}$ the associated norm. Under second order stationarity assumptions, the mean function μ is the same at any point of the domain and it is supposed to be unknown

$$\mathbb{E}(Y_i) = \mu, \quad \forall \mathbf{x}_i \in \mathcal{D}.$$

Moreover, the linear covariance operator $C_{ij} : \mathcal{H} \to \mathcal{H}$ between Y_i at location $\mathbf{x}_i$ and Y_j at location $\mathbf{x}_j = \mathbf{x}_i + \mathbf{h}$ is defined as follows

$$C_{ij}(f) = \mathbb{E}[(Y_i - \mu) \otimes (Y_j - \mu)(f)] = \mathbb{E}(\langle f, Y_i - \mu \rangle_{\mathcal{H}} (Y_j - \mu)), \quad f \in \mathcal{H}.$$

This spatial covariance only depends on the vector $\mathbf{h}$ connecting the functional variable pair and is invariant for any translation of that pair into the domain. The operator C_{ij} is self-adjoint, positive, continuous and hence it is Hilbert–Schmidt (HS) and hence compact.

Following the ideas developed in Cuevas et al. [9] and Cardot et al. [10], we seek to estimate Y_0, the curve at unknown location $\mathbf{x}_0$, with the linear model

$$\widehat{Y}_0 = \sum_{i=1}^{n} B_i(Y_i), \tag{1}$$

with $B_i : \mathcal{H} \to \mathcal{H}$ being linear operators given by

$$B_i(Y_i)(t) = \langle \beta_i(\cdot, t), Y_i \rangle_{\mathcal{H}}, \quad t \in \tau. \tag{2}$$

where the functions β_i are unknown bivariate weighting functions. We suppose that the operators B_i belong to $\mathcal{B}(\mathcal{H})$ the space of HS linear operators from $\mathcal{H}$ to $\mathcal{H}$ equipped with the inner product

$$\forall (A, B) \in \mathcal{B}^2(\mathcal{H}), \quad \langle A, B \rangle_{\mathcal{B}} = \sum_{k,l} \langle A(u_k), u_l \rangle_{\mathcal{H}} \langle B(u_k), u_l \rangle_{\mathcal{H}}$$

where the functions u_k form an orthonormal basis of $\mathcal{H}$. In that case, the HS norm of B_i verifies $\|B_i\|^2_{\mathcal{B}} = \sum_{k,l} \langle B_i(u_k), u_l \rangle^2_{\mathcal{H}} < \infty, i = 1, \ldots, n$.

Now, looking for an unbiased estimator $\widehat{Y}_0$ of Y_0

$$\mathbb{E}\left(\widehat{Y}_0 - Y_0\right) = 0$$

leads to the following condition

$$\left[\sum_{i=1}^{n} B_i\right](\mu) = \mu. \tag{3}$$

As the mean function μ is unknown, the previous condition must be extended to the more general constraint

$$\left[\sum_{i=1}^{n} B_i\right](f) = f, \quad \forall f \in \mathcal{H}.$$

This constraint involves the existence of an operator K which must satisfy

$$K(f) = f, \quad \forall f \in \mathcal{H}$$
$$K \in \mathcal{B}(\mathcal{H}).$$

The first property is connected to the properties of the RKHS $\mathcal{H}$. Indeed, $\mathcal{H}$ possesses a reproducing kernel function κ defined as

$$\kappa : \tau \times \tau \to \mathbb{R}$$
$$(s, t) \mapsto \kappa(s, t)$$

which checks the following conditions

$$\begin{aligned} &\forall t \in \tau \quad \kappa(\cdot, t) \in \mathcal{H} \\ &\forall t \in \tau, \forall f \in \mathcal{H} \quad \langle f, \kappa(., t)\rangle_{\mathcal{H}} = f(t). \end{aligned} \tag{4}$$

Thus, the operator K, deduced from the reproducing kernel function κ is a linear operator by properties of the inner product on $\mathcal{H}$. It belongs to $\mathcal{B}(\mathcal{H})$ only if $\|K\|_{\mathcal{B}}^2 < \infty$.

Therefore, the search for the unbiased estimator $\widehat{Y}_0$ must be performed by minimizing

$$\mathbb{E}\left\|\widehat{Y}_0 - Y_0\right\|_{\mathcal{H}}^2 \tag{5}$$

under the constraint

$$\sum_{i=1}^{n} B_i = K.$$

Replacing condition (3) in (5) gives

$$\begin{aligned} \mathbb{E}\left\|\widehat{Y}_0 - Y_0\right\|_{\mathcal{H}}^2 &= \mathbb{E}\left\|\widehat{Y}_0 - \mu - Y_0 + \mu\right\|_{\mathcal{H}}^2 \\ &= \mathbb{E}\left\|\sum_{i=1}^{n} B_i(Y_i - \mu) - (Y_0 - \mu)\right\|_{\mathcal{H}}^2. \end{aligned}$$

Centering Y_i gives $Z_i = Y_i - \mu$ and we obtain

$$\mathbb{E}\left\|\sum_{i=1}^{n} B_i(Z_i) - Z_0\right\|_{\mathcal{H}}^2 = \sum_i \sum_j \mathbb{E}\left\langle B_i(Z_i), B_j(Z_j)\right\rangle_{\mathcal{H}} + \mathbb{E}\langle Z_0, Z_0\rangle_{\mathcal{H}} - 2\sum_i \mathbb{E}\langle B_i(Z_i), Z_0\rangle_{\mathcal{H}}. \tag{6}$$

Using the Kronecker product notation and reminding that

$$\forall B \in \mathcal{B}(\mathcal{H}), \forall (f, g) \in \mathcal{H}^2, \quad \langle B(f), g\rangle_{\mathcal{H}} = \langle B, f \otimes g\rangle_{\mathcal{B}}$$

the expression (6) becomes

$$\mathbb{E}\left\|\widehat{Y}_0 - Y_0\right\|_{\mathcal{H}}^2 = \sum_i \sum_j \left\langle B_i, B_j\, \mathbb{E}\left(Z_i \otimes Z_j\right)\right\rangle_{\mathcal{B}} + \langle K, \mathbb{E}(Z_0 \otimes Z_0)\rangle_{\mathcal{B}} - 2\sum_i \langle B_i, \mathbb{E}(Z_i \otimes Z_0)\rangle_{\mathcal{B}}.$$

Since $C_{ij} = \mathbb{E}\left(Z_i \otimes Z_j\right)$, we now seek for a solution $\left(\widetilde{B}_1, \ldots, \widetilde{B}_n\right)'$ of the optimization problem

$$\min_{(B_1,\ldots,B_n)' \in \mathcal{B}^n(\mathcal{H})} \sum_i \sum_j \left\langle B_j C_{ij}, B_i\right\rangle_{\mathcal{B}} + \langle K, C_{00}\rangle_{\mathcal{B}} - 2\sum_i \langle C_{i0}, B_i\rangle_{\mathcal{B}}$$

under the constraint

$$\sum_{i=1}^{n} B_i = K.$$

Following the Lagrange multipliers method, construct the function

$$F(B_1, \ldots, B_n, \Lambda) = \sum_i \sum_j \left\langle B_j C_{ij}, B_i\right\rangle_{\mathcal{B}} + \langle K, C_{00}\rangle_{\mathcal{B}} - 2\sum_i \langle C_{i0}, B_i\rangle_{\mathcal{B}} + 2 \times \left\langle \Lambda, \sum_i B_i - K\right\rangle_{\mathcal{B}}$$

where $\Lambda \in \mathcal{B}(\mathcal{H})$ is a Lagrange multiplier. The solution $\left(\widetilde{B}_1, \ldots, \widetilde{B}_n, \widetilde{\Lambda}\right)'$ is given by differentiating F and makes its $(n+1)$ terms equal to zero. In that case, the following equalities hold

$$\begin{cases} \left\|\sum_j B_j C_{ij} + \Lambda - C_{i0}\right\|_{\mathcal{B}} = 0, \quad i = 1, \ldots, n \\ \left\|\sum_i B_i - K\right\|_{\mathcal{B}} = 0. \end{cases}$$

We finally obtain the equivalent functional kriging system of $n+1$ equations:

$$\begin{cases} \sum_{j=1}^{n} B_j C_{ij} + \Lambda = C_{i0}, \quad i = 1, \ldots, n \\ \sum_{i=1}^{n} B_i = K. \end{cases} \tag{7}$$

This system can be solved if one can find an admissible estimator of the covariances C_{ij} for which the properties of a covariance operator hold.

3. Cokriging on basis coefficients

A comfortable way to estimate the weighting operators B_i is to consider a basis expansion approach. We suppose in this section that $\mathcal{H}$ is a finite p-dimensional Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ such that

$$\forall (f, g) \in \mathcal{H}^2, \quad \langle f, g \rangle_{\mathcal{H}} = \int_{\tau} f(t)\, g(t)\, \mathrm{d}t.$$

Each function Y_i can be expressed in terms of a linear combination of basis functions $\left(\phi_1, \ldots, \phi_p\right)$ of $\mathcal{H}$

$$Y_i(t) = \sum_{k=1}^{p} \alpha_k(\mathbf{x}_i)\, \phi_k(t) = \boldsymbol{\alpha}_i' \boldsymbol{\Phi}(t)$$

where $\boldsymbol{\alpha}_i = \left(\alpha_1(\mathbf{x}_i), \ldots, \alpha_p(\mathbf{x}_i)\right)'$ is the p- vector of coefficients at location $\mathbf{x}_i$ and $\boldsymbol{\Phi}(t) = \left(\phi_1(t), \ldots, \phi_p(t)\right)'$ the p-vector of basis functions. The weighting functions $\beta_i(s, t)$ can also be expanded into the ϕ-basis such that

$$\begin{aligned} \beta_i(s, t) &= \sum_{k=1}^{p} \sum_{l=1}^{p} b_{kl}^{i} \phi_k(s)\, \phi_l(t) \\ &= \boldsymbol{\Phi}'(s)\, \mathbf{B}_i \boldsymbol{\Phi}(t) \end{aligned}$$

for each (s, t) belonging to $\tau \times \tau$, where $\mathbf{B}_i$ is the $p \times p$ matrix of coefficients b_{kl}^{i} at location $\mathbf{x}_i$. Replacing these expressions in (1) gives the equation of the spatial linear model:

$$\widehat{Y}_0(t) = \sum_{i=1}^{n} \boldsymbol{\alpha}_i' \mathbf{W} \mathbf{B}_i \boldsymbol{\Phi}(t) = \widehat{\boldsymbol{\alpha}}_0' \boldsymbol{\Phi}(t). \tag{8}$$

where $\mathbf{W}$ is the matrix of the inner products

$$\mathbf{W} = \int_{\tau} \boldsymbol{\Phi}(t)\, \boldsymbol{\Phi}'(t)\, \mathrm{d}t.$$

It is always possible to impose

$$\mathbf{W} = \mathbf{I}$$

for suitable choice of orthonormal basis functions. Therefore, the reproducing kernel function $\kappa(s, t)$ is defined as

$$\kappa(s, t) = \sum_{k=1}^{p} \phi_k(s)\, \phi_k(t)$$

and verifies conditions (4) because any finite dimensional Hilbert space of functions has a reproducing kernel for any set $\left(\phi_1, \ldots, \phi_p\right)$ defining an orthonormal basis in $\mathcal{H}$ [11].

The stationarity hypotheses of the random functions Y_i expressed into an orthonormal ϕ-basis reduce to the classical multivariate stationarity assumptions over the coefficients of the expansion

$$\begin{cases} \mathbb{E}(\boldsymbol{\alpha}_i) = \mathbf{a}, \quad \forall \mathbf{x}_i \in \mathcal{D} \\ \mathbf{C}_{ij} = \mathbb{E}\left[(\boldsymbol{\alpha}_i - \mathbf{a}) \left(\boldsymbol{\alpha}_j - \mathbf{a}\right)' \right]. \end{cases}$$

where the mean $\mathbf{a}$ is a p-vector of coefficients, $\mathbf{C}_{ij}$ the $p \times p$ cross-covariance matrix with entries $cov\left(\alpha_k(\mathbf{x}_i), \alpha_l\left(\mathbf{x}_j\right)\right)$, $k, l = 1, \ldots, p$. The estimation of $\widehat{Y}_0$ in (8) is achieved through cokriging on coefficients

$$\widehat{\boldsymbol{\alpha}}_0 = \sum_{i=1}^{n} \mathbf{B}_i' \boldsymbol{\alpha}_i \tag{9}$$

where the $p \times p$ matrices $\mathbf{B}_i$, with entries $\left\{b_{kl}^{i}\right\}$, are weighting matrices which minimizes

$$\text{trace}\left(\text{var}\left(\widehat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0\right)\right). \tag{10}$$

The condition of unbiasedness given by

$$\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0\right) = 0$$

is satisfied by choosing weights that fulfill the constraints

$$\sum_{i=1}^{n} b_{kl}^{i} = \delta_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise} \end{cases}$$

which can be expressed in matrix form as $\sum_{i=1}^{n} \mathbf{B}_i = \mathbf{I}$. Note that the matrix form of the kernel operator K reduces to the identity matrix, in that case.

The constrained optimization of (10) provides solution $\left(\widetilde{\mathbf{B}}_1, \ldots, \widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Lambda}}\right)'$ solving the linear system (7) which now reads

$$\begin{pmatrix} \mathbf{C}_{11} & \cdots & \mathbf{C}_{n1} & \mathbf{I} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}_{1n} & \cdots & \mathbf{C}_{nn} & \mathbf{I} \\ \mathbf{I} & \cdots & \mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_n \\ \boldsymbol{\Lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{01} \\ \vdots \\ \mathbf{C}_{0n} \\ \mathbf{I} \end{pmatrix} \tag{11}$$

where $\boldsymbol{\Lambda}$ is a $p \times p$ matrix of Lagrange multipliers with entries λ_{kl}. This matrix equation is equivalent to a standard ordinary cokriging system on the basis coefficients. This system is written in the isotopic case *i. e.* when all the variables are available at all sampling points. It represents the discrete form of the functional kriging system (7). Once the solution $\left(\widetilde{\mathbf{B}}_1, \ldots, \widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Lambda}}\right)'$ has been computed, any function Y_0 is reconstructed with $\widehat{Y}_0(t) = \sum_{i=1}^{n} \boldsymbol{\alpha}_i' \mathbf{B}_i \boldsymbol{\Phi}(t)$.

4. Choice of the basis

We propose an approach that uses a regularization step provided by a spline smoothing in order to obtain the decomposition of curves into an orthonormal polynomial basis. Consider a function expressed in term of linear combination of known polynomial basis functions $\phi_1, \ldots, \phi_q$

$$y(t) = \sum_{k=1}^{q} \alpha_k \phi_k(t).$$

If the basis functions are orthonormal, the coefficients are computed in the following way

$$\alpha_k = \langle y, \phi_k \rangle_{\mathcal{H}} = \int_{\tau} y(t)\, \phi_k(t)\, \mathrm{d}t, \quad k = 1, \ldots, q. \tag{12}$$

The determination of each polynomial coefficient involves the integral (12) to be computed which can easily be achieved by quadrature rule. The simplest kind of quadrature rule is given by

$$\int_{\tau} y(t)\, \phi_k(t)\, \mathrm{d}t = \sum_{j=1}^{p} w\left(\widetilde{t}_j\right) y\left(\widetilde{t}_j\right) \phi_k\left(\widetilde{t}_j\right) + R_p(y).$$

The inner product in $\mathcal{H}$ is approximated by a finite sum involving p values of y and ϕ_k at suitably distinct nodes $\widetilde{t}_1, \ldots, \widetilde{t}_p$, called quadrature points for known weight function $w(t)$. The formula above is said to have degree of exactness d if the remainder $R_p(y) = 0$ for $y \in \mathbb{P}_d$, the set of polynomials of degree d. In the case above, the quadrature rule is interpolatory if $d = q - 1$ [12].

The situation is different with sampled curves since the function y can be evaluated only at location where sampling occurred. The quadrature approach cannot be performed in that case. Actually, each profile y arrives as a vector of m noisy sampled points $\mathbf{y} = (y(t_1), \ldots, y(t_m))'$ on a mesh varying from one curve to another one. A profile y can be considered such that

$$y(t) = g(t) + \varepsilon(t), t \in \tau$$

where g is a smooth function and $\varepsilon(t)$ is a random process such that $\mathbb{E}[\varepsilon(t)] = 0$ and $\mathbb{E}[\varepsilon(t)\, \varepsilon(s)] = \sigma^2$ for $s = t, = 0$, otherwise. It is desired to reconstruct the function g using the sampled vector $\mathbf{y}$. Using smoothing cubic splines, one wants to find $g_{m,\theta}$ to be the minimizer of the penalized sum of squares:

$$\frac{1}{m} \sum_{j=1}^{m} \left(f\left(t_j\right) - y\left(t_j\right)\right)^2 + \theta \int_{\tau} \left(f''(u)\right)^2 \mathrm{d}u$$

where the parameter θ controls the tradeoff between the smoothness of the solution as measured by $\int_{\tau} \left(f''(u)\right)^2 \mathrm{d}u$ and the empirical mean square error to the data computed by $\frac{1}{m} \sum_{j=1}^{m} \left(f\left(t_j\right) - y\left(t_j\right)\right)^2$.

The smoothing parameter θ is commonly estimated by cross-validation. However, a good value of θ can be found if the variance σ^2 is known, which may be regarded as the known variance of the measuring instruments. In that case, Craven and Wahba [13] proposed an unbiased estimator of the true mean square error

$$R(\theta) = \frac{1}{m} \sum_{j=1}^{m} \left(g_{m,\theta}\left(t_j\right) - g\left(t_j\right)\right)^2.$$

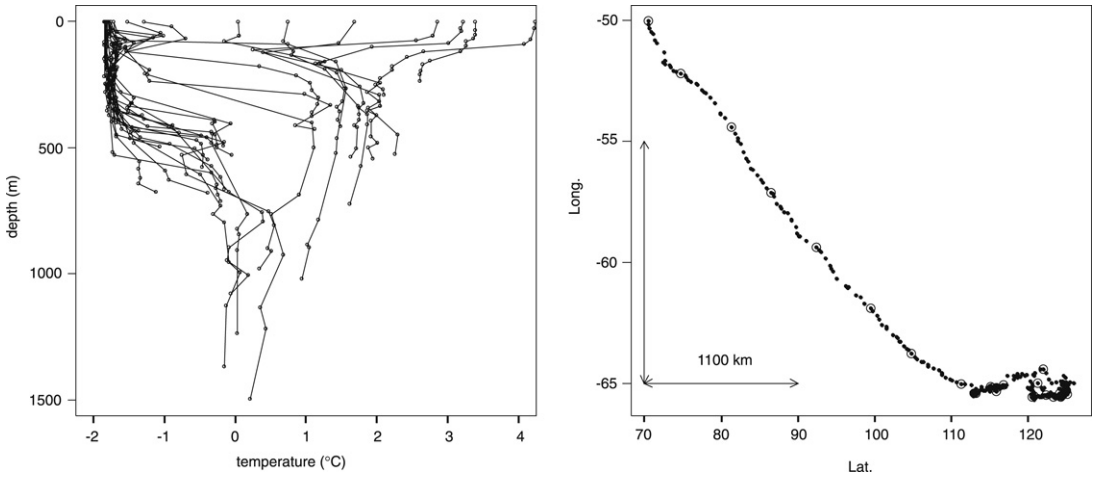

Fig. 1. Example of temperature profiles sampled by elephant seal and spatial location of sampled curves along trajectory (white circles). The cruise duration is 1 month. During the travel, the animal crosses and samples heterogeneous water structures.

As $g_{m,\theta}$ is a linear function of $\mathbf{y}$, define $\mathbf{A}(\theta)$ the $m \times m$ matrix satisfying

$$\mathbf{g}_{m,\theta} = \mathbf{A}(\theta)\,\mathbf{y}$$

where $\mathbf{g}_{m,\theta} = \left(g_{m,\theta}(t_1), \ldots, g_{m,\theta}(t_m)\right)'$. Then, an unbiased estimate of $\mathbb{E}[R(\theta)]$ is given by $\widehat{R}(\theta)$ defined by

$$\widehat{R}(\theta) = \frac{1}{m}\left\|(\mathbf{I} - \mathbf{A}(\theta))\,\mathbf{y}\right\|^2 - \frac{\sigma^2}{m}\mathrm{trace}\left[(\mathbf{I} - \mathbf{A}(\theta))^2\right] + \frac{\sigma^2}{m}\mathrm{trace}\left(\mathbf{A}^2(\theta)\right)$$

where $\|\cdot\|$ denotes the usual Euclidean norm. Therefore, a relevant choice of θ is the minimizer of $\widehat{R}(\theta)$ which can easily be computed [13].

We now dispose of an estimate $g_{m,\theta}$, a smoothed version of y whose shape no longer depends on the variability of the measuring devices. As this curve can be valued at any $t \in \tau$, the coefficients associated to the projection of the spline into the polynomial basis are computed as

$$\alpha_k = \sum_{j=1}^{p} w\left(\widetilde{t}_j\right) g_{m,\theta}\left(\widetilde{t}_j\right) \phi_k\left(\widetilde{t}_j\right) + R_p\left(g_{m,\theta}\right).$$

As the number p of coefficients increases, the remainder $R_p\left(g_{m,\theta}\right)$ tends to zero. The spline being a smooth curve, it is expected that the number of coefficients p necessary to match the entire variations of the curves is fairly small. Moreover, one can even use more quadrature points than data values ($p > m$) and still achieve a good polynomial fit. The choice of the number of coefficients p depends on the shape of the curves that form the sample at hand and will be discussed in the next section.

5. Dealing with real data

The above method is illustrated with data in oceanography where marine mammals operate as samplers. As a matter of fact, the southern Antarctic ocean is probably one of the least accessible areas on Earth. This place plays a key role in heat exchanges between the ocean and the atmosphere and very little oceanographic data is available. Since the 90's, scientists have been interested in the possibility to explore unknown parts of Antarctic ocean using elephant seals equipped with Argos transmitters including pressure, temperature and salinity sensors. The elephant seal becomes a valuable auxiliary for operational oceanography and enables scientists to study the hydrology of the southern ocean and the animal behavior [14,15].

Fig. 1 displays temperature profiles sampled in Antarctic ocean by an elephant seal traveling from the Kerguelen Islands to the Antarctic continental shelf. During one month, 438 temperature profiles are recorded each time the elephant seal dives at different discrete locations in space and depth. A fast examination of the left panel shows some peculiarities from this dataset. A profile y_i is recorded in the form of a discrete set of m_i points $\left\{y_i(t_1), \ldots, y_i\left(t_{m_i}\right)\right\}$. Comparing two profiles, the sampling points are located at different depths. Because of limited data storage, the sampling device is optimized to record a temperature value only in case of strong variations. Thus, a temperature curve is sampled well in areas where the local curvature is high, and few data points are available when the water column temperature is homogeneous. Moreover, each profile does not reach the same depth. The elephant seal does not dive each time to the same level and the sampling strategy is strongly dependent of its eating behavior. For this reasons, we limit the study of the temperature field between

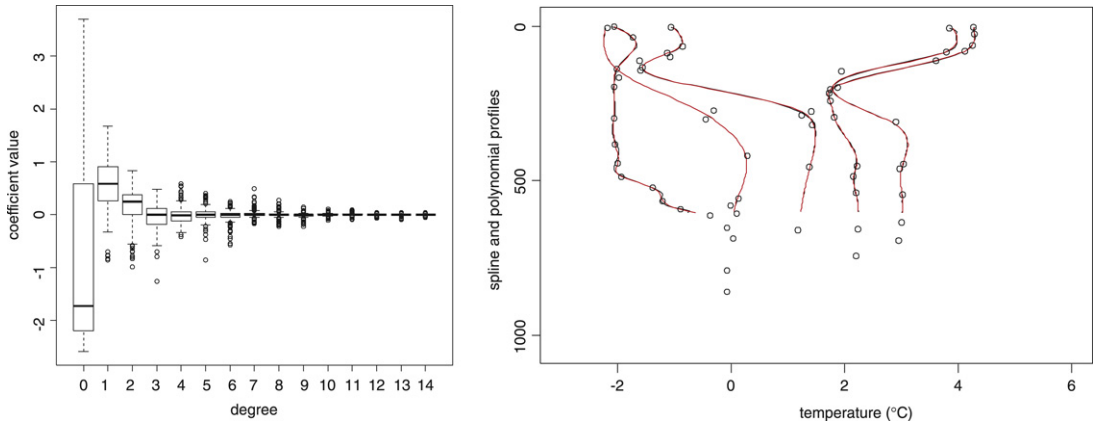

Fig. 2. Boxplot of the polynomial coefficients estimated with the 438 spline profiles. As the polynomial degree increases, the variability between coefficients decreases. The right panel shows a sample of smoothing splines and the associated polynomials. All curves are superimposed taking $p = 11$ polynomial coefficients. In the worst case, polynomials slightly oscillate at both ends of the spline.

0 m to the isobath of 600 m since this area contains the main part of the sampling points. In that case, each sampled profile is constituted of at least 10 temperature values. As is usual in oceanography, we also consider that the variations in water temperature structures are negligible during the traveling period.

5.1. Construction of the curves

The construction of the set E of functions reduces to the estimation of the polynomial coefficients of each function. The choice of the number of basis functions is fixed as follows:

(1) Evaluate cubic splines for each sampled profile, ranging on different depth, and get the collection of smooth curves $\left\{g_{m_1,\theta_1}, \ldots, g_{m_n,\theta_n}\right\}$ taking into account the known variability of the sampling devices. This variability value is the same for all profiles and it is provided by the manufacturer of the sampling device (here $\sigma^2 = 0.01$).

(2) Fix a tolerance error R_{tol} which corresponds to an arbitrary small value of the maximum accepted error between the spline curve and the fitted polynomial (here R_{tol}=1$e-08$).

(3) Compute by quadrature the polynomial degree and the coefficients for the whole set of sampled functions over a common support τ by successively increasing the number p of basis coefficients until $R_p\left(g_{m_i,\theta_i}\right) \leq R_{tol}, \forall i = 1, \ldots, n$.

(4) Store the n sets of p coefficients to form the sample $E = \{\boldsymbol{\alpha}_i, i = 1, \ldots, n\}$ of coefficient vectors.

We choose to express each function into an orthonormal Legendre polynomial basis but other choices may be relevant as well. This choice has been guided by the simplicity of the weighting function ($w(t) = 1, \forall t \in [-1, 1]$) and because the quadrature points and weights are easily obtained by a simple recurrence relation [12]. With $p = 11$ coefficients, polynomials are almost superimposed with splines (Fig. 2). Now, estimation of a functional profile may be achieved at any location $\mathbf{x}_0$ along the trajectory, using the kriging estimator in (9).

5.2. Covariance estimation

The linear system (11) possesses a solution if the left matrices of covariances and cross-covariances can be estimated and inverted. As we only dispose of a unique spatial sample of functions $\{y_1, \ldots, y_n\}$, the estimation of the covariance structures is realized through the fit of a linear coregionalization model for which the general statistical framework is the following. Each function y_i at position $\mathbf{x}_i$ can be seen as a p-vector $\boldsymbol{\alpha}_i$ of coefficients. We assume that these vectors $\boldsymbol{\alpha}_i$, $i = 1, \ldots, n$ are realizations of a multivariate random field $\boldsymbol{\alpha}(\mathbf{x}) = \left(\alpha_1(\mathbf{x}), \ldots, \alpha_p(\mathbf{x})\right)'$ of p components, where $\mathbf{x} \in \mathcal{D}$. The quantity $\boldsymbol{\alpha}(\mathbf{x})$ denotes the p values of the components of the random function Y at spatial position $\mathbf{x}$. The linear model of coregionalization supposes that the matrix $\mathbf{C}(\mathbf{h})$ of cross-covariances with entries $cov\left(\alpha_k(\mathbf{x}), \alpha_l(\mathbf{x}+\mathbf{h})\right)$, $k, l = 1, \ldots, p$ is proportional to a small number s of real correlation functions $\rho_u(\mathbf{h})$ such that

$$\mathbf{C}(\mathbf{h}) = \sum_{u=1}^{s} \mathbf{P}_u \rho_u(\mathbf{h})$$

where $\mathbf{P}_u$ are positive semi-definite coregionalization matrices. In practice, the estimation of the matrices $\mathbf{P}_u$ is carried out through weighted least squares fitting of variogram models to experimental data [7,8]. In this study, we used two covariance

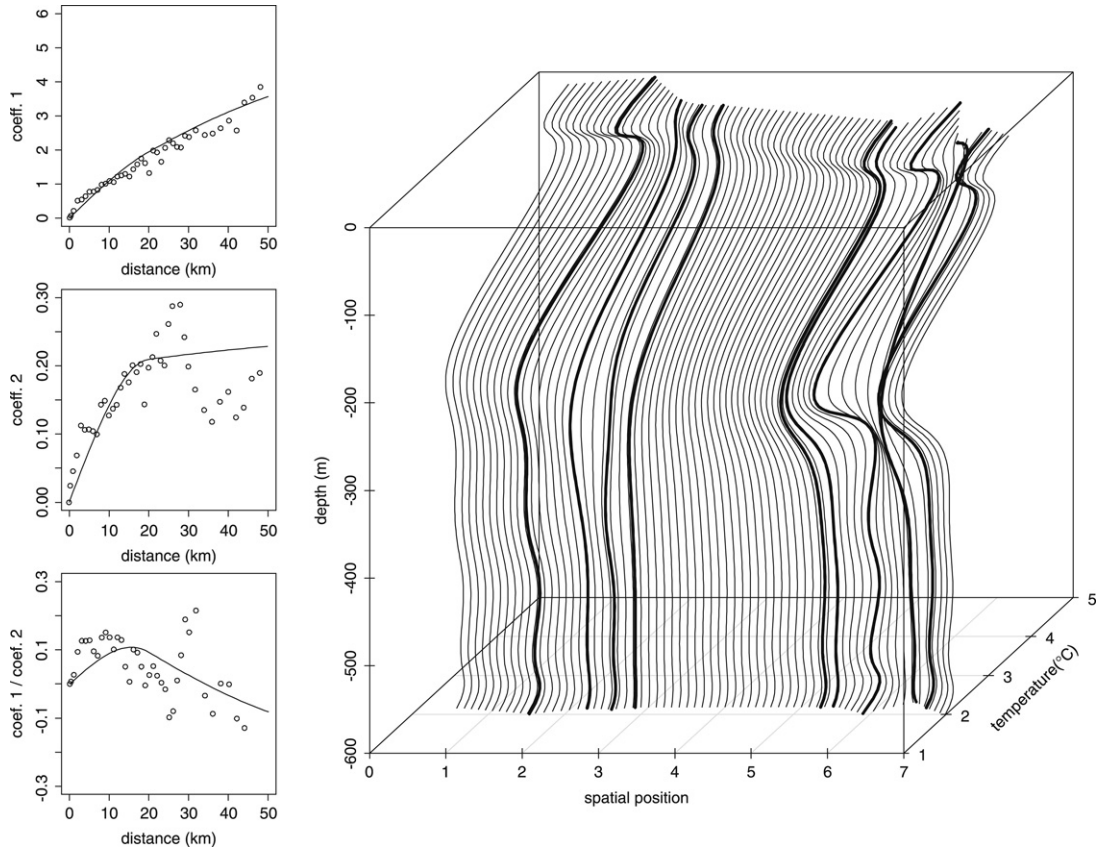

Fig. 3. Example of least squares fitting of a multivariate nested variogram model on the 2 first polynomial coefficients and associated estimated temperature profiles. The covariance structure is decomposed on two scales (20 km and 100 km) which provides an obtainable model for estimating the profiles at any locations of the sampled domain (spatial position in km). Bold curves are original polynomial profiles. Thin curves show the predicted curves with cokriging along a linear part of the trajectory of the elephant seal.

structures arising from the same variogram function $\gamma(\mathbf{h})$ of very common use (spherical model), whose equation is

$$\gamma(\mathbf{h}) = \begin{cases} \dfrac{3|\mathbf{h}|}{2r} - \dfrac{1}{2}\left(\dfrac{|\mathbf{h}|}{r}\right)^3 & \text{for } |\mathbf{h}| \le r \\ 1 \quad \text{for } |\mathbf{h}| \ge r \end{cases}$$

The range r of each model has been fixed by the practitioner (here 20 km and 100 km) so as to provide the most graphically satisfactory fit. The choice of the model and the number of basis structures have been chosen as discussed in [8]. For the sake of simplicity, we chose to display only curves obtained by kriging along a linear part of the animal trajectory (between *lat*. 55 S and *lat*. 60 S), but the entire dataset has been used to fit the covariograms. The coregionalization approach with two different scales provides an acceptable estimation of the field of temperature profiles ranging from 0 to 600 m (Fig. 3).

6. Discussion

The early ideas of kriging for functional data were first developed in Goulard and Voltz [2]. In the framework of interpolating curves, these authors define and study the properties of two linear unbiased predictors

$$\widehat{Y}_0(t) = \sum_{i=1}^{n} \beta_i(t)\, Y_i(t)\,,$$

where $\beta_i(t)$ is either a constant function at position $\mathbf{x}_i$ or a real function. The work of Delicado et al. [4] also provides a comparison between both these estimators in the context of the study of periodic functions. The authors propose a useful matrix framework for functional kriging.

In the present work, the problem of kriging curves has been considered in its broadest form considering the estimator

$$\widehat{Y}_0(t) = \sum_{i=1}^{n} \langle \beta_i(\cdot, t)\,, Y_i \rangle_{\mathcal{H}}$$

whose integral form will depend on the choice of the RKHS and its associated scalar product. In the finite dimensional case proposed here, the estimator reduces to

$$\widehat{Y}_0(t) = \sum_{i=1}^{n} \int_{\tau} \beta_i(s,t)\, Y_i(s)\, \mathrm{d}s.$$

The estimation of the weighting functions $\beta_i(\cdot,\cdot)$ revolves around three main steps: (i) the (not obvious) choice of the space $\mathcal{H}$ that spans the sampled functions, (ii) the choice of a polynomial basis estimated by quadrature and (iii) the covariance estimations by coregionalization.

Under stationarity assumptions, if the mean function μ is known (or well estimated), the spatial linear functional model can be extended to the space $\mathcal{L}^2(\tau)$ of square-integrable functions over τ: the problem of simple kriging is fixed. However, looking at the more general case where the mean function is unknown (ordinary kriging), some constraints must be imposed for the estimator $\widehat{Y}_0$ to be unbiased. These constraints have been integrated into the model by a reproducing kernel approach which implies stronger regularity conditions on the functions than those of the elements of $\mathcal{L}^2(\tau)$. A Hilbert space of functions owns a reproducing kernel only when the functions are sufficiently regular. The main idea developed here is that the reproducing kernels allow one to introduce symmetric functions $\kappa(\cdot,\cdot)$ which behave like the identity operator but which in addition belong to the same space when one argument has been fixed. In our case, the regularity assumption holds since, at this scale, the ocean may be considered as a continuous medium where profiles of an environmental variable are fairly smooth. In the $\mathcal{L}^2(\tau)$ case, the identity operator leaves any function unchanged but it is not an integral operator of $\mathcal{L}^2(\tau)$. In that case, the unbiasedness condition reads $\sum_i B_i = I$. However, this condition cannot be fulfilled since a sum of HS operators cannot be equal to the identity operator which is not HS.

Anyway, one could imagine the use of more refined RKHS spaces rather than a finite dimensional Hilbert space with a standard scalar product. For instance, the spline formalism can be embedded into the RKHS theory as shown in [16,17]. This introduces the second point: why use a polynomial basis rather than a smoothing spline basis straightforwardly? The choice of a finite dimensional Hilbert space with an orthonormal basis is justified by the fact that it allows one to find a simple expression of the minimization problem (11) and that the functional kriging problem comes down to a standard multivariate kriging on coefficients. Apart from this consideration, the use of a smoothing spline basis is banned here because it is impossible to find a common basis between profiles that have not been sampled at the same depths. B-spline basis have also proved their worth and could provide an interesting alternative. Their are currently the most usual way to deal with basis in the field of Functional Data Analysis [1]. However, the choice of the number of nodes and their position is not an easy task as discussed in [18,19]. An interesting way proposed by [20] with free-nodes splines should be explored further.

The coupled use of spline smoothing and polynomial estimation by quadrature is interesting because the number of polynomial coefficients is not constrained by the number of sampling points which should not be the case with direct polynomial regression. Indeed, a low number of sampled points per curve does not allow a reasonable polynomial fit if one needs a high polynomial degree in order to describe complex shape variations. The estimation of polynomial coefficients by quadrature provides a global parameterization of the curves while keeping the smoothness of their shape obtained by the good local properties of spline smoothing. Moreover, the use of orthonormal polynomials is quite simple and describes most of the shape variations of the curves with a fairly small number of coefficients which makes the fitting of covariograms easier. Finally, the properties provided by the orthonormality of the basis functions avoid ill-posed problems since both metric and reproducing kernel reduce to the identity.

Another procedure that might be investigated using polynomials and quadrature is the possibility to bypass the calculus of coefficients before kriging. Once the spline smoothing has been done, it is easy to choose an arbitrary number p of quadrature points $\left(\widetilde{t}_1, \ldots, \widetilde{t}_p\right)$ and to execute standard 3D-kriging straightforwardly on the set of spline vectors $\left(g_{m_i,\theta_i}\left(\widetilde{t}_1\right), \ldots, g_{m_i,\theta_i}\left(\widetilde{t}_p\right)\right)', i = 1, \ldots, n$. This procedure is equivalent to the one proposed by the basis approach if care is taken to weight each level $\left(\widetilde{t}_1, \ldots, \widetilde{t}_p\right)$ by the function $w(t)$ whose values depend on the selected quadrature scheme. A similar procedure has been proposed in [1] in the context of functional PCA.

The last point concerns the estimation of covariance structures by a coregionalization approach which is the most comfortable way to provide an obtainable model for spatial covariance. We first assume that all coefficients of the functional variable are generated by a same set of processes that act at different spatial scales. This implies that the shape of the covariance functions of the coregionalization matrix only depends of a small number of elementary structures weighted with specific coefficients, which is a fairly strong hypothesis. The choice of the basis structures is open for the practitioner and is included within the range of positive semi-definite models more often met in the literature. However, this choice remains a difficult task, under stationarity assumptions, as shown by fitted variograms on Fig. 3 and must be improved further.

Acknowledgments

The authors wish to thank Frederic Bailleul and Christophe Guinet (Centre d'Etude Biologique de Chize, CNRS, France) for providing access to the elephant seal dataset. Their work was funded by the CNES-TOSCA and the ANR project IPSOS-SEAL (French National Research Agency). The authors also thank Prof. J.C. Poggiale, Dr. M. Gauduchon and the members of the STAPH group for fruitful discussions and improvements brought to this work.

References

[1] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, New-York, 2005.
[2] M. Goulard, M. Voltz, Geostatistical interpolation of curves: A case study in soil science, in: A. Soares (Ed.), in: Geostatistics Troia'92, vol. 2, Kluwer Academic Publishers, 1992, pp. 805–816.
[3] S. Dabo-Niang, A.-F. Yao, Kernel regression estimation for continuous spatial processes, Math. Methods Statist. 16 (2007) 298–317.
[4] P. Delicado, R. Giraldo, J. Mateu, Point-wise Kriging for Spatial Prediction of Functional Data, in: S.Dabo-Niang, F. Ferraty (eds.), Proceedings of the First International Workshop on Functional and Operatorial Statistics, Springer Series in Contributions to Statistics, Berlin, 2008, pp. 135–142.
[5] W. Meiring, Oscillations and time trends in stratospheric ozone levels: A functional data analysis approach, J. Amer. Statist. Assoc. 102 (2007) 788–802.
[6] C.C. Eriksen, T.J. Osse, R.D. Light, T. Wen, T.W. Lehman, P.L. Sabin, J.W. Ballard, A.M. Chiodi, Seaglider: A long-range autonomous underwater vehicle for oceanographic research, IEEE Oceanic Eng. 26 (2001) 424–436.
[7] H. Wackernagel, Multivariate Geostatistics: An Introduction with Applications, Springer, New-York, 2003.
[8] M. Goulard, M. Voltz, Linear coregionalization model: Tools for estimation and choice of multivariate variograms, Math. Geol. 24 (1992) 269–286.
[9] A. Cuevas, M. Febrero, R. Fraiman, Linear functional regression: The case of fixed design and functional response, Canadian J. Statist. 30 (2002) 285–300.
[10] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, Statist. Probab. Lett. 45 (1999) 11–22.
[11] A. Berlinet, C. Thomas-Agnan, Reproducing Kernel Hilbert Spaces in Probability and Statistics, Kluwer Academic Publishers, 2004.
[12] W. Gautschi, Orthogonal Polynomials: Computation and Approximation, Oxford University Press, Oxford, 2004.
[13] P. Craven, G. Wahba, Smoothing Noisy Data with Splines Functions, Numer. Math. 31 (1979) 377–403.
[14] F. Bailleul, J-.B. Charrassin, R. Ezraty, F. Girard-Ardhuin, C.R. McMahon, I.C. Field, Ch. Guinet, Southern elephant seals from Kerguelen Islands confronted by Antarctic Sea ice. Changes in movements and in diving behaviour, Deep-Sea Res. II 54 (2007) 343–355.
[15] F. Bailleul, J-.B. Charrassin, P. Monestiez, F. Roquet, M. Biuw, Ch. Guinet, Successful foraging zones of southern elephant seals from the Kerguelen Islands in relation to oceanographic conditions, Phil. Trans. R. Soc. B 362 (2007) 2169–2181.
[16] Ph. Besse, J.O. Ramsay, Principal Components Analysis of Sampled Functions, Psychometrika 51 (1986) 285–311.
[17] N.D. Pearce, M.P. Wand, Penalized splines and reproducing kernel methods, J. Amer. Statist. Assoc. 60 (2006) 233–240.
[18] J.O. Ramsay, Monotone regression splines in action, Statist. Sci. 3 (1988) 425–461.
[19] P. Eilers, B. Marx, Flexible smoothing with B-splines and penalties, Statist. Sci. 11 (1996) 89–121.
[20] S. Zhou, X. Shen, Spatially adaptive regression splines and accurate knot selection schemes, J. Amer. Statist. Assoc. 96 (2001) 247–259.