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 $\|B_i\|_{\mathcal{B}}^2 = \sum_{k,l} \langle B_i(u_k), u_l \rangle_{\mathcal{H}}^2 < \infty, i = 1, \ldots, n$.

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

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