

Prediction of nonstationary functional data: Universal Kriging in a Hilbert Space

Alessandra Menafoglio, Matilde Dalla Rosa and Piercesare Secchi

1 Introduction

In an increasing number of studies collected data are curves. When functional data are spatially dependent, the inferential tools developed in the framework of Functional Data Analysis turn out to be somewhat inappropriate, since the dependence structure existing among observations is not taken into account by the model and the estimation procedures. Geostatistical techniques [2] for infinite-dimensional data has recently achieved particular attention in the literature, especially with respect to the problem of prediction for stationary random fields [5, 7, 3].

The aim of this work is to provide an extension of kriging methodology to non-stationary functional random fields. The problem is faced both from a theoretical point of view, establishing a coherent frame of definitions and hypotheses, and from a computational one, developing algorithms focused on spatial prediction of functional variables at unsampled locations. The proposed procedure is applied to daily mean temperatures curves observed in 35 meteorological stations located in Canada's Maritimes Provinces.

2 Model and Estimation Problem

Let us consider a functional random field $\{\chi_s, s \in D \subset \mathbb{R}^d\}$ such that, for each $s \in D$, χ_s is a random element of a separable Hilbert space H with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$, whose points are functions $x: \mathcal{T} \subset \mathbb{R} \rightarrow \mathbb{R}$. Let $\chi_{s_1}, \dots, \chi_{s_n}$ be a functional dataset, corresponding to the observation of the process in n locations $s_1, \dots, s_n \in D$: our aim is the prediction of the realization χ_{s_0} in an unsampled site $s_0 \in D$, through a Universal Kriging (UK) predictor, which is the best linear unbiased

Alessandra Menafoglio and Piercesare Secchi

MOX, Department of Mathematics, Politecnico di Milano, Italy, e-mail: alessandra.menafoglio@mail.polimi.it; piercesare.secchi@polimi.it

Matilde Dalla Rosa

Eni S.p.A., Exploration & Production Division, Italy, e-mail: matilde.dalla.rosa@eni.com

predictor (BLUP) $\chi_{s_0}^* = \sum_{i=1}^n \lambda_i^* \chi_{s_i}$, whose weights $\lambda_1^*, \dots, \lambda_n^* \in \mathbb{R}$ solve:

$$\min_{\lambda_1, \dots, \lambda_n \in \mathbb{R}} \mathbb{E}[\|\chi_{s_0}^* - \chi_{s_0}\|^2] \quad \text{subject to} \quad \mathbb{E}[\chi_{s_0}^* - \chi_{s_0}] = 0. \quad (1)$$

Assume for $\{\chi_s, s \in D \subset \mathbb{R}^d\}$ the following dichotomy model:

$$\chi_s = m_s + \delta_s, \quad (2)$$

where the drift m_s describes the spatial mean variation through a linear model, while the residual term δ_s is supposed to be a zero-mean, globally second-order stationary and isotropic random field [6] with covariance function $C(\cdot)$, i.e.:

$$\mathbb{E}[\chi_s(t)] = m_s(t) = \sum_{l=0}^L a_l(t) f_l(s), \quad f_0(s) = 1 \quad \forall s \in D, \quad t \in \mathcal{T}; \quad a_l \in H \quad \forall l = 0, \dots, L;$$

$$\text{Cov}(\delta_{s_i}, \delta_{s_j}) = \mathbb{E}[\langle \delta_{s_i}, \delta_{s_j} \rangle] = C(h), \quad \forall s_i, s_j \in D, \quad h = \text{dist}(s_i, s_j).$$

Assuming that $\{f_l\}_{l \geq 0}$ are known and $\{a_l\}_{l \geq 0}$ are independent from the spatial location, the expression of the UK predictor $\chi_{s_0}^*$ can be easily derived since problem (1) reduces to the following linear system (in block-matrix form):

$$\left(\begin{array}{c|c} \gamma(h_{i,j}) & f_l(s_i) \\ \hline f_l(s_j) & 0 \end{array} \right) \left(\begin{array}{c} \lambda_j \\ \mu_l \end{array} \right) = \left(\begin{array}{c} \gamma(h_{0,i}) \\ f_l(s_0) \end{array} \right), \quad (3)$$

where $\{\mu_l\}_{l \geq 0}$ are Lagrange multipliers, while $\gamma(\cdot)$ denotes the trace-semivariogram function of the residual process:

$$\gamma(h_{i,j}) = \text{Var}[\delta_{s_i} - \delta_{s_j}] = \mathbb{E}[\|\delta_{s_i} - \delta_{s_j}\|^2], \quad s_i, s_j \in D, \quad h_{i,j} = \text{dist}(s_i, s_j),$$

that plays the same key role in functional geostatistics as in the finite-dimensional case.

An estimation of $\gamma(\cdot)$ is thus necessary in order to determine the UK predictor $\chi_{s_0}^*$. As in classical geostatistics, this estimation can be performed in two steps: first the computation of an empirical estimate from (an estimate of) the vector of residuals $\delta_s = (\delta_{s_i})$, then the fitting of a variogram valid model.

Although the drift coefficients are not directly included in system (3), their estimation is needed since the residual process is in general unobserved. In order to take into account the spatial dependence among observations, we propose to estimate the drift through the generalized least squares (GLS) estimator $\hat{m}_s^{GLS}$ minimizing the functional Mahalanobis distance between the observations vector $\chi_s = (\chi_{s_i})$ and drift estimates one $\hat{m}_s = (\hat{m}_{s_i})$:

$$\hat{m}_s^{GLS} = \underset{\{\hat{m}_s = (\hat{m}_{s_i}); \hat{m}_{s_i} \in H\}}{\text{argmin}} \sum_{i=1}^n \|\Sigma^{-1/2}(\chi_s - \hat{m}_s)\|_i^2; \quad \Sigma = \text{Cov}(\chi_s) \in \mathbb{R}^{n \times n}. \quad (4)$$

In [6] the explicit formula for the unique solution $\hat{m}_s^{GLS}$ of the estimation problem (4) is provided ($\hat{m}_s^{GLS} = \mathbb{F}_s(\mathbb{F}_s^T \Sigma^{-1} \mathbb{F}_s)^{-1} \mathbb{F}_s \Sigma^{-1} \chi_s$, $(\mathbb{F}_s)_{il} = f_l(s_i)$), deriving furthermore its properties and showing that it coincides with the best linear unbiased estimator (BLUE) for the mean vector m_s .

3 Assessing the Drift for Universal Kriging Prediction

Consistently with the established theoretical results, a three-steps procedure for the prediction of non-stationary spatial dependent functional data is proposed: drift model selection, GLS drift estimation, Universal Kriging prediction.

The first step is needed whenever the family $\{f_l\}_{l \geq 0}$ is ‘a priori’ unknown: in such a case, we propose to adopt a predictive criterion, selecting, among polynomial models of order lower than two, the one minimizing the prediction sum of squared error $SSE = \mathbb{E}[\|\chi_{s_0} - \chi_{s_0}^*\|^2]$, $s_0 \in D$, assessed through a cross-validation method.

Having selected the family $\{f_l\}_{l \geq 0}$, in order to compute the GLS drift estimation an iterative algorithm is necessary, since both the drift estimator $\hat{m}_s^{GLS}$ and the Universal Kriging system (3) depend substantially on the residual covariance structure, that can be assessed only once an estimation of the residual process —obtained by difference from the estimate $\hat{m}_s^{GLS}$ — is available. Therefore we propose to initialize the procedure to the ordinary least squares estimate $\hat{m}_s^{GLS(0)} = \hat{m}_s^{GLS}|_{\Sigma=\mathbb{I}}$, computing at each generic step $k \geq 1$ the residual estimate $\hat{\delta}_s^{GLS(k)} = \chi_s - \hat{m}_s^{GLS(k-1)}$, the related trace-variogram structure $\hat{\gamma}^{(k)}(\cdot)$ and updating the drift estimate on the basis of the structure of spatial dependence currently available: $\hat{m}_s^{GLS(k)} = \hat{m}_s^{GLS}|_{\Sigma=\hat{\Sigma}^{(k)}}$. Once the convergence has been reached, the final estimate of the variogram model can be used to solve the Universal Kriging system (3), deriving the desired prediction.

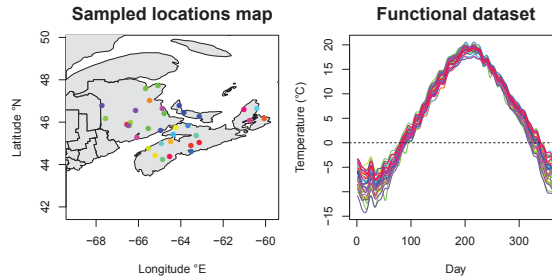
Algorithms validation has been carried out through their application to simulated data, verifying convergence properties and deriving robustness with respect to the functional form of the valid variogram model. The predicting power of the proposed methodology has been finally tested with satisfactory results in different scenarios with different data sample size.

4 Case Study

The proposed methodology is applied to the Canada’s Maritime Provinces Temperatures dataset, that collects daily mean temperatures data, observed in 35 meteorological stations located in Canada’s Maritimes Provinces [1] (Fig. 1).

As a matter of innovation, geostatistical analysis is performed modeling the non-stationary mean behavior of the process through a drift term with a non-Euclidean metric on the spatial domain. Since no information is ‘a priori’ available on the

Fig. 1 Canada’s Maritime Provinces Temperatures functional dataset. On the left: geographical locations. On the right: temperature curves. The color of each sampled point on the map corresponds with the color of the related functional data on the right.



deterministic variability of the process, the drift model selection is performed among polynomials of order lower than 2, singling out as optimal drift model a quadratic form ($\{f_l\}_{l=0,\dots,3} = \{1, x^2, y^2, xy\}$). Five iterations of the drift estimation algorithm are sufficient for its convergence, leading finally to GLS drift estimation maps (Fig. 2, top panels) and Universal Kriging maps (Fig. 2, bottom panels).

Besides being consistent with respect to the seasonal reference maps [8], the obtained results are climatically interpretable. The exposition of the Maritimes region towards the sea plays a key role indeed, due to influence on temperatures of the alternation of Atlantic warm-humid currents with freezing streams coming from the internal Canadian regions.

The introduction of the drift term in the model improves the predictive power of the method with respect to previous analyses of the same data presented in the literature (e.g., [4]). Moreover, the regularizing effect typical of kriging is partially mitigated by the presence of the drift term, that turns the prediction from a strongly data driven behavior to a model driven one. The resulting prediction is thus precise also in peripheral regions, catching the local structures as well (Fig. 2, July panel).

References

1. Canada's Maritime Provinces Temperatures dataset: <http://www.docentes.unal.edu.co/rgiral-doh/docs/> (30th January 2012).
2. Cressie N.: Statistics for Spatial data. John Wiley & Sons, New York (1993).
3. Delicado P., Giraldo R., Comas C., Mateu J.: Statistics for spatial functional data. *Environmetrics* **21**, 224–239 (2010).
4. Giraldo R., Delicado P., Mateu J.: Geostatistics for functional data: An ordinary kriging approach. Technical report. Universitat Politècnica de Catalunya, <http://hdl.handle.net/2117/1099> (2008).
5. Goulard M., Voltz M.: Geostatistical Interpolation of Curves: A Case Study in Soil Science. In: *Geostatistics Tróia '92*, ed. A. Soares, Dordrecht: Kluwer Academic, pp. 805–816 (1993).
6. Menafoglio A.: Geostatistica per elementi di uno spazio di Hilbert: teoria e applicazioni ai dati funzionali. Master thesis. Politecnico di Milano, Milano (2011).
7. Monestiez P., Nerini D.: A Cokriging Method for Spatial Functional Data with Applications in Oceanology. In: S. Dabo-Niang and F. Ferraty, *Functional and operatorial statistics*, pp. 237–242. Springer (2008).
8. The Atlas of Canada: <http://atlas.nrcan.gc.ca/> (30th January 2012).

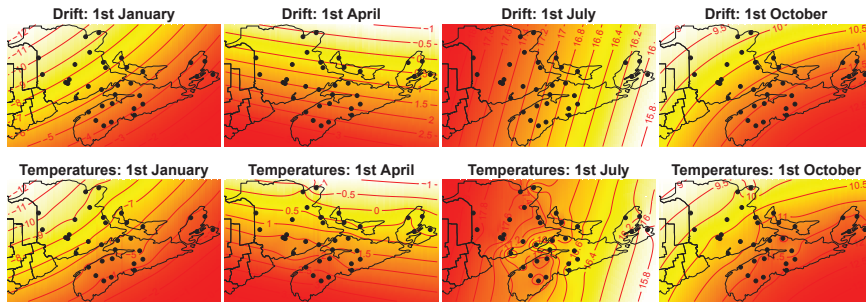


Fig. 2 GLS drift estimation maps (top panels) and UK prediction maps (bottom panels).