PROCEEDINGS OF SIMAI 2016

THE XIII BIANNUAL CONGRESS OF SIMAI

13-16 SEPTEMBER 2016

MILAN, ITALY

Editors: L. Bonaventura, L. Formaggia, E. Miglio, N. Parolini, A. Scotti and C. Vergara

SIMAI
SOCIETÀ ITALIANA DI MATEMATICA APPLICATA E INDUSTRIALE

Conference logo design by Anna Scotti. Typographic composition made in LaTeX by Mattia Penati.

SIMAI
SOCIETÀ ITALIANA DI MATEMATICA APPLICATA E INDUSTRIALE
http://www.simai.eu

ISBN 978-88-6493-035-0

Preface

Since 1992, the Italian Society for Applied and Industrial Mathematics (SIMAI) holds a biennial congress gathering contributions of researchers from academia and industry working on industrial and applied mathematics problems.
This book collects the abstracts of the talks and plenary lectures given at the SIMAI Congress 2016 that took place in Milano, Italy, from September 13 to September 16, 2016.
We are extremely satisfied that so many people have shown their interest in this meeting. In addition to 6 invited plenary lectures, we had more than 360 contributions from Italy and many other European countries, organized into 64 minisymposia. The contributions contained in this book cover both theoretical aspects and practical applications of mathematics and scientific computing. Topics include the analysis of evolution and dissipative processes, stochastic modeling, numerical methodologies such as computational optimization, advanced numerical methods for PDEs, conservation laws and inverse problems, optimal control, model reduction and high-performance-computing, as well as statistical methodologies for the treatment of complex data and signals.
A wide range of applications is covered from life science and biology to geophysics, from image processing to petroleum engineering and quantitative finance.
We would like to thank all participants for their valuable contributions. In particular we mention the fundamental contribution of the minisymposia organizers.
Special thanks are due to the invited speakers: P. Antonietti (Politecnico di Milano), A. Buffa (IMATI-CNR), A. Pontremoli (Dallara Automobili), Wil Schilders (TU Eindhoven), A. Quarteroni (EPFL), and G. Toscani (University of Pavia), for contributing to the success of the conference with the high quality of their contributions.
We gratefully acknowledge the support of the industrial sponsors: Mathesia, MOXOFF, Noesis and SpingerNature.
Finally we would like to thank the Politecnico di Milano for hosting the Congress and, in particular, the Eventimate Team (Anna Rho and Laura Guarino) for the logistic support in the organization of the conference and Luca Lo Curto for the technical support. Moreover we thank and all volunteers (mainly post-doc and PhD students) for their help during the meeting.
We believe that the wide range of applications and the scientific quality of the contributions collected in this book represent the best evidence of the important role that the industrial and applied mathematics can play in our society.
We believe that this book gives an up-to-date description of the state of the art of the research in industrial and applied mathematics in Italy.

Milano, Italy
September 2016

The Organizing Committee

Back to Table of Contents

Functional Kriging Uncertainty Assessment

Rosaria Ignaccolo
Università degli Studi di Torino
Lungo Dora Siena 100A, 10153 - Torino Italy
rosaria.ignaccolo@unito.it

Maria Franco-Villoria
Università degli Studi di Torino
Lungo Dora Siena 100A, 10153 - Torino Italy
maria.francovilloria@unito.it

Geostatistical techniques for functional data were introduced by Goulard and Voltz (1993) [4], but have only been developed recently. Several papers consider ordinary and universal kriging models to predict a curve at an unmonitored site under the assumption of a constant or longitude and latitude dependent mean (see e.g. [2, 3, 8, 1, 7]) or kriging with external drift [5], where scalar and functional exogenous variables are introduced. However, uncertainty evaluation of a predicted curve remains an open issue. Given the difficulty to derive sampling distributions for functional data, prediction band derivation can be approached using resampling methods. To evaluate uncertainty of a predicted curve, we adapt two semi-parametric bootstrap approach for spatially correlated data proposed by [10] and [6] to the functional data case. The approach is illustrated by means of a simulation study.

1 Functional Kriging with External Drift (FKED)

Let $\Upsilon_s = \{Y_s(t); t \in T\}$ be a functional random variable observed at location $s \in D \subseteq \mathbb{R}^d$, whose realization is a function of $t \in T$, T compact subset of $\mathbb{R}$. Assume that we observe a sample of curves Υ_{s_i}, for $s_i \in D$, $i = 1, \ldots, n$, that take values in a separable Hilbert space of square integrable functions. The set $\{\Upsilon_s, s \in D\}$ constitutes a functional random field or a *spatial functional process* [2], that can be non-stationary and whose elements are supposed to follow the model $\Upsilon_s = \mu_s + \epsilon_s$. The term μ_s is interpreted as a drift describing a spatial trend while ϵ_s represents a residual random field that is zero-mean, second-order stationary and isotropic. At the generic site s_i, $i = 1, \ldots, n$, and at point t, the model can be rewritten as a functional concurrent linear model $Y_{s_i}(t) = \mu_{s_i}(t) + \epsilon_{s_i}(t)$ with the drift

$$\mu_{s_i}(t) = \alpha(t) + \sum_p \gamma_p(t) C_{p,i} + \sum_q \beta_q(t) X_{q,i}(t) \tag{1}$$

where $\alpha(t)$ is a functional intercept, $C_{p,i}$ and $X_{q,i}$ are the p^{th} and q^{th} scalar and functional covariates at site s_i with coefficients $\gamma_p(t)$ and $\beta_q(t)$ and $\epsilon_{s_i}(t)$ represents the residual spatial functional process $\{\epsilon_s(t), t \in T, s \in D\}$ at site s_i. Once the Functional Regression Model (1) has been fitted by means of a GAM representation (for details see [5]), the functional residuals $e_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t)$ can be used to predict the residual curve at a new site s_0 via ordinary kriging for functional data [3], according to which $\hat{e}_{s_0}(t) =$

Work supported by "Futuro in Ricerca" 2012 Grant (project no. RBFR12URQJ) provided MIUR.

$\sum_{i=1}^{n} \lambda_i e_{s_i}(t)$, with $\lambda_i \in \mathbb{R}$. Prediction at the new site s_0 is obtained by adding up, as in the classical regression kriging, the two terms, i.e. $\hat{Y}_{s_0}(t) = \hat{\mu}_{s_0}(t) + \hat{e}_{s_0}(t)$, where $\hat{\mu}_{s_0}(t) = \hat{\alpha}(t) + \sum_p \hat{\gamma}_p(t) C_{p,0} + \sum_q \hat{\beta}_q(t) X_{q,0}(t)$ depends on the covariate values $C_{p,0}$ and $X_{q,0}(\cdot)$ at the site s_0.

2 Uncertainty evaluation

To evaluate the uncertainty of a predicted curve $\hat{Y}_{s_0}(t)$ at a new site s_0, we consider two semi-parametric bootstrap approaches for spatially correlated data proposed by [10] and [6] and extend them to the functional context. Suppose that $\hat{Y}_{s_0}(t) - Y_{s_0}(t)$ follows the distribution F_n, a $1-\alpha$ prediction interval for $Y_{s_0}(t)$ can be built as $(\hat{Y}_{s_0}(t) - q_{1-\alpha/2}, \hat{Y}_{s_0}(t) - q_{\alpha/2})$, with q_α the α^{th}-quantile of the unknown distribution F_n. The idea is to construct B bootstrap replicates $\{\hat{Y}^{*j}_{s_0}, Y^{*j}_{s_0}\}_{j=1}^{B}$ and approximate F_n by $\hat{F}^*_n$, the empirical distribution of $\{\hat{Y}^{*j}_{s_0} - Y^{*j}_{s_0}\}_{j=1}^{B}$. The bootstrapping algorithm can be summarized as follows:

1. Estimate and remove the drift following Model (1) to obtain $\epsilon_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t)$.

2. Estimate the functional residuals covariance matrix Σ through the estimated trace-semivariogram:

$$\hat{v}(h) = \frac{1}{2\,|N(h)|} \sum_{i,j \in N(h)} \int_T \left(e_{s_i}(t) - e_{s_j}(t)\right)^2 dt$$

   where $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$. A parametric model (e.g. Matérn) can be fitted to the points $(h_g, \hat{v}(h_g)), g = 1, \ldots, G$, as in classical geostatistics. Using Cholesky decomposition, $\hat{\Sigma} = \hat{L}\hat{L}^T$ and the functional residuals can be transformed:

$$\zeta_{n\times M} = (\zeta(s_1), \ldots, \zeta(s_n))' = \hat{L}^{-1}_{n\times n} \left(Y_{n\times M} - \hat{\mu}_{n\times M}\right).$$

3. Generate B bootstrap samples with size $n+1$, $\zeta^*_{n+1} = (\zeta^*(s_1), \ldots, \zeta^*(s_n), \zeta^*(s_{n+1}))'$ from $\zeta(s_1), \ldots, \zeta(s_n)$.

4. Create the augmented covariance matrix $\hat{\Lambda} = \begin{bmatrix} \hat{\Sigma} & \hat{c}_n^T \\ \hat{c}_n & \hat{\sigma}^2 \end{bmatrix}$, where $\hat{c}_n = \{\hat{C}(s_i - s_0)\}_{i=1}^{n}$, $\hat{C}$ is the estimated covariance function and $\hat{\sigma}^2 = \hat{C}(0)$ is the estimated sill. Use Cholesky decomposition so that $\hat{\Lambda} = \hat{R}\hat{R}^T$ and transform the bootstrap samples ζ^*_{n+1} as

$$(e^*(s_1), \ldots, e^*(s_n), e^*(s_0))' = \hat{R}_{(n+1)\times(n+1)} \zeta^*_{(n+1)\times M}.$$

5. The final bootstrap sample is determined as $Y^*_{s_i}(t) = \hat{\mu}_{s_i}(t) + e^*_{s_i}(t)$, $i = 1, \ldots, n$ and $Y^*_{s_0}(t) = \hat{\mu}_{s_0}(t) + e^*_{s_0}(t)$.

The bootstrap samples $\{Y^{*j}_{s_1}, \ldots, Y^{*j}_{s_n}\}_{j=1}^{B}$ are then fed into the FKED method to obtain B prediction curves $\hat{Y}^{*j}_{s_0}$ and the differences $\{\hat{Y}^{*j}_{s_0} - Y^{*j}_{s_0}\}_{j=1}^{B}$ are considered. The prediction interval for $Y_{s_0}(t)$ can be written as $(\hat{Y}_{s_0}(t) - q^*_{1-\alpha/2}, \hat{Y}_{s_0}(t) - q^*_{\alpha/2})$, with q^*_α the α-percentile of $\hat{F}^*_n$, that can be obtained ordering the curves. There is no gold standard for ordering functional data. Here we consider two ordering techniques based on band depth and L^2 distance. Band depth [9] can be defined for any set of k curves (here $k = 2$). The sample band depth (BD) of $y(t)$ can be calculated as the proportion of bands delimited by two

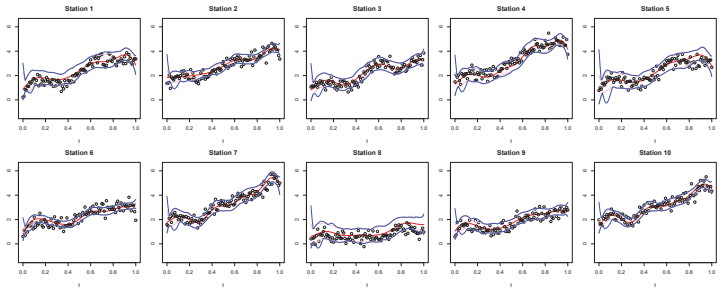

Figure 1: Original data (black dots), FKED predicted curve (solid red line), 95% prediction band (pink) based on L^2 distance (pink) and on *MBD* (blue) for $n = 50$, $\sigma^2 = 0.25$, $\phi = 1.5$.

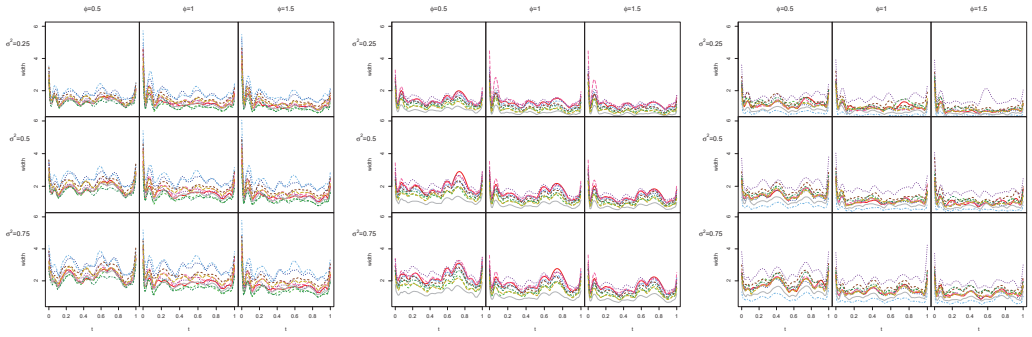

Figure 2: Band width (depth) for $n = 25$ (left), $n = 50$ (middle) and $n = 90$ (right).

curves containing the whole curve $y(t)$ [9]; here we use the modified band depth (*MBD*), that takes into account whether a portion of the curve is in the band (for details see [9]). The lower/upper limits of a (depth based) 95% prediction band are obtained by taking the pointwise (w.r.t. t) minimum/maximum of the 95% deepest curves (in the case of band depth) or of the 95% curves closest to the zero curve (in the case of using L^2 distance)

3 Simulation study

We aim to analyse the impact that trend complexity, spatial structure (via the covariance function parameters of the functional residual random field) and ordering technique have on the performance of the bootstrapping method when increasing the number of sites. Data were simulated using cubic B-splines on a spatial irregular grid (n locations) on $D = [0, 2] \times [0, 3]$ and curve domain $T = [0, 1]$. The residual functional random field was built as $e_s(t) = \sum_{j=1}^{10} \xi_j(s) B_j(t)$, where $B_j(t)$ is the j^{th} basis function evaluated at $t \in T$. The spatially correlated spline coefficients $\{\xi_j(s), s \in D\}$ were generated for each j in $1, \ldots, 10$ using the same exponential covariance function with range and scale parameters $\phi \in (0.5, 1, 1.5)$ and $\sigma^2 \in (0.25, 0.50, 0.75)$ respectively, resulting in 9 different scenarios. The drift was obtained as $m_s(t) = \alpha(t) + \beta_1(t) lon + \beta_2(t) lat$, where lon and lat are the spatial coordinates, $\alpha(t)$ is a functional intercept and $\beta_1(t), \beta_2(t)$ are functional coefficients that can be expressed in terms of B-spline basis (whose coefficients can be chosen to determine the complexity of the drift). Finally, simulated observations were built as

$$Y_s^{sim}(t) = m_s(t) + e_s(t) + \xi_s(t)$$

where $\xi(t) = \{\xi_{s_1}(t), \ldots, \xi_{s_n}(t)\} \sim N_n(0, 0.09)$ is a vector of random errors for each fixed $t \in [0, 1]$. For each simulation scenario, we generated functional data at $n = 25, 50$ and 90 nested locations. Additionally data were generated at 10 more sites used as validation stations. The FKED model (Section 1) was applied to each simulated data set to predict curves at the 10 validation sites. For each validation station $B = 500$ predictions were obtained following Section 2 and 95% prediction bands were produced using both distance and *MBD*. An example can be seen in Figure 1. To evaluate the performance of the proposal, we consider two different indicators: the width of the resulting 95% prediction interval and the proportion of the simulated curve within the interval. Figure 2 summarizes (depth) band width for all sample sizes and simulation scenarios. As one would expect, band width decreases with increasing sample size. Moreover, band width increases with increasing σ^2 and decreases slightly with increasing ϕ for a fixed value of σ^2. The depth-based band is practically always wider than the distance-based one. In terms of coverage (figure not shown here), the performance appears good and improves with increasing sample size.

4 Discussion

We propose a semi-parametric bootstrap approach that allows the construction and evaluation of simultaneous prediction bands - over T - for the functional kriging predictor with a non-constant width. The simulation study shows that the proposed technique has a good performance. We are currently investigating the effect of more complex drifts as well as alternative ways of evaluating the performance of the proposal.

References

[1] Caballero, W., Giraldo, R. and Mateu, J. 2013 *A universal kriging approach for spatial functional data*. Stoch. Env. Res. Risk Assess. 27(7), 1553–1563.

[2] Delicado, P., Giraldo, R., Comas, C. and Mateu, J. 2010 *Statistics for spatial functional data: some recent contributions*. Environmentrics, 21, 224–239.

[3] Giraldo, R., Delicado, P. and Mateu, J. 2011 *Ordinary kriging for function-valued spatial data*. Environmental and Ecological Statistics, 18(3), 411–426.

[4] Goulard, M. and Voltz, M. 1993 Geostatistical interpolation of curves: A case study in soil science. In *Geostatistics Troia '92*(ed. A. Soares), vol. 2, pp. 805–816, Kluwer Academic, Dordrecht.

[5] Ignaccolo, R., Mateu, J. and Giraldo, R. 2014 *Kriging with external drift for functional data for air quality monitoring*. Stoch. Env. Res. Risk Assess. 28(5), 1171–1186.

[6] Iranpanah, N., Mohammadzadeh, M. and Taylor, C.C. 2011 *A comparison of block and semi-parametric bootstrap methods for variance estimation in spatial statistics*. Comput. Statist. Data Anal. 55, 578–587.

[7] Menafoglio, A., Secchi, P. and Dalla Rosa, M. 2013 *A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space*. Electron. J. Stat. 7, 2209–2240.

[8] Nerini, D., Monestiez, P. and Manté C. 2010 *Cokriging for spatial functional data*. J. Multivariate Anal. 101, 409–418.

[9] Lopez-Pintado, S. and Romo, J. 2009 *On the concept of depth for Functional Data.* J. Amer. Statist. Assoc. 104(486), 718–734.

[10] Schelin, L. and Sjöstedt-de Luna, S. 2010 *Kriging prediction intervals based on semiparametric bootstrap.* Math. Geosci. 42, 985–1000