Spatial Statistics 21 (2017) 130–148

Contents lists available at ScienceDirect

Spatial Statistics

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

Bootstrap based uncertainty bands for prediction in functional kriging

Maria Franco-Villoria *, Rosaria Ignaccolo

Department of Economics and Statistics "Cognetti de Martiis", University of Torino, Italy

ARTICLE INFO

Article history:
Received 11 April 2017
Accepted 13 June 2017
Available online 23 June 2017

Keywords:
B-splines
Band depth
Functional data modeling
Generalized additive models
Geostatistics
Trace-variogram

ABSTRACT

The increasing interest in spatially correlated functional data has led to the development of appropriate geostatistical techniques that allow to predict a curve at an unmonitored location using a functional kriging with external drift model that takes into account the effect of exogenous variables (either scalar or functional). Nevertheless uncertainty evaluation for functional spatial prediction remains an open issue. We propose a semi-parametric bootstrap for spatially correlated functional data that allows to evaluate the uncertainty of a predicted curve, ensuring that the spatial dependence structure is maintained in the bootstrap samples. The performance of the proposed methodology is assessed via a simulation study. Moreover, the approach is illustrated on a well known data set of Canadian temperature and on a real data set of PM_{10} concentration in the Piemonte region, Italy. Based on the results it can be concluded that the method is computationally feasible and suitable for quantifying the uncertainty around a predicted curve. Supplementary material including R code is available online.

1. Introduction

Kriging is a well known prediction method in the geostatistics community (see e.g. Chiles and Delfiner, 2012); it allows to predict a (scalar) random field or spatial process $\{Z(s), s \in D \subseteq \mathbb{R}^2\}$ in a new spatial location s_0 given a set of observed values $Z = (Z(s_1), \ldots, Z(s_n))$, taking into account the underlying correlation structure. Spatially dependent functional data (see e.g. the last two chapters of the book by Horváth and Kokoszka, 2012) have received increasing interest over the last few

* Corresponding author.
E-mail address: maria.francovilloria@unito.it (M. Franco-Villoria).

http://dx.doi.org/10.1016/j.spasta.2017.06.005

years. Geostatistical techniques for functional data were first introduced in the pioneering work of Goulard and Voltz (1993), but the development of such techniques is rather recent. The simplest case would be that of ordinary kriging, which allows to predict a curve at an unmonitored site under the assumption of a constant mean (see e.g. Delicado et al., 2010; Giraldo et al., 2011; Nerini et al., 2010). The case of a mean function that depends on longitude and latitude was considered in Caballero et al. (2013), Menafoglio et al. (2013) and Reyes et al. (2015). In their work, Ignaccolo et al. (2014) consider more complex forms of non-stationarity, where the mean function may depend on exogenous variables (either scalar or functional), developing the so called kriging with external drift – or regression kriging – in a functional data setting.

While much effort has been put in prediction, the uncertainty of a predicted curve remains an open issue, since there is no functional version of the kriging variance. The lack of a distribution function in the functional framework leads to the use of resampling methods for confidence band calculation. In this context, Cuevas et al. (2006) consider the standard bootstrap and a smoothed version of it to obtain confidence intervals for location estimators; an informal discussion on the asymptotic validity of the bootstrap approach in a functional framework can also be found in their paper. Goldsmith et al. (2013) use a bootstrap approach to account for the uncertainty in Functional Principal Components decomposition in estimating the functional mean and constructing a confidence band for it. Further, Ferraty et al. (2010) propose using "wild bootstrapping" in the case of a nonparametric regression model with scalar response and functional covariate and derive asymptotic results; Rana et al. (2016) extend this last work to the case of α-mixing dependence. The recent paper by González-Rodríguez and Colubi (2017) shows the consistency of some bootstrap approaches for separable Hilbert-valued random elements but under the assumption of independence.

While the bootstrap theory is well established for independent data, in the spatial data setting a bootstrap procedure needs to mimic the data generating mechanism in order to reproduce the spatial dependence structure in the bootstrap samples. Mostly by extending bootstrap for time series, several variants of spatial subsampling and spatial block bootstrap methods have been proposed in the literature (see chapter 12 in Lahiri, 2003 for a good overview on resampling methods for spatial data). In classical geostatistics, it is common to assume a decomposition of data variability in large- and small-scale components, so that a "semi-parametric" bootstrap method as described in Cressie (1993, p. 493) – and inspired by Freedman and Peters (1984) and Solow (1985) – seems appropriate. The latter consists in transforming the residuals of a regression model (i.e. after estimation of the large-scale component) to remove the spatial dependence structure (small-scale) so that resampling can be done on uncorrelated data, to then re-introduce the spatial correlation on bootstrapped samples and finally add up the large-scale component. Recently, Iranpanah et al. (2011) compare the semi-parametric bootstrap with a moving block bootstrap for variance estimation of estimators in a simulation study, and point out some advantages of the semi-parametric approach in terms of precision and accuracy of the estimator. A semi-parametric bootstrap approach has also been considered in Schelin and de Luna (2010), where the focus is on the ordinary kriging predictor for data whose distribution is not necessarily Gaussian, and indeed their proposal does not need any distributional assumptions about the data generating process. While Iranpanah et al. (2011) consider the presence of a non-constant mean structure too, the main difference between the two proposals is related to the considered statistics: Iranpanah et al. (2011) suggest to create the bootstrap distribution of the spatial predictor, while Schelin and de Luna (2010) construct a bootstrap distribution for the contrast defined as the difference between the spatial predictor and the unknown value (the one we want to predict).

The literature available considers either prediction bands for functional data in the case of independent observations or prediction intervals for spatially correlated data but in a scalar framework. In this sense, we fill the gap by providing a solution for spatially correlated functional data. In this framework, however, mimicking the data generating process is expected to be more difficult than in the scalar case. In this paper, we consider the case of functional kriging with external drift (FKED) developed in Ignaccolo et al. (2014) and extend it to take into account spatial correlation when estimating the drift functional coefficients by means of an iterative algorithm. To evaluate the uncertainty of a predicted curve, we propose to extend the semi-parametric bootstrap approach for spatially correlated data introduced by Schelin and de Luna (2010) to the case of functional data, with the addition of a functional drift in the kriging model (a scalar drift was considered in Iranpanah et al.,