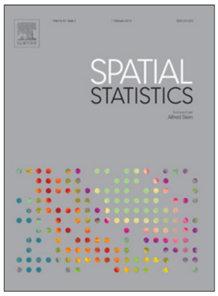
Accepted Manuscript



P-spline smoothing for spatial data collected worldwide

Fedele Greco, Massimo Ventrucci, Elisa Castelli

PII: S2211-6753(17)30294-4
DOI: https://doi.org/10.1016/j.spasta.2018.08.008
Reference: SPASTA 323

To appear in: *Spatial Statistics*

Received date : 14 November 2017
Accepted date : 18 August 2018

Please cite this article as: Greco F., Ventrucci M., Castelli E., P-spline smoothing for spatial data collected worldwide. *Spatial Statistics* (2018), https://doi.org/10.1016/j.spasta.2018.08.008

P-spline smoothing for spatial data collected worldwide

Fedele Greco[a], Massimo Ventrucci[a,*], Elisa Castelli[b]

[a]*Department of Statistical Sciences, University of Bologna, via delle Belle Arti n. 41, 40126 Bologna, Italy*
[b]*Institute of Atmospheric Sciences and Climate (CNR-ISAC), Via Piero Gobetti 101, 40129 Bologna, Italy*

Abstract

Spatial data collected worldwide from a huge number of locations is frequently used in environmental and climate studies. Spatial modelling for this type of data presents both methodological and computational challenges. In this work we illustrate a computationally efficient non-parametric framework in order to model and estimate the spatial field while accounting for geodesic distances between locations. The spatial field is modelled via penalized splines (P-splines) using intrinsic Gaussian Markov Random Field (GMRF) priors for the spline coefficients. The key idea is to use the sphere as a surrogate for the Globe, then build the basis of B-spline functions on a geodesic grid system. The basis matrix is sparse as is the precision matrix of the GMRF prior, thus computational efficiency is gained by construction. We illustrate the approach with a real climate study, where the goal is to identify the Intertropical Convergence Zone using high-resolution remote sensing data.

Keywords: smoothing, intrinsic Gaussian Markov Random field, P-spline, geodesic, ITCZ

1. Introduction

High-resolution spatial data collected worldwide, usually by means of remote sensing techniques, is wide-spread in environmental and climate studies: most of the statistical methods developed in modelling this kind of data use the sphere as a surrogate for the Globe. Modelling data collected at a global scale presents both methodological and computational challenges. The traditional toolkit for a spatial data modeller when dealing with geostatistical datasets and aiming to make predictions at unmonitored locations would suggest to apply kriging techniques (see, e.g., Banerjee et al. (2014)). These rely on the assumption of a smooth Gaussian Random Field (GRF), continuous

*Corresponding author: massimo.ventrucci@unibo.it