Spatial Statistics 14 (2015) 505–531

Contents lists available at ScienceDirect

Spatial Statistics

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

Does non-stationary spatial data always require non-stationary random fields?

Geir-Arne Fuglstad [a,*], Daniel Simpson [a], Finn Lindgren [b], Håvard Rue [a]

[a] *Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

[b] *Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom*

A R T I C L E I N F O

Article history:
Received 2 July 2015
Accepted 2 October 2015
Available online 17 October 2015

Keywords:
Annual precipitation
Penalized maximum likelihood
Non-stationary spatial modelling
Stochastic partial differential equations
Gaussian random fields
Gaussian Markov random fields

A B S T R A C T

A stationary spatial model is an idealization and we expect that the true dependence structures of physical phenomena are spatially varying, but how should we handle this non-stationarity in practice? We study the challenges involved in applying a flexible non-stationary model to a dataset of annual precipitation in the conterminous US, where exploratory data analysis shows strong evidence of a non-stationary covariance structure.

The aim of this paper is to investigate the modelling pipeline once non-stationarity has been detected in spatial data. We show that there is a real danger of over-fitting the model and that careful modelling is necessary in order to properly account for varying second-order structure. In fact, the example shows that sometimes non-stationary Gaussian random fields are not necessary to model non-stationary spatial data.

1. Introduction

There are, in principle, two sources of non-stationarity present in any dataset: the non-stationarity in the mean and the non-stationarity in the covariance structure. Classical geostatistical models

* Corresponding author.
E-mail address: fuglstad@math.ntnu.no (G.-A. Fuglstad).

http://dx.doi.org/10.1016/j.spasta.2015.10.001
2211-6753/

based on stationary Gaussian random fields (GRFs) ignore the latter, but include the former through covariates that capture important structure in the mean. The focus of non-stationary spatial modelling is non-stationarity in the covariance structure. However, it is impossible to separate the non-stationarity in the mean and the non-stationarity in the covariance structure based on a single realization, and even with multiple realizations it is challenging.

The Karhunen–Loève expansion states that under certain conditions a GRF can be decomposed into an infinite linear combination of orthogonal functions, which is weighted by independent Gaussian variables with decreasing variances. For a single realization these orthogonal functions will be confounded with the covariates in the mean, and the mean structure and the covariance structure cannot be separated. This can give apparent long range dependencies and global non-stationarity if spatially varying covariates are missing in the mean. Such spurious global non-stationarity and its impact on the local estimation of non-stationarity is an important topic in the paper.

A high degree of flexibility in the covariance structure combined with weak information about the covariance structure in the data makes overfitting by interpreting ostensible patterns in the data as non-stationarity a critical issue. However, the most important point from an applied viewpoint is the computational costs of running a more complex model versus the scientific gain. Non-stationarity in the mean is computationally cheap, whereas methods for non-stationarity in the covariance structure are much more expensive. This raises two important questions: How much do we gain by including non-stationarity in the covariance structure? What sort of non-stationarity, if any, is most appropriate for the problem at hand?

The computational cost of non-stationary models usually comes from a high number of highly dependent parameters that makes it expensive to run MCMC methods or likelihood optimizations, but the challenges with non-stationary models are not only computational. Directly specifying non-stationary covariance functions is difficult and we need other ways of constructing models. Additionally, we need to choose where to put the non-stationarity. Should we have non-stationarity in the range, the anisotropy, the marginal variance, the smoothness or the nugget effect? And how do we combine it all to a valid covariance structure?

1.1. Non-stationarity

Most of the early literature on non-stationary methods deals with data from environmental monitoring stations where multiple realizations are available. In this situation it is possible to calculate the empirical covariances between observed locations, possibly accounting for temporal dependence, and finding the required covariances through, for example, shrinkage towards a parametric model (Loader and Switzer, 1989) or kernel smoothing (Oehlert, 1993). It is also possible to deal efficiently with a single realization with the moving window approach of Haas (1990a,b, 1995), but this method does not give valid global covariance structures.

However, the most well-known method from this time period is the deformation method of Sampson and Guttorp (1992), in which an underlying stationary process is made non-stationary by applying a spatial deformation. The original formulation has been extended to the Bayesian framework (Damian et al., 2001, 2003; Schmidt and O'Hagan, 2003), to a single realization (Anderes and Stein, 2008), to covariates in the covariance structure (Schmidt et al., 2011) and to higher dimensional base spaces (Bornn et al., 2012).

Another major class of non-stationary methods is based on the process convolution method developed by Higdon (1998) and Higdon et al. (1999). In this method a spatially varying kernel is convolved with a white noise process to create a non-stationary covariance structure. Paciorek and Schervish (2006) looked at a specific case where it is possible to find a closed form expression for a Matérn-like covariance function and Neto et al. (2014) used a kernel that depends on wind direction and strength to control the covariance structure. The process convolution methods have also been extended to dynamic multivariate processes (Calder, 2007, 2008) and spatial multivariate processes (Kleiber and Nychka, 2012).

It is possible to take a different approach to non-stationarity, where instead of modelling infinite-dimensional Gaussian processes one uses a linear combination of basis functions and models the covariance matrix of the coefficients of the basis functions (Nychka et al., 2002, 2015). One