Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

High resolution simulation of nonstationary Gaussian random fields

William Kleiber

Department of Applied Mathematics, University of Colorado, Boulder, CO, United States

ARTICLE INFO

Article history:
Received 20 March 2015
Received in revised form 21 January 2016
Accepted 13 March 2016
Available online xxxx

Keywords:
Circulant embedding
Deformation
Nonstationary
Simulation
Stationary
Warping

ABSTRACT

Simulation of random fields is a fundamental requirement for many spatial analyses. For small spatial networks, simulations can be produced using direct manipulations of the covariance matrix. Larger high resolution simulations are most easily available for stationary processes, where algorithms such as circulant embedding can be used to simulate a process at millions of locations. We discuss an approach to simulating high resolution nonstationary Gaussian processes that relies on generating a stationary random field followed by a nonlinear deformation to produce a nonstationary field. A spatially varying variance coefficient accounts for local scale effects. The nonstationary covariance function is estimated nonparametrically, and the deformation function is then estimated in a variational framework. We illustrate the proposed approach on synthetic datasets, a challenging temperature dataset over the state of Colorado and a regional climate model over North America.

1. Introduction

Spatial analyses typically involve three common goals: the first is surface estimation based on incomplete or noisy observations, the second is interpretation of a spatial model to gain scientific insight of a particular process, and the third is simulation. Simulation plays an important role for both of the first two goals; in surface estimation simulation can be used to quantify predictive uncertainty via conditional simulation and can also yield field realizations consistent with a partially observed process. In the second goal, simulation can give insight into the statistical properties of a process, such as its spatial length scale, smoothness, level crossings or extrema.

Suppose interest focuses on simulating a random process $Z(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d$. Without loss of generality we assume $Z(\mathbf{s})$ is a mean zero process; otherwise we would simulate a mean zero field onto which we add a nontrivial mean function. We additionally assume $Z(\mathbf{s})$ is a Gaussian process, whose stochastic behavior is then fully known once the covariance function $C(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2))$ is specified.

Simulation of random fields can be either unconditional or conditional. Unconditional simulation is simply direct simulation of the process $Z(\mathbf{s})$. Conditional simulation, on the other hand, is the simulation of $Z(\mathbf{s})$ conditioned on some observations in order to generate plausible realizations that are consistent with the partially observed process. In this manuscript, we mainly focus on unconditional simulation, noting that conditional simulation can be directly formulated via unconditional simulations (Journel, 1974).

It seems there are at least three possible ways to simulate nonstationary field—first, use a stochastic model that allows for nonstationarity and directly simulate from this (e.g., a nonstationary Matérn); this requires a Cholesky decomposition of the covariance matrix which is infeasible for large simulation grids. The second is to explicitly build such a process from a set of

E-mail address: william.kleiber@colorado.edu.

http://dx.doi.org/10.1016/j.csda.2016.03.005
0167-9473/

basis functions, e.g., using a Karhunen–Loéve expansion or polynomial chaos expansion. The final method is one explored in this manuscript, where a mapping is developed between the nonstationary process and a stationary process, whence simulations involve only the straightforward stationary generation. We note that there is some literature on simulation methods for nonstationary intrinsic random functions, although we do not focus on these here (Stein, 2001, 2002, and references therein).

Our proposal relies on the spatial deformation approach of Sampson and Guttorp (1992) to achieve a mapping between the nonstationary and stationary plane. The basic idea is that in the stationary plane high resolution simulations are easy to produce from some existing method such as circulant embedding, and then the inverse transformation results in a high dimensional nonstationary simulation. Note we are primarily concerned with simulation, whereas Sampson and Guttorp (1992) and ensuing literature has focused almost exclusively on modeling and kriging. The type of nonstationarity that can be captured by deformation is either known as stationary isotropic reducible or stationary reducible (Perrin and Senoussi, 1999, 2000). Deformation has successfully been used to capture a large number of physical processes (Guttorp et al., 1992; Monestiez et al., 1993; Guttorp and Sampson, 1994; Brown et al., 1994; Guttorp et al., 1994; Meiring, 1995), but not all types of nonstationarity can be reduced to stationarity this way, e.g., a Matérn covariance with spatially varying smoothness cannot be (Stein, 2005; Paciorek and Schervish, 2006). Note some similarities with the time-deformation method in economics (Barndorff-Nielsen and Shepard, 2006).

We consider two motivating environmental examples: the first is application in the field of stochastic weather generators, while the second involves statistical emulation or analysis of a regional climate model. Stochastic weather generators are probabilistic models whose simulations behave statistically similarly to observations (Wilks and Wilby, 1999). These simulators are used primarily in the hydrologic and climate sciences to perform downscaling or impact assessments (Semenov and Barrow, 1997). Typically, weather realizations are required on a grid, sometimes requiring simulation over very large geographical regions at high resolutions (Serinaldi and Kilsby, 2014). Following the technical development of our approach, we illustrate its implementation on a challenging temperature dataset over the state of Colorado. For the second example, we consider the problem of stochastically simulating fields of regional climate model (RCM) output that are consistent with RCM runs. These products are crucial for climate forecasting or model emulation, that is, using a statistical model as a fast surrogate for a computationally expensive physical climate model. RCMs are usually run coupled with a general circulation model (GCM) in order to better represent local nonstationarities that are driven by local geographical effects that are not well represented in a coarse GCM grid. The ability to quickly generate stochastic realizations from a high resolution nonstationary process is of fundamental importance for both of these applications.

2. Simulation of random fields

Our approach to nonstationary random field simulation relies on fast simulation algorithms for stationary or isotropic random fields. We begin this section with a brief overview of some classic algorithms for stationary simulation, Schlather (2012) and Kroese and Botev (2013) give nice recent overviews of some of these approaches among others.

2.1. Stationary simulation

Momentarily suppose the random field $Z(\mathbf{s})$ is stationary, that is, $C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{s}_1 - \mathbf{s}_2)$ is a function of the lag vector separating two spatial locations. The spectral method is a traditional approach to approximately simulating stationary random fields. The spectral method (Shinozuka and Jan, 1972) relies on using the spectral representation of a random field requiring continuity of the covariance function,

$$Z(\mathbf{s}) = \mathrm{Re}\left[\int \exp(2\pi \mathrm{i}\boldsymbol{\omega}'\mathbf{s})\mathrm{d}Y(\boldsymbol{\omega})\right] \tag{1}$$

where $\boldsymbol{\omega} \in \mathbb{R}^d$ and $\mathrm{d}Y(\boldsymbol{\omega})$ is a complex-valued Gaussian measure with zero mean and whose pointwise variance is $\mathcal{F}(C)(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}$, $\mathcal{F}$ denoting the Fourier transform. Simulations of $Z(\mathbf{s})$ can then be approximately generated by using a discrete approximation to the integral representation (1). The turning bands method, originally suggested by Matheron (1973), relies on simulating a multidimensional isotropic Gaussian random field by summing simulations from one-dimensional processes that have been embedded in more than one dimension (Mantoglou and Wilson, 1982; Dietrich, 1995; Gneiting, 1996, 1999). In particular, realizations for $d = 2, 3$ are obtained via

$$Z(\mathbf{s}) = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Z_i(\mathbf{s} \cdot \mathbf{e}_i)$$

where $\{Z_i(\cdot)\}_{i=1}^L$ are mutually independent one-dimensional processes, $\{\mathbf{e}_i\}_{i=1}^L$ are unit vectors and $\cdot$ indicates the dot product. The key difficulty turns out to be identifying the one-dimensional covariance function that corresponds to the desired multidimensional covariance structure, with the most common two-dimensional case requiring solving an Abel integral equation (Gneiting, 1998). For stationary fields whose covariance can be represented as a convolution, $C(\mathbf{h}) = \int g(\mathbf{s})g(\mathbf{s}+\mathbf{h})\mathrm{d}\mathbf{s}$, one can use the so-called random coin (or dilution) method to generate approximate realizations (Chilès and Delfiner, 1999; Schlather, 2012). Ehm et al. (2004) examined conditions on compactly supported covariance functions

Please cite this article in press as: Kleiber, W., High resolution simulation of nonstationary Gaussian random fields. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.03.005