Spatial Statistics 18 (2016) 105–124

Contents lists available at ScienceDirect

Spatial Statistics

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

Multi-level restricted maximum likelihood covariance estimation and kriging for large non-gridded spatial datasets

Julio E. Castrillón-Candás [a], Marc G. Genton [b,*], Rio Yokota [c]

[a] SRI Center for Uncertainty Quantification in Computational Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[b] Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[c] Tokyo Institute of Technology Global Scientific and Computing Center, 2-12-1 i7-2 O-okayama Meguro-ku, 152-8550, Tokyo, Japan

ARTICLE INFO

Article history:
Received 9 October 2015
Accepted 19 October 2015
Available online 10 November 2015

Keywords:
Fast Multipole Method
Hierarchical basis
High performance computing
Sparsification of covariance matrices

ABSTRACT

We develop a multi-level restricted Gaussian maximum likelihood method for estimating the covariance function parameters and computing the best unbiased predictor. Our approach produces a new set of multi-level contrasts where the deterministic parameters of the model are filtered out thus enabling the estimation of the covariance parameters to be decoupled from the deterministic component. Moreover, the multi-level covariance matrix of the contrasts exhibits fast decay that is dependent on the smoothness of the covariance function. Due to the fast decay of the multi-level covariance matrix coefficients only a small set is computed with a level dependent criterion. We demonstrate our approach on problems of up to 512,000 observations with a Matérn covariance function and highly irregular placements of the observations. In addition, these problems are numerically unstable and hard to solve with traditional methods.

* Corresponding author.
E-mail addresses: uvel@alum.mit.edu (J.E. Castrillón-Candás), marc.genton@kaust.edu.sa (M.G. Genton), rioyokota@gsic.titech.ac.jp (R. Yokota).

http://dx.doi.org/10.1016/j.spasta.2015.10.006
2211-6753/

1. Introduction

Consider the following model for a Gaussian spatial random field Z:

$$Z(\mathbf{s}) = \mathbf{m}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + \epsilon(\mathbf{s}), \qquad \mathbf{s} \in \mathbb{R}^d, \tag{1}$$

where $\mathbf{m} \in \mathbb{R}^p$ is a known function of the spatial location $\mathbf{s}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown vector of coefficients, and ϵ is a stationary mean zero Gaussian random field with parametric covariance function $C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}) = \mathrm{cov}\{\epsilon(\mathbf{s}), \epsilon(\mathbf{s}')\}$ having an unknown vector $\boldsymbol{\theta} \in \mathbb{R}^w$ of parameters. We observe the data vector $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^{\mathrm{T}}$ at locations $\mathbb{S} := \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, where $\mathbf{s}_1 \neq \mathbf{s}_2 \neq \mathbf{s}_3 \neq \cdots \neq \mathbf{s}_{n-1} \neq \mathbf{s}_n$, and wish to: 1) estimate the unknown vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$; and 2) predict $Z(\mathbf{s}_0)$, where $\mathbf{s}_0$ is a new spatial location. These two tasks are particularly challenging when the sample size n is large.

To address the estimation part, let $\mathbf{C}(\boldsymbol{\theta}) = \mathrm{cov}(\mathbf{Z}, \mathbf{Z}^{\mathrm{T}}) \in \mathbb{R}^{n\times n}$ be the covariance matrix of $\mathbf{Z}$ and assume it is nonsingular for all $\boldsymbol{\theta} \in \mathbb{R}^w$. Define $\mathbf{M} = \big(\mathbf{m}(\mathbf{s}_1) \ldots \mathbf{m}(\mathbf{s}_n)\big)^{\mathrm{T}} \in \mathbb{R}^{n\times p}$ and assume it is of full rank, p. The model (1) leads to the vectorial formulation

$$\mathbf{Z} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{\epsilon}$ is a Gaussian random vector, $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}))$. Then the log-likelihood function is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det\{\mathbf{C}(\boldsymbol{\theta})\} - \frac{1}{2}(\mathbf{Z} - \mathbf{M}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{Z} - \mathbf{M}\boldsymbol{\beta}), \tag{3}$$

which can be profiled by generalized least squares with

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{M}\}^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{Z}. \tag{4}$$

A consequence of profiling is that the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ then tends to be biased. A solution to this problem is to use restricted maximum likelihood (REML) estimation which consists in calculating the log-likelihood of $n - p$ linearly independent contrasts, that is, linear combinations of observations whose joint distribution does not depend on $\boldsymbol{\beta}$, from the set $\mathbf{Y} = \{\mathbf{I}_n - \mathbf{M}(\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}\}\mathbf{Z}$. In this paper, we propose a new set of contrasts that lead to significant computational benefits (with good accuracy) when computing the REML estimator of $\boldsymbol{\theta}$ for large sample size n.

To address the prediction part, consider the best unbiased predictor $\hat{Z}(\mathbf{s}_0) = \lambda_0 + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}$ where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^{\mathrm{T}}$. The unbiasedness constraint implies $\lambda_0 = 0$ and $\mathbf{M}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{m}(\mathbf{s}_0)$. The minimization of the mean squared prediction error $E[\{Z(\mathbf{s}_0) - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}\}^2]$ under the constraint $\mathbf{M}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{m}(\mathbf{s}_0)$ yields

$$\hat{Z}(\mathbf{s}_0) = \mathbf{m}(\mathbf{s}_0)^{\mathrm{T}}\hat{\boldsymbol{\beta}} + \mathbf{c}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{Z} - \mathbf{M}\hat{\boldsymbol{\beta}}), \tag{5}$$

where $\mathbf{c}(\boldsymbol{\theta}) = \mathrm{cov}\{\mathbf{Z}, Z(\mathbf{s}_0)\} \in \mathbb{R}^n$ and $\hat{\boldsymbol{\beta}}$ is defined in (4). In this paper, we propose a new transformation of the data vector $\mathbf{Z}$ leading to a decoupled multi-level description of the model (1) without any loss of structure. This multi-level representation leads to significant computational benefits when computing the kriging predictor $\hat{Z}(\mathbf{s}_0)$ in (5) for large sample size n.

Previous work has been performed to maximize (3). The classical technique is to compute a Cholesky factorization of $\mathbf{C}$. However, this requires $\mathcal{O}(n^2)$ memory and $\mathcal{O}(n^3)$ computational steps, thus is impractical for large scale problems.

Under special structures of the covariance matrix, i.e., fast decay of the covariance function, a tapering technique can be used to sparsify the covariance matrix and thus increase memory and computational efficiency (Furrer et al., 2006; Kaufman et al., 2008). These techniques are good when applicable but tend to be restrictive. For a review of various approaches to spatial statistics for large datasets, see Sun et al. (2012). Recently we have seen the advent of solving the optimization problem (3) from a computational numerical perspective. Anitescu et al. (2012) developed a matrix-free approach for computing the maximum of the log-likelihood (3) based on a stochastic programming reformulation. This method relies on Monte Carlo approximation of the derivative of the score function with respect to the covariance parameters $\boldsymbol{\theta}$ to compute the maximization (3). The authors show promising results for a grid geometry of the placement of the observations. However for a non-grid geometry the cost of computing the preconditioner becomes $\mathcal{O}(n^2)$ and it is not clear how many