Spatial Statistics 12 (2015) 31–49

Contents lists available at ScienceDirect

Spatial Statistics

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

Equivalent kriging

William Kleiber[a,*], Douglas W. Nychka[b]

[a] Department of Applied Mathematics, University of Colorado, Boulder, CO, United States

[b] Geophysical Statistics Project, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, United States

ARTICLE INFO

Article history:
Received 31 July 2014
Accepted 22 January 2015
Available online 9 February 2015

Keywords:
Equivalent kernel
Kriging
Massive dataset

ABSTRACT

Most modern spatially indexed datasets are very large, with sizes commonly ranging from tens of thousands to millions of locations. Spatial analysis often focuses on spatial smoothing using the geostatistical technique known as kriging. Kriging requires covariance matrix computations whose complexity scales with the cube of the number of spatial locations, making analysis infeasible or impossible with large datasets. We introduce an approach to kriging in the presence of large datasets called *equivalent kriging*, which relies on approximating the kriging weight function using an equivalent kernel, requiring presence of a nontrivial nugget effect. Resulting kriging calculations are extremely fast and feasible in the presence of massive spatial datasets. We derive closed form kriging approximations for multiresolution classes of spatial processes, as well as under any stationary model, including popular choices such as the Matérn. The theoretical justification for equivalent kriging also leads to a correction term for irregularly spaced observations that also reduces edge effects near the domain boundary. For large sample sizes, equivalent kriging is shown to outperform covariance tapering in an example. Equivalent kriging is additionally illustrated on multiple simulated datasets, and a monthly average precipitation dataset whose size prohibits traditional geostatistical approaches.

* Corresponding author.
E-mail address: william.kleiber@colorado.edu (W. Kleiber).

http://dx.doi.org/10.1016/j.spasta.2015.01.004
2211-6753/

1. Introduction

In the era of big data, spatially indexed datasets are especially prone to size-induced limitations. Indeed, modern atmospheric, hydrologic, ecological and environmental datasets are increasingly large and complex, often involving data sited at between thousands and millions of locations. A major goal when faced with such complex and noisy data is estimating the underlying physical process whose observations are subject to noise. In geostatistics, the main tool used for surface estimation is kriging, which is the linear predictor that minimizes predictive squared loss, assuming known model parameters.

The main obstacle for kriging on large datasets is solving a linear system of equations involving the spatial covariance matrix. This covariance matrix is usually dense and unstructured, and has size that scales as the square of the number of spatial locations. Over the past decade there have been a number of proposed approaches to kriging on large datasets. Many of the most popular techniques rely on a low rank representation for the spatial covariance matrix. For instance, fixed rank kriging achieves low rank by representing spatial covariances via a small set of basis functions in the observation domain (Cressie and Johannesson, 2008). Similarly, predictive processes use a conditional expectation representation at a small set of knots in the observation domain that leads to a low rank type setup (Banerjee et al., 2008). An alternative approach is covariance tapering, using a compactly supported function to impose sparsity in the covariance matrix (Furrer et al., 2006; Kaufman et al., 2008). One of the criticisms of low rank ideas is that they tend to capture low frequency behavior quite well, but are unable to model well high frequency behavior (Finley et al., 2009). To overcome this problem, an idea that retains computational feasibility is to use a low rank representation of spatial covariance, and superimpose a high frequency term that is generated by a compactly supported covariance; Sang and Huang (2012) named this approach a full scale approximation, see also Stein (2008). Finally, a simple alternative is to window the data and perform kriging locally; Stein (2014) found this approach to be favorable to low rank methods in approximating likelihoods.

A more recent idea involves approximating a Gaussian random field by a Gaussian Markov random field (Lindgren et al., 2011). This approach is computationally extremely fast for very large datasets, but is designed for processes with Matérn covariances, and can only approximate the restrictive subclass whose smoothnesses are integer plus one half values. A somewhat related approach is a multiresolution representation of the underlying stochastic field, a specific class of which has been developed very recently by Nychka et al. (in press), which they term LatticeKrig. A computationally expensive step common to many models is evaluating the likelihood (or Bayesian posterior) to determine variance and covariance parameters; some approaches to likelihood approximations have been proposed, involving an approximate gridding of the observations and using techniques for regular lattices (Fuentes, 2007). Sun et al. (2012) give an overview of some of the aforementioned techniques and others.

We propose a novel approach to kriging over large datasets called equivalent kriging. Equivalent kriging relies on approximating the kriging weights using an equivalent kernel via ideas that have previously been confined to the spline literature (Silverman, 1984). This approximation's primary limitation is that it is only valid with a nonzero nugget effect, akin to spline smoothing. The equivalent kernel is available in closed form for multiresolution processes, and has a representation as a Hankel transform for kriging with any isotropic covariance function. Specifically, we can approximate kriging under a Matérn covariance with an arbitrary smoothness, improving upon many of the previously proposed techniques. We explore both gridded and irregularly spaced data situations. Estimation can proceed by cross-validation or generalized cross-validation, as the smoothing matrix is quickly computable using the equivalent kernel approximation. We follow the technical discussion with data examples, empirically illustrating the computational advantages of equivalent kriging over traditional kriging.

As a suggestion of the timing improvements of equivalent kriging over traditional kriging, Fig. 1 illustrates a simple example. The goal is to smooth a set of noisy observations on an $n \times n$ grid by kriging or equivalent kriging using an exponential covariance model with a nugget effect. The grid is on $[0, 2\pi]^2$ with the exponential scale set to unity. Timing comparisons are shown in seconds for between approximately $n^2 = 700$ and 10 000 total locations. For even moderately large datasets, equivalent