Spatial Statistics 16 (2016) 152–167

Contents lists available at ScienceDirect

Spatial Statistics

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

A comparison of plug-in predictors for lognormal kriging

Christien Thiart *, Linda M. Haines

Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa

ARTICLE INFO

Article history:
Received 15 October 2015
Accepted 6 March 2016
Available online 16 March 2016

Keywords:
Lognormal kriging
'plug-in' estimators
Empirical mean squared prediction error
Optimal point prediction
Ordinary kriging
Unknown covariance parameters

ABSTRACT

When skewed spatial data are encountered, a common procedure is to invoke lognormal kriging. The resultant naïve lognormal predictors are however biased. A family of optimal predictors which address this unbiasedness has been introduced into the literature by De Oliveira (2006). The behaviour of this family is well documented in the case of known covariance parameters. In contrast, when the covariance parameters are not known and must be estimated, the behaviour of the predictors is unknown. In the present paper, the performance of a subset of the family of predictors formulated by De Oliveira (2006), together with a predictor introduced by Abt et al. (1999), with 'plug-in' estimates of the covariance parameters introduced into the appropriate theoretical formulae, is investigated by means of a simulation study. Results based on eighteen scenarios describing a range of covariance structures and three performance criteria are presented and, in addition, results for a real world example are reported. The main features to emerge from the study are that the naïve predictor is negatively biased, that the lognormal predictor is representative of key predictors in the class introduced by De Oliveira (2006) and that the empirical mean squared prediction error severely underestimates the true value.

* Corresponding author.
E-mail addresses: christien.thiart@uct.ac.za (C. Thiart), linda.haines@uct.ac.za (L.M. Haines).

http://dx.doi.org/10.1016/j.spasta.2016.03.002

1. Introduction

Kriging is used extensively in the modelling of spatial data. The approach is rooted in the notion of minimizing the mean squared prediction error at an unobserved location. The only assumption made in the broad-based derivation of kriging is that of stationarity in the underlying error structure. When the data are assumed to follow a multivariate Gaussian distribution, the expectation of the best predictor is both linear and unbiased. When the data are not multivariate Gaussian however, the predictor may not be linear or unbiased. It is then common practice to adopt a transformation of the non-normal data to yield data which follow a Gaussian or an approximate Gaussian distribution. For skewed spatial data, it is usual to assume that the observations are log-normally distributed and can therefore be transformed by taking logarithms to create a pseudo-Gaussian random field. Kriging can then be performed on the transformed data and best linear unbiased predictors within the Gaussian setting obtained. This procedure is commonly known as lognormal kriging and is, in essence, a special case of trans-Gaussian kriging (Matheron, 1974; Journel, 1980; Cressie, 1993; Chiles and Delfiner, 1999; Diggle and Ribeiro, 2007). Many researchers advocate back-transforming the Gaussian predictors to yield lognormal predictors but some do not agree with this approach, preferring to leave the results within the Gaussian setting.

The moments of the log-normal, that is log-Gaussian, distribution are well-documented and can be formulated explicitly in terms of the moments of the underlying Gaussian distribution (Aitchison and Brown, 1957). Thus it is straightforward to show that the lognormal predictor obtained directly by taking the exponential of the corresponding Gaussian predictor is biased. As a consequence, other lognormal predictors which in some sense counter this bias, have been proposed and analytic expressions for the associated biases and mean squared prediction errors (MSPEs) have been derived under the assumption that the covariance parameters embedded in the model are known (Cressie, 2006; De Oliveira, 2006). In practice however the covariance parameters are invariably not known and must be estimated within the Gaussian framework. Lognormal predictors and the attendant mean squared prediction errors can then be evaluated by 'plugging' estimates of the covariance parameters into the appropriate expressions derived under the assumption that the parameters are known. A good predictor in the Gaussian setting does not necessarily transform to a good predictor in the lognormal setting however and, furthermore, the behaviour of the empirical mean squared prediction errors associated with the lognormal predictors would seem to be unclear. These important points are emphasized in the book by Rossi and Deutsch (2014).

De Oliveira (2006) introduced a family of lognormal point predictors which are optimal under a range of criteria and which are expressed in terms of an unknown mean but known covariance parameters. In the present study, the properties of five predictors belonging to this family and a sixth predictor introduced by Abt et al. (1999) are examined for settings in which the covariance parameters of the underlying model are not known and must be estimated. More specifically, the performance of the six lognormal predictors with respect to bias and mean squared prediction error under eighteen scenarios based on the parameters of an exponential covariance model is investigated by means of a simulation study. The paper is organized as follows. The basic notions which underpin lognormal kriging are presented in the section entitled Theory. The simulation study is described in the next section and the results are summarized and discussed in the two sections which then follow. A case study is reported in the penultimate section and some conclusions and pointers for further research are given in the final section.

2. Theory

2.1. Setting

Let $\boldsymbol{Z} = [Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)]^T$ denote a vector of observed responses of an attribute Z at n locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, in a domain $\mathcal{D}$ and suppose that a prediction of the response at an unobserved location $\boldsymbol{s}_0$, that is $Z_0 = Z(\boldsymbol{s}_0)$, is required. Assume that the vector $[Z_0, \boldsymbol{Z}]^T$ can be expressed as the exponential of a vector $[Y_0, \boldsymbol{Y}]^T$, that is $[Z_0, \boldsymbol{Z}]^T = \exp([Y_0, \boldsymbol{Y}]^T)$, where $[Y_0, \boldsymbol{Y}]^T = [Y(\boldsymbol{s}_0), Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)]^T$ follows