



Investigating ‘optimal’ kriging variance estimation using an analytic and a bootstrap approach

by C. Thiart*, M.Z. Ngwenya†, and L.M. Haines*

Synopsis

Kriging is an interpolation technique for predicting unobserved responses at target locations from observed responses at specified locations. Kriging predictors are best linear unbiased predictors (BLUPs) and the precision of the BLUP is assessed by the mean square prediction error (MSPE), commonly known as the kriging variance. Both the BLUP and the MSPE depend on the covariance function describing the spatial correlation between locations and on specific parameters. The parameters are usually treated as known, whereas in practice they invariably have to be estimated and the empirical BLUP (that is, the EBLUP) so obtained. The empirical or estimated mean square prediction error (EMSPE), or the so called ‘plug-in’ kriging variance estimator, underestimates the true kriging variance of the EBLUP, at least in general. In this paper five estimators for the kriging variance of the EBLUP are considered and compared by means of a simulation study in which a Gaussian distribution for the responses, an exponential structure for the covariance function, and three levels of spatial correlation – weak, moderate, and strong – are adopted. The Prasad-Rao estimator obtained using restricted or residual maximum likelihood (REML) is recommended for moderate and strong spatial correlation and the Kacker-Harville estimator for weak correlation in the random fields.

Keywords

kriging, covariance parameters, mean square prediction error.

Introduction

The problem of accommodating correlation between observations made at different locations permeates design, analysis, and prediction for spatial data. In two seminal papers, Krige formulated a statistical model for such data, which comprises a deterministic component reflecting the underlying trends and an error component capturing correlations between observations taken at separate locations (Krige, 1951), and invoked his model to obtain predictions at unobserved locations (Krige, 1962). Krige’s work provided the basis for best linear unbiased prediction within the context of spatial data and was formalized more rigidly and mathematically by Matheron (1963). An excellent overview of the historical development of the methodology, broadly referred to as kriging, is provided by Cressie (1990).

The kriging approach has been adapted seamlessly to accommodate more modern statistical practice and is used extensively in a

vast range of areas. A particular problem, that of using best linear unbiased prediction to predict a response at an unobserved location when the parameters describing the correlation structure between locations are unknown is, however, not fully resolved. To be specific, an obvious solution to this problem is to ‘plug’ an available estimate of the parameters into the expression for the predictor with parameters known. Unfortunately, the distribution of the resultant predictor, termed the empirical best linear unbiased predictor or EBLUP, is intractable, and more particularly the mean square prediction error associated with the EBLUP cannot be expressed analytically.

In the present study, five approaches to approximating the mean square prediction error of the EBLUP, the ‘plug-in’, those of Kacker and Harville (1984) and Prasad and Rao (1990), and two taken from den Hertog *et al.* (2006) and based on bootstrapping, are introduced and compared using a simulation study that accommodates weak, moderate, and strong spatial correlation. The paper is organized as follows. Basic notions and formulae underpinning the methodology are presented first. The simulation study is then described, and the attendant results summarized. Finally, some broad conclusions and pointers for further research are given.

Preliminaries

Kriging basics

Suppose that observations are made on an attribute Z at the n spatial locations x_i , $i = 1, \dots, n$, in a designated geographic region D . Let $Z(x_g)$ denote the observed response at a

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

† Biometry Unit, Agricultural Research Council, Stellenbosch, South Africa.

© The Southern African Institute of Mining and Metallurgy, 2014. ISSN 2225-6253.

Investigating ‘optimal’ kriging variance estimation using an analytic and a bootstrap approach

generic location x_g where $x_g \in D \subset R^2$ and let $Z(x)$ where $x = (x_1, \dots, x_n)$ denote the $n \times 1$ vector of responses at the n locations. Suppose further that interest centres on predicting the response $Z(x_0)$ at a location x_0 in D . Then it is usual to obtain such a predicted value by introducing a weighted sum of the observations of the form

$$\hat{Z}(x_0) = \sum_{i=1}^{N(x_0)} \lambda_i \cdot Z(x_i) = \boldsymbol{\lambda}^T \mathbf{Z}(x), \quad [1]$$

where λ_i is the weight associated with the i th observation

$Z(x_i)$ with $\sum_{i=1}^{N(x_0)} \lambda_i = 1$, $\boldsymbol{\lambda}$ is the $n \times 1$ vector $(\lambda_1, \dots, \lambda_{N(x_0)})$ and $N(x_0)$ is the number of points in a search neighbourhood around the prediction location x_0 .

Now, assuming second-order stationarity, the kriging model can be formulated as

$$\mathbf{Z}(x) = \mathbf{F}(x)\boldsymbol{\beta} + \mathbf{e}(x) \quad [2]$$

where $\mathbf{F}(x)$ is an $n \times (k + 1)$ matrix with rows which usually depend on the spatial coordinates alone and/or other explanatory variables, $\boldsymbol{\beta}$ is a vector of $k + 1$ unknown parameters, and $\mathbf{e}(x)$ is an $n + 1$ vector of error terms with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The covariance between any two points is taken here to be a function of the Euclidean distance between the two points and the vector of parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$, corresponding, in order, to the nugget, the partial sill, and the range. Other functions for the covariance between points, such as the exponential and the Matérn, which depend on more or fewer parameters, can also be invoked (Diggle and Ribeiro, 2007). The covariance between any two points, x_i and x_j is denoted as $Cov[Z(x_i), Z(x_j)] = C(x_i - x_j; \boldsymbol{\theta})$ and for $i, j = 1, \dots, n$ are the elements of the variance-covariance matrix.

The best linear unbiased predictor (BLUP) at location x_0 under the kriging model in Equation [2] is obtained by minimizing the mean square prediction error, $E[\boldsymbol{\lambda}^T \mathbf{Z}(x) - Z(x_0)]^2$, with respect to the weights $\boldsymbol{\lambda}$ under the condition of unbiasedness, that is $E[\hat{Z}(x_0)] = Z(x_0)$. Thus, assuming the parameter $\boldsymbol{\theta}$ is known, the kriging weight vector $\boldsymbol{\lambda}$ is given by

$$\boldsymbol{\lambda} = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{c}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{F} \Delta^{-1} (\mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{c}(\boldsymbol{\theta}) - \mathbf{f}_0)$$

where $\mathbf{c}(\boldsymbol{\theta})$ is the $n \times 1$ vector $(C(x_0 - x_1; \boldsymbol{\theta}), \dots, C(x_0 - x_n; \boldsymbol{\theta}))$, $\Delta = \mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{F}$ and $\mathbf{f}_0 \equiv \mathbf{F}^T \boldsymbol{\lambda}$. It then follows that the BLUP at x_0 is given by

$$\hat{Z}(x_0; \boldsymbol{\theta}) = \mathbf{f}_0^T \hat{\boldsymbol{\beta}} + \mathbf{c}(\boldsymbol{\theta})^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\mathbf{Z}(x) - \mathbf{F} \hat{\boldsymbol{\beta}})$$

and its mean square prediction error, denoted MSPE and termed the kriging variance, by

$$\begin{aligned} Var[\hat{Z}(x_0; \boldsymbol{\theta}) - Z(x_0)] &= C(\mathbf{0}; \boldsymbol{\theta}) - \\ &\mathbf{c}(\boldsymbol{\theta})^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{c}(\boldsymbol{\theta}) + [\mathbf{f}_0 - \mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{c}(\boldsymbol{\theta})]^T \\ &(\mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{F})^{-1} [\mathbf{f}_0 - \mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{c}(\boldsymbol{\theta})] \end{aligned} \quad [3]$$

where $\hat{\boldsymbol{\beta}}$ is the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$, that is

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{F})^{-1} \mathbf{F}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{Z}(x).$$

In practice, however, the parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$ are rarely known and must be estimated. Early studies adopted

least squares estimators of $\boldsymbol{\theta}$ based on fitting appropriate nonlinear models to the semivariogram data. However, a more modern approach involves maximizing the log-likelihood or the restricted log-likelihood with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to give the maximum likelihood estimator (MLE), $\hat{\boldsymbol{\theta}}_{MLE}$, or the restricted maximum likelihood (REML) estimator, $\hat{\boldsymbol{\theta}}_{REML}$, of $\boldsymbol{\theta}$ respectively. The kriging predictor at the location x_0 is then obtained by ‘plugging’ an estimate of $\boldsymbol{\theta}$, denoted generically $\hat{\boldsymbol{\theta}}$, into the expression for the BLUP to give the ‘empirical’ BLUP, termed the EBLUP, as

$$\hat{Z}_E(x_0; \hat{\boldsymbol{\theta}}) = \mathbf{f}_0^T \hat{\boldsymbol{\beta}} + \mathbf{c}(\hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1} (\mathbf{Z}(x) - \mathbf{F} \hat{\boldsymbol{\beta}}).$$

In the same spirit, the kriging variance or MSPE of the predictor at x_0 can be estimated by plugging $\hat{\boldsymbol{\theta}}$ into the expression for the MSPE with $\boldsymbol{\theta}$ known, that is into Equation [3], to give the ‘empirical’ MSPE, denoted EMSPE and more commonly referred to as the plug-in kriging variance estimator.

Kriging variance estimators

It is well known that the plug-in estimator of the kriging variance underpredicts the MSPE of the EBLUP, at least in general (Zimmerman and Cressie, 1992, den Hertog *et al.* 2006). More specifically, the MSPE with $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ known is given in Equation [3] and can be expressed succinctly as

$$m_1(\boldsymbol{\theta}) = E[\{\hat{Z}(x_0; \boldsymbol{\theta}) - Z(x_0)\}^2].$$

Thus, for $\boldsymbol{\theta}$ unknown, the EMSPE, that is the plug-in kriging variance, is given by $m_1(\hat{\boldsymbol{\theta}})$. Strictly, however, the mean square prediction error associated with the EBLUP is given by

$$\begin{aligned} m_2(\boldsymbol{\theta}) &= E[\{\hat{Z}(x_0; \hat{\boldsymbol{\theta}}) - Z(x_0)\}^2] = \\ &m_1(\boldsymbol{\theta}) + Var[\hat{Z}(x_0; \hat{\boldsymbol{\theta}}) - \hat{Z}(x_0; \boldsymbol{\theta})] \end{aligned}$$

where the trailing term in the second equality is algebraically intractable and therefore $m_2(\hat{\boldsymbol{\theta}})$ cannot be evaluated. A number of approaches, both algebraic and computational, to the construction of estimators of the EBLUP kriging variance, which are approximate and which to some extent redress the negative bias in the EMSPE, have been reported. In this study the properties of four such estimators, namely the Kacker-Harville estimator, the Prasad-Rao estimator, and two bootstrap estimators, are investigated. These estimators are summarized as follows.

Kacker and Harville (1984) used a Taylor series expansion of the trailing term in $m_2(\boldsymbol{\theta})$ to derive an estimator of the EBLUP kriging variance, albeit approximate, as

$$m_{2(KH)}(\boldsymbol{\theta}) \approx m_1(\boldsymbol{\theta}) + tr[\mathbf{A}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})] \quad [4]$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = Var\left[\frac{\partial \hat{Z}(x_0; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \text{ and } \mathbf{B}(\boldsymbol{\theta}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T].$$

In practice $\hat{\boldsymbol{\theta}}$ can be plugged into the terms $m_1(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$ of Equation [4] and the term $\mathbf{B}(\boldsymbol{\theta})$, which is the mean square error of $\hat{\boldsymbol{\theta}}$ and is unknown, can be approximated by the inverse of the observed Fisher information matrix for $\boldsymbol{\theta}$, that is the inverse of the Hessian $\mathbf{H}(\hat{\boldsymbol{\theta}})$. Thus the Kacker-Harville estimator can be evaluated as

$$m_{2(KH)}(\hat{\boldsymbol{\theta}}) = m_1(\hat{\boldsymbol{\theta}}) + tr[\mathbf{A}(\hat{\boldsymbol{\theta}})\mathbf{H}(\hat{\boldsymbol{\theta}})^{-1}]$$

Investigating ‘optimal’ kriging variance estimation using an analytic and a bootstrap approach

but, since this is essentially a plug-in estimator, it still remains negatively biased. To redress this bias, Prasad and Rao (1990) invoked some intricate theory to develop a further correction to $m_2(\theta)$ and, based on their results, introduced an estimator which approximates the MSE of the EBLUP as

$$m_{2(PR)}(\hat{\theta}) = m_1(\hat{\theta}) + 2tr[A(\hat{\theta})H(\hat{\theta})^{-1}].$$

Both the Kacker-Harville and the Prasad-Rao estimators reduce the negative bias of the plug-in kriging variance, but the improvement depends on sample size, spatial autocorrelation, Gaussian assumptions, an unbiased estimate of θ , and on the assumption that the covariance function is linear in the elements of θ (Prasad and Rao, 1990; Harville and Jeske, 1992; Zimmerman and Cressie, 1992). However, the latter assumption is clearly not valid for a range of widely used covariance models, such as the Gaussian, the exponential, the spherical, and the Matérn. This observation prompted Wang and Wall (2003) to suggest bootstrapping as an alternative method for estimating the MSPE of the EBLUP.

Two bootstrap estimators of $m_2(\theta)$ are investigated in the present study. Specifically, suppose that observations $Z(x)$ at the n locations, x_1, \dots, x_n , are taken from a Gaussian random field and that predictors at m new or test locations, x_{01}, \dots, x_{0m} , are required. The MLE or REML estimators of β and θ , denoted $\hat{\beta}$ and $\hat{\theta}$ respectively, can immediately be obtained from the data and, following den Hertog et al. (2006), bootstrap procedures based on the estimated distribution $Z(x) \sim G(F\hat{\beta}, \Sigma(\hat{\theta}))$ can then be implemented.

The bootstrap estimator of the MSPE of the EBLUP based on the assumption of a joint or unconditional Gaussian distribution for the responses at the observed and the test locations, termed the unconditional bootstrap estimator (UBE) and denoted $m_{2(UBE)}(\hat{\theta})$, is introduced here and can be obtained by implementing the following algorithm.

Unconditional bootstrap estimation

1. Sample from $G(F\hat{\beta}, \Sigma(\hat{\theta}))$ at the observed locations x_1, \dots, x_n and at the test locations x_{01}, \dots, x_{0m} simultaneously
2. Use the bootstrap sample at the n observed locations to fit a kriging model and use this fitted model to compute predictions at the m test locations x_{01}, \dots, x_{0m}
3. Calculate the squared prediction error at each test point
4. Repeat steps 1 to 3 B times. Then compute the unconditional bootstrap estimate of the MSPE, that is $m_{2(UBE)}(\hat{\theta})$, at each test point. Stop.

An alternative bootstrap estimator to the UBE, termed the conditional bootstrap estimator (CBE) and denoted $m_{2(CBE)}(\hat{\theta})$, is based on the fact that the conditional distribution of the responses $Z(x_0)$ at the test locations, x_{01}, \dots, x_{0m} , given observations $Z(x)$ at the n locations, x_1, \dots, x_n under a Gaussian assumption is well known. Specifically,

$$Z(x_0)|Z(x) \sim G\left(\hat{\mu}_0 + \Sigma_{01}(\hat{\theta})\Sigma_{11}^{-1}(\hat{\theta})(Z(x) - \hat{\mu}), \Sigma_{00}(\hat{\theta}) - \Sigma_{01}(\hat{\theta})\Sigma_{11}^{-1}(\hat{\theta})\Sigma_{10}(\hat{\theta})\right)$$

where the vectors $\hat{\mu} = F\hat{\beta}$ and $\hat{\mu}_0 = F_0\hat{\beta}$, with $\hat{\beta}$ evaluated at $\hat{\theta}$, are the vectors of estimated means at the n observed and the m test locations respectively and the variance matrix for $x = (x_1, \dots, x_n)$ and x_{01}, \dots, x_{0m} , partitioned conformably with these vectors, is given by

$$\Sigma(\hat{\theta}; x, x_0) = \begin{bmatrix} \Sigma_{11}(\hat{\theta}) & \Sigma_{10}(\hat{\theta}) \\ \Sigma_{01}(\hat{\theta}) & \Sigma_{00}(\hat{\theta}) \end{bmatrix}.$$

The conditional bootstrap estimator is also introduced in the present study and can be obtained by implementing the following algorithm.

Conditional bootstrap estimation

1. Use the kriging model, $G(F\hat{\beta}, \Sigma(\hat{\theta}))$ fitted to the observations $Z(x)$ to make predictions at the m test locations x_{01}, \dots, x_{0m}
2. Take a bootstrap sample at the test locations x_{01}, \dots, x_{0m} conditional on the observations $Z(x)$, that is, sample from the conditional distribution $Z(x_0)|Z(x)$
3. Calculate the squared differences of the predictions and the bootstrap realizations obtained from steps 1 and 2 respectively
4. Repeat steps 2 and 3 B times. Then compute the conditional bootstrap estimate of the MSPE, that is, $m_{2(CBE)}(\hat{\theta})$, at each test point. Stop.

Simulation study

In order to investigate the performance of the five estimators for the MSPE of the EBLUP introduced in this paper, namely the plug-in, the Kacker-Harville, the Prasad-Rao, the unconditional bootstrap, and the conditional bootstrap estimators, a simulation study incorporating conditions of spatial correlation specified by three Gaussian random fields was undertaken (Ngwenya, 2009).

Settings for the simulations are summarized as follows. The observed locations were taken as a regular grid of 112 points in a square domain $D = [0,15] \times [0,15]$ and are specified by the set

$$S = \{x = (u, v) : u = 2k, v = 2l + 1 \text{ and } u = 2l + 1, v = 2k \text{ for } k = 1, \dots, 7, l = 0, 1, \dots, 7\}.$$

In addition, a set of five points, $\{x_{01}, \dots, x_{05}\} = \{(3,1), (13,0), (8,8), (3,13), (15,15)\}$ within the domain D , was chosen to specify the prediction locations (Figure 1). The five prediction points were located so as to enable the effective study of location and data configuration on the various $m_2(\theta)$

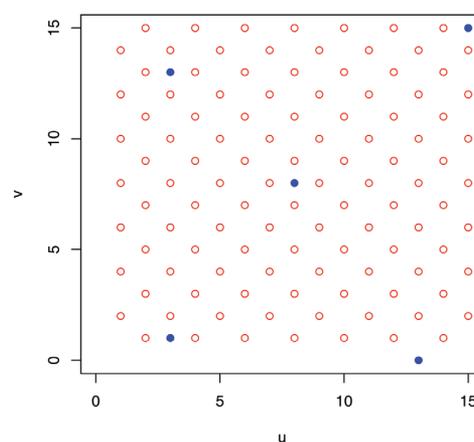


Figure 1—Spatial configuration of observed 112 points in domain D and 5 prediction locations. Observation locations are indicated by open circles and prediction locations by solid circles

Investigating ‘optimal’ kriging variance estimation using an analytic and a bootstrap approach

estimators. We have chosen a point in the middle of the grid, one inside the grid but close to the edge of the domain D , two on the edge of the grid, and one outside the grid but still in D .

An ordinary kriging model with $F(\mathbf{x}) = \mathbf{1}$ and a Gaussian distribution for the error terms, namely

$$\mathbf{Z}(\mathbf{x}) = \mathbf{1}\beta + \mathbf{e}(\mathbf{x}) \quad \text{with } \mathbf{e}(\mathbf{x}) \sim G(\mathbf{0}, \Sigma(\boldsymbol{\theta})),$$

was adopted. More specifically, the regression parameter β was taken to be zero, thus defining a simple kriging model, and an exponential covariance structure

$$C(\mathbf{x}_i - \mathbf{x}_j; \boldsymbol{\theta}) = \theta_1 + \theta_2(1 - e^{-3h/\theta_3})$$

where h denotes the Euclidean distance between the locations \mathbf{x}_i and \mathbf{x}_j , θ_1 the nugget, θ_2 the partial sill, and θ_3 the range, was adopted. Fixed values of $\theta_1 = 0.25$ and $\theta_2 = 1$ were taken and three values for the range, namely $\theta_3 = 1.5, 2.5,$ and 4.2 corresponding to weak, moderate, and strong spatial correlations respectively, were introduced. Thus three data-sets or random fields, denoted $\mathbf{Z}_1, \mathbf{Z}_2$ and \mathbf{Z}_3 , with increasing spatial correlation were created.

It is sensible to compare each of the five estimators of the MSPE of the EBLUP introduced in this study with the true value, $m_2(\boldsymbol{\theta})$. As noted earlier, no analytical expression for $m_2(\boldsymbol{\theta})$ is available. However Monte Carlo simulation to approximate the true value $m_2(\boldsymbol{\theta})$ to a chosen degree of accuracy can readily be performed. The procedure is summarized as follows.

Monte Carlo procedure to approximate $m_2(\boldsymbol{\theta})$

1. Set $r = 1$
2. Generate a spatial process over the grid of observation and prediction locations using the true values of the parameters β and $\boldsymbol{\theta}$
3. For the r th realization fit a kriging model by using the data at the observation locations to obtain estimates $\hat{\beta}_r$ of β and $\hat{\boldsymbol{\theta}}_r$ of $\boldsymbol{\theta}$
4. Calculate $[\hat{Z}(\mathbf{x}_0; \hat{\boldsymbol{\theta}}_r) - Z_r(\mathbf{x}_0)]^2$, the squared difference between the predicted value $\hat{Z}(\mathbf{x}_0; \hat{\boldsymbol{\theta}}_r)$ and the observed value $Z_r(\mathbf{x}_0)$ at each of the prediction locations, \mathbf{x}_0
5. Repeat steps 2 to 4, updating r to $r + 1$, R times and then compute the average squared differences between the predicted and simulated values as

$$\frac{1}{R} \sum_{r=1}^R [\{\hat{Z}(\mathbf{x}_0; \hat{\boldsymbol{\theta}}_r) - Z_r(\mathbf{x}_0)\}^2]$$

for each of the prediction locations, \mathbf{x}_0 . Stop.

A pilot study to determine the number of simulations R that gives $m_2(\boldsymbol{\theta})$ to an accuracy of at least ± 0.0001 for each of the three random fields, $\mathbf{Z}_1, \mathbf{Z}_2$, and \mathbf{Z}_3 , was performed. A maximum value of $R = 250\,000$ for the full Monte Carlo simulations was found and used in all cases.

The plug-in, the Kacker-Harville, and the Prasad-Rao estimators of the MSPE of the EBLUP at the $m = 5$ prediction locations for each of the random fields $\mathbf{Z}_1, \mathbf{Z}_2$, and \mathbf{Z}_3 , were based on 1000 realizations at the observation locations. Specifically, for each realization the MLE and REML estimates of $\boldsymbol{\theta}$, denoted generically as $\hat{\boldsymbol{\theta}}$, were found and plugged into the appropriate expressions. The resultant estimates were then averaged over the 1000 realizations to give the

estimators $m_1(\hat{\boldsymbol{\theta}})$, $m_{2(KH)}(\hat{\boldsymbol{\theta}})$ and $m_{2(PR)}(\hat{\boldsymbol{\theta}})$ to a degree of accuracy determined by the number of simulations. The bootstrap estimators were based on $B = 1000$ bootstrap iterations for each of the 1000 realizations of the observed and, in the case of the unconditional bootstrap estimator, the observed and the prediction locations. The unconditional bootstrap method is computationally expensive since $1000 \times 1000 \times 2$ (MLE and REML) = 200 000 kriging models had to be fitted to the simulated data. This consideration limited the number of realizations.

Results

The values the MSPE of the BLUP, $m_1(\boldsymbol{\theta})$, the MSPE of the EBLUP, $m_2(\boldsymbol{\theta})$, and the EMSPE $m_1(\hat{\boldsymbol{\theta}})$ together with the values of the four alternative kriging variance estimators, namely $m_{2(KH)}(\hat{\boldsymbol{\theta}})$, $m_{2(PR)}(\hat{\boldsymbol{\theta}})$, $m_{2(UBE)}(\hat{\boldsymbol{\theta}})$ and $m_{2(CBE)}(\hat{\boldsymbol{\theta}})$, under both MLE and REML estimation of the covariance parameters, were computed at each prediction location across the three random fields, $\mathbf{Z}_1, \mathbf{Z}_2$, and \mathbf{Z}_3 . The R programming language (R Development Core Team, 2012) was used to calculate these values and the results are tabulated in Ngwenya (2009). For conciseness, only the relative biases in the estimators of the MSPE of the EBLUP are presented here, with relative bias defined to be

$$\frac{E\{m_{2,E}(\hat{\boldsymbol{\theta}})\} - m_2(\boldsymbol{\theta})}{m_2(\boldsymbol{\theta})}$$

where $m_{2,E}(\hat{\boldsymbol{\theta}})$, is a generic estimator of $m_2(\boldsymbol{\theta})$. Specifically, the relative biases under MLE and REML estimation of the covariance parameters $\boldsymbol{\theta}$ are summarized in Table 1(a) and 1(b) together with $m_2(\boldsymbol{\theta})$ values obtained by the Monte Carlo procedure. In addition, the absolute values of the percentage relative biases are represented diagrammatically in Figures 2(a) and 2(b) respectively.

Certain features of the relative biases in the MSPEs of the EBLUP are worth noting. In particular, the performance of the conditional bootstrap estimator $m_{2(CBE)}(\hat{\boldsymbol{\theta}})$ is almost indistinguishable from that of the plug-in estimator $m_1(\hat{\boldsymbol{\theta}})$ under both MLE and REML estimation of $\boldsymbol{\theta}$ and at each point across the three random fields. These estimators seriously underestimate $m_2(\boldsymbol{\theta})$, with the degree of underestimation increasing with an increase in the spatial correlation exhibited by the random fields.

The Kacker-Harville and unconditional bootstrap estimators, $m_{2(KH)}(\hat{\boldsymbol{\theta}})$ and $m_{2(UBE)}(\hat{\boldsymbol{\theta}})$ respectively, also exhibit very similar behaviour, but only when the spatial correlation is moderate to strong. It is interesting to note here that Zimmerman (2006), in simulation studies on optimal network designs, found that design criteria based on $m_{2(KH)}(\hat{\boldsymbol{\theta}})$ and $m_{2(UBE)}(\hat{\boldsymbol{\theta}})$ lead to similar designs. However, when the spatial correlation is weak, $m_{2(KH)}(\hat{\boldsymbol{\theta}}_{MLE})$ has a propensity to lead to overestimates of $m_2(\boldsymbol{\theta})$, while $m_{2(UBE)}(\hat{\boldsymbol{\theta}}_{MLE})$ exhibits a high degree of accuracy, with an absolute relative bias as small as 0.3% at the prediction location (15, 15). In contrast, the estimators $m_{2(KH)}(\hat{\boldsymbol{\theta}}_{REML})$ and $m_{2(UBE)}(\hat{\boldsymbol{\theta}}_{REML})$ are more evenly matched when the spatial correlation is weak.

Values of $m_2(\boldsymbol{\theta})$, at the points (3,1), (15,15), and (13,0), which lie at an edge, a corner, and outside of the area circum-

Investigating ‘optimal’ kriging variance estimation using an analytic and a bootstrap approach

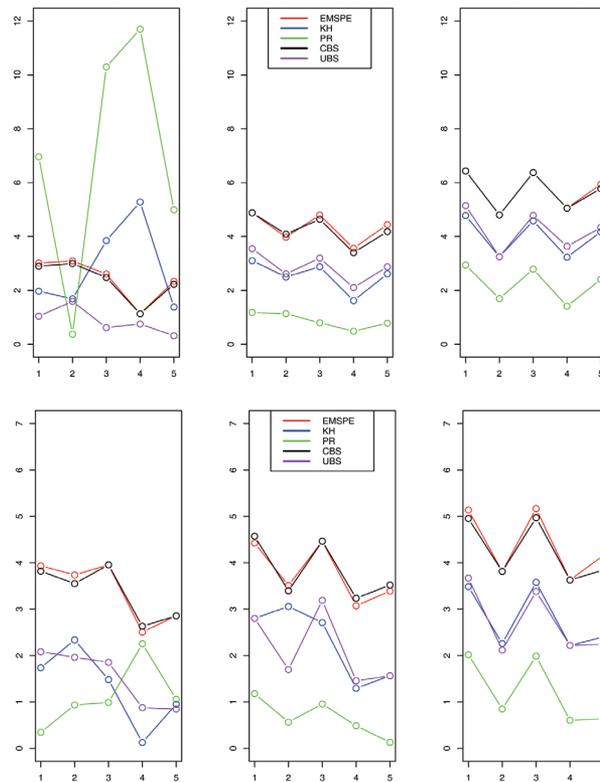


Figure 2—Graphical comparisons of the absolute relative bias of the five estimators of $m_2(\theta)$ at the five prediction locations in the three fields, Z_1 (left panel), Z_2 (middle panel), and Z_3 (right panel). All parameters for the estimators reported here were estimated via (a) MLE (top row) and (b) REML (bottom row)

scribed by the grid of observed locations, are greater than those for the prediction points (8,8) and (3,13), which are surrounded by observed locations. However it is also clear from the results summarized in Table I that this feature of the values of $m_2(\theta)$ for the prediction points is not mirrored in the corresponding values of the absolute relative biases. Indeed, there seems to be no obvious relationship between these measures of bias and the location of the prediction points and the spatial configuration of the grid, in accord with the findings of Zimmerman and Cressie (1992).

The Prasad-Rao estimator $m_{2(PR)}(\hat{\theta})$ does not seem to exhibit behaviour that is very close to any of the other four estimators of the MSPE of the EBLUP investigated in this study. Specifically, when the spatial correlation is weak, the estimator exhibits a large positive bias of up to 11.7% with $\hat{\theta}_{MLE}$ and 2.3% with $\hat{\theta}_{REML}$. This observation can be explained by the fact that, for weak spatial correlation, the large sample covariance matrix overestimates the variability in the parameter estimates $\hat{\theta}$, thereby inflating the trailing term in the expression for $m_{2(PR)}(\hat{\theta})$ (Abt, 1999). The same explanation can be given for the fact that $m_{2(KH)}(\hat{\theta})$ has a propensity to be positively biased when the spatial correlation is weak. In contrast, for moderate and strong spatial correlation, the Prasad-Rao estimator $m_{2(PR)}(\hat{\theta})$ exhibits the smallest absolute relative bias at all prediction locations when compared to the other four estimators of $m_2(\theta)$ in accord with the findings of Zimmerman and Cressie (1992). In fact, the Prasad-Rao estimator performs well with both $\hat{\theta}_{MLE}$ and $\hat{\theta}_{REML}$ in these situations, with biases as small as 0.5% and 0.1% respectively, and is clearly to be preferred.

Conclusions

Some broad conclusions based on the simulation study are now drawn, but should be taken strictly within the context of that study. Thus the performance of all of the five estimators of the mean square prediction error examined here depend somewhat sensitively on the method of estimation, that is MLE or REML, and in most cases REML should be the desired choice. The plug-in estimator consistently underestimates the MSPE of the EBLUP, in accord with more general findings in the literature, and is clearly unsatisfactory. The Prasad-Rao estimator performs optimally for the random fields with moderate and strong spatial correlation but would seem to be erratic for the field with weak correlation. In the latter case the Kacker-Harville estimator is to be preferred overall. It is interesting to note that the absolute relative biases of the prediction points do not appear to be related in any meaningful way to the location of those points in the domain of the simulation experiment.

Empirical studies on the nature and performance of approximate estimators for the MSPE of the EBLUP based on simulation or real data can be construed as being endless. Clearly, therefore, there is scope for further theoretical research into constructing reliable and robust estimators of this MSPE. One such approach could involve third-, fourth-, and higher-order approximations of appropriate Taylor series expansions. However, such derivations are intricate and nontrivial and indeed are fraught with subtle problems.

Acknowledgments

The authors would like to thank the reviewers for their

Investigating ‘optimal’ kriging variance estimation using an analytic and a bootstrap approach

Table 1

Approximate relative bias of the five kriging variance estimators at each point under (a) MLE and (b) REML of the covariance parameters across the Gaussian random fields Z_1 , Z_2 , and Z_3 used in the simulation study. For each field the estimator with the smallest relative bias at each point is indicated in bold. Also shown are the values of $m_2(\theta)$ obtained by the Monte Carlo procedure

I(a)	Random field	x_0	$m_2(\theta)$	$\hat{\theta}_{MLE}$				
				$m_1(\hat{\theta})$	$\sigma^2(\hat{\theta})_{KH}$	$\sigma^2(\hat{\theta})_{PR}$	$\sigma^2(\hat{\theta})_{UBS}$	$\sigma^2(\hat{\theta})_{CBS}$
	Z_1 weak	(3,1)	0.862	-0.030	0.020	0.070	-0.010	-0.029
		(13,0)	1.067	-0.031	-0.017	0.004	-0.016	-0.030
		(8,8)	0.806	-0.026	0.038	0.103	-0.006	-0.025
		(3,13)	0.795	-0.011	0.053	0.117	0.008	-0.011
		(15,15)	0.941	-0.023	0.014	0.003	-0.022	-0.022
	Z_2 moderate	(3,1)	0.676	-0.049	-0.031	-0.012	-0.036	-0.049
		(13,0)	0.880	-0.040	-0.025	-0.011	-0.026	-0.041
		(8,8)	0.625	-0.048	-0.029	-0.008	-0.032	-0.046
		(3,13)	0.617	-0.036	-0.016	0.005	-0.021	-0.034
		(15,15)	0.765	-0.044	-0.026	0.008	-0.029	-0.042
	Z_3 strong	(3,1)	0.544	-0.064	-0.048	-0.029	-0.051	-0.064
		(13,0)	0.708	-0.048	-0.032	-0.017	-0.032	-0.048
		(8,8)	0.502	-0.064	-0.046	-0.028	-0.048	-0.064
		(3,13)	0.495	-0.051	-0.032	-0.014	-0.036	-0.051
		(15,15)	0.623	-0.059	-0.042	-0.024	-0.043	-0.058
I(b)		x_0	$m_2(\theta)$	$\hat{\theta}_{REML}$				
				$m_1(\hat{\theta})$	$\sigma^2(\hat{\theta})_{KH}$	$\sigma^2(\hat{\theta})_{PR}$	$\sigma^2(\hat{\theta})_{UBS}$	$\sigma^2(\hat{\theta})_{CBS}$
	Z_1 weak	(3,1)	0.864	-0.039	-0.017	0.003	-0.021	-0.038
		(13,0)	1.070	-0.037	0.023	-0.009	-0.020	-0.036
		(8,8)	0.809	-0.040	0.015	0.010	-0.019	-0.040
		(3,13)	0.798	-0.025	0.001	0.023	-0.009	-0.026
		(15,15)	0.945	-0.029	-0.010	0.011	-0.008	-0.029
	Z_2 moderate	(3,1)	0.678	-0.044	-0.028	-0.012	-0.028	-0.046
		(13,0)	0.883	-0.035	-0.031	-0.006	-0.017	-0.034
		(8,8)	0.627	-0.045	-0.027	-0.010	-0.032	-0.045
		(3,13)	0.618	-0.031	-0.013	0.005	-0.015	-0.032
		(15,15)	0.767	-0.034	-0.016	0.001	-0.016	-0.035
	Z_3 strong	(3,1)	0.545	-0.051	-0.035	-0.020	-0.037	-0.050
		(13,0)	0.708	-0.038	-0.023	-0.008	-0.021	-0.038
		(8,8)	0.503	-0.052	-0.036	-0.020	-0.034	-0.050
		(3,13)	0.496	-0.036	-0.022	-0.006	-0.022	-0.036
		(15,15)	0.624	-0.042	-0.024	-0.006	-0.022	-0.038

helpful comments. They would also like to thank the University of Cape Town and the National Research Foundation (NRF) of South Africa, grants (UID) 85518 and (UID) 85456, for financial support. Any opinion, finding, and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept liability in this regard.

References

- ABT, M. 1999. Estimating the prediction mean squared error of a Gaussian stochastic process with exponential correlation structure. *Scandinavian Journal of Statistics*, vol. 26. pp. 563–578.
- CRESSIE, N. 1990. The origins of kriging. *Mathematical Geology*, vol. 22. pp. 239–252.
- DEN HERTOG, D., KLEIJNEN, J.P.C., and SIEM A.Y.D. 2006. The correct kriging variance estimated by bootstrapping. *Journal of the Operational Research Society*, vol. 57. pp. 400–409.
- DIGGLE, P.J. and RIBEIRO, P.J. 2007. *Model-based Geostatistics*. Springer, New York.
- HARVILLE, D.A. and JESKE, D.R. 1992. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, vol. 87. pp. 724–731.
- KACKER, R.N. and HARVILLE, D. 1984. Approximations for standard errors of estimators of fixed and random effect in mixed linear models. *Journal of*

- the American Statistical Association*, vol. 79. pp. 853–862.
- KRIGE, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, vol. 52. pp. 119–139.
- KRIGE, D.G. 1962. Effective pay limits for selective mining. *Journal of the South African Institute of Mining and Metallurgy*, vol. 62. pp. 345–363.
- MATHERON, G. 1963. Principles of geostatistics. *Economic Geology*, vol. 58. pp. 1246–1266.
- NGWENYA, M.Z. 2009. Investigating ‘Optimal’ Kriging Variance Estimation: Analytic and Bootstrap Estimators. Master’s thesis, University of Cape Town, South Africa.
- PRASAD, N.G.N. and RAO, J.N.K. 1990. On the estimation of the mean square error of small area predictors. *Journal of the American Statistical Association*, vol. 85. pp. 163–171.
- R DEVELOPEMENT CORE TEAM. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- WANG, F. and WALL, M.M. 2003. Incorporating parameter uncertainty into prediction intervals for spatial data modelled via parametric variogram. *Journal of Agricultural, Biological and Environmental Statistics*, vol. 8. pp. 296–309.
- ZIMMERMAN, D.L. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, vol. 17. pp. 635–652.
- ZIMMERMAN, D.L. and CRESSIE, N. 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, vol. 44. pp. 27–43. ◆