Ann. Inst. Statist. Math.
Vol. 44, No. 1, 27–43 (1992)

MEAN SQUARED PREDICTION ERROR IN THE SPATIAL LINEAR MODEL WITH ESTIMATED COVARIANCE PARAMETERS*

DALE L. ZIMMERMAN[1] AND NOEL CRESSIE[2]

[1] *Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, U.S.A.*
[2] *Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.*

(Received April 5, 1989; revised July 2, 1990)

Abstract. The problem considered is that of predicting the value of a linear functional of a random field when the parameter vector $\boldsymbol{\theta}$ of the covariance function (or generalized covariance function) is unknown. The customary predictor when $\boldsymbol{\theta}$ is unknown, which we call the EBLUP, is obtained by substituting an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in the expression for the best linear unbiased predictor (BLUP). Similarly, the customary estimator of the mean squared prediction error (MSPE) of the EBLUP is obtained by substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in the expression for the BLUP's MSPE; we call this the EMSPE. In this article, the appropriateness of the EMSPE as an estimator of the EBLUP's MSPE is examined, and alternative estimators of the EBLUP's MSPE for use when the EMSPE is inappropriate are suggested. Several illustrative examples show that the performance of the EMSPE depends on the strength of spatial correlation; the EMSPE is at its best when the spatial correlation is strong.

Key words and phrases: Best linear unbiased prediction, generalized covariances, geostatistics, kriging, spatial models.

1. Introduction

Many data sets in a variety of applied sciences consist of observations $y_1, y_2, \dots, y_n$ taken at corresponding known locations (here assumed to be points) $\boldsymbol{t}_1, \boldsymbol{t}_2, \dots, \boldsymbol{t}_n$ in d-dimensional Euclidean space $\mathbb{R}^d$; usually $d = 2$ or $d = 3$. A frequently successful approach to the analysis of such spatial data is to act as though they were derived from a real-valued stochastic process $\mathcal{F} = \{Y_{\boldsymbol{t}} : \boldsymbol{t} \in D\}$, where $D \subset \mathbb{R}^d$. The process $\mathcal{F}$ is called a random field. In contrast to an approach that takes the observations to be independently and identically distributed, the random field approach allows for spatial dependence among the observations, i.e., dependence that is related to the locations (both actual and relative) of the observations. However, with this approach, the data are regarded as a portion of a

* This research was partially supported by a University of Iowa Old Gold Fellowship (Zimmerman) and by the NSF under grant DMS-8703083 (Cressie).

single realization of $\mathcal{F}$; thus, if inference is to be generally possible, it is necessary to make some (second-order) stationarity assumptions.

Common stationarity assumptions are that $\mathcal{F}$ is either (a) second-order stationary or (b) an intrinsic random function of order k (IRF-k) for some integer $k \geq 0$ (Matheron (1973)). In case (b), $E(Y_t) = \boldsymbol{x}_t'\boldsymbol{\beta}$, where $\boldsymbol{x}_t$ is a $q \times 1$ vector whose elements are mixed monomials of degree $\leq k$ in the coordinates of $\boldsymbol{t}$ and $\boldsymbol{\beta}$ is a $q \times 1$ parameter vector; in case (a) the same expression for $E(Y_t)$ holds but with $q = 1$ and $\boldsymbol{x}_t \equiv 1$. Such random fields are referred to as spatial linear models. Assumptions (a) or (b) imply that there exists a parametric function $G(\cdot\,;\boldsymbol{\theta})$ of a d-dimensional vector such that

$$\operatorname{cov}\Big(\sum_i \lambda_i Y_{\boldsymbol{s}_i}, \sum_j \nu_j Y_{\boldsymbol{u}_j}\Big) = \sum_i \sum_j \lambda_i \nu_j G(\boldsymbol{s}_i - \boldsymbol{u}_j;\boldsymbol{\theta}),$$

either for all $\{\lambda_i\}$ and $\{\nu_j\}$ [in case (a)] or for all $\{\lambda_i\}$ and $\{\nu_j\}$ satisfying $E(\sum_i \lambda_i Y_{\boldsymbol{s}_i}) = E(\sum_j \nu_j Y_{\boldsymbol{u}_j}) = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^q$ [in case (b)] (see, e.g., Matheron (1973)). The dimension q is equal either to one [in case (a)] or to the number of mixed monomials in d variables of degree $\leq k$ [in case (b)], and the function G is either the covariance function [in case (a)] or the generalized covariance function of order k [in case (b)]. In case (b), if $k = 0$ then $q = 1$ and $G(\boldsymbol{s} - \boldsymbol{u})$ is equal to the negative of the semivariogram $(1/2)\operatorname{var}(Y_{\boldsymbol{s}} - Y_{\boldsymbol{u}})$; an example of such a process is one-dimensional Brownian motion, for which $G(s-u) = -|s-u|$. An example of a one-dimensional intrinsic random function of higher order ($k \geq 1$) is the process $W_s \equiv \int_0^s (s-t)^{k-1} W_t dt/(k-1)!$, which has generalized covariance function (of order k) $G(s-u) = (-1)^{k-1}|s-u|^{2k+1}/(2k+1)!$ (Matheron (1973), Section 2.3).

In this paper, we consider the problem of predicting the value y_0 of a linear functional $l(\cdot)$ of a random field $\mathcal{F}$ satisfying either assumption (a) or (b). How spatial prediction should be accomplished in this framework depends on what one is willing to assume is known about G, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. We consider this problem under each of the following alternative states of knowledge:

State 1. Form of G is known, $\boldsymbol{\theta}$ is known, $\boldsymbol{\beta}$ is unknown.

State 2. Form of G is known, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are unknown, and $\boldsymbol{\theta}$ is restricted to a known set Θ.

Under State 1, in which case $\boldsymbol{\theta}$ is known, the best linear unbiased predictor (BLUP) of y_0 exists, and expressions for it and its mean squared prediction error (MSPE) are well known. However, in practice $\boldsymbol{\theta}$ is rarely known; consequently, of the two states considered, the most appropriate one is usually State 2. But how should one predict under State 2? The natural and, by now, classical approach to prediction under State 2 takes the expressions for the BLUP and its MSPE derived under State 1, and simply substitutes in an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$, yielding the EBLUP and EMSPE, respectively. The "E" added to "BLUP" and "MSPE" can be regarded as an abbreviation for either "empirical" (following Harville and Jeske (1992)) or "estimated."

The goal of this article is to give properties of the State-2 prediction approach just described, with emphasis on the appropriateness of the EMSPE as an estimator of the EBLUP's MSPE. In particular, we obtain an inequality that, under certain conditions, relates the BLUP's MSPE to the EBLUP's MSPE and to the expectation of EMSPE. Reinforcing that the purpose of this research is to obtain good estimators of the EBLUP's MSPE, we propose some alternative estimators. We illustrate our results with four examples. The examples suggest that when spatial correlation is weak, the EMSPE tends to underestimate the EBLUP's actual MSPE but the alternative estimators may be useful; the examples also suggest that the EMSPE can perform well in the presence of strong spatial correlation.

Throughout this article, we shall assume that the order k of the IRF has been correctly identified using, for example, the unbiased nonparametric procedure proposed by Cressie (1987). If the chosen k is too small, then there are monomial mixtures in $E(Y_t)$ of order larger than k, and the estimated variograms and generalized covariance functions tend to be inflated (e.g., Starks and Fang (1982)), which in turn make estimators of MSPE's larger.

The effect of estimated variance and covariance parameters on mean squared error of estimation or prediction has been considered previously in the context of time series models (Yamamoto (1976), Reinsel (1980), Fuller and Hasza (1981)), random and mixed linear models (Khatri and Shah (1981), Reinsel (1984), Kackar and Harville (1984)), the heteroscedastic regression model (Bement and Williams (1969), Carroll *et al.* (1988)), and the general linear model (Toyooka (1982), Rothenberg (1984), Eaton (1985), Harville (1985), Harville and Jeske (1992)). The present paper follows closely the development of Harville (1985) and Harville and Jeske (1992), and much of it may be viewed as an extension of their results to models with generalized covariances.

2. The State-1 predictor and its MSPE

Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and $\boldsymbol{y}^* = (y_0, \boldsymbol{y}')'$, let $\boldsymbol{K}$ denote the $n \times n$ matrix whose ij-th element is $G(\boldsymbol{t}_i - \boldsymbol{t}_j; \boldsymbol{\theta})$, let $\boldsymbol{k}$ denote the $n \times 1$ vector whose i-th element is $\int G(\boldsymbol{t}_i - \boldsymbol{v}; \boldsymbol{\theta}) l(d\boldsymbol{v})$, and let $k_0 = \int\int G(\boldsymbol{v} - \boldsymbol{w}; \boldsymbol{\theta}) l(d\boldsymbol{v}) l(d\boldsymbol{w})$. Here, $l(d\boldsymbol{v})$ is the measure corresponding to the linear functional $l(\mathcal{F})$ to be predicted. Define

$$\boldsymbol{\Sigma} = \begin{bmatrix} k_0 & \boldsymbol{k}' \\ \boldsymbol{k} & \boldsymbol{K} \end{bmatrix}.$$

We assume that the set Θ to which the true value $\boldsymbol{\theta}$ is restricted is that set for which $\boldsymbol{\Sigma}$ is conditionally positive definite, i.e., the set for which $\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} > 0$ for all $\boldsymbol{\lambda}$ satisfying $E(\boldsymbol{\lambda}'\boldsymbol{y}^*) = 0$. To emphasize the dependence of $\boldsymbol{K}$, $\boldsymbol{k}$, k_0 and $\boldsymbol{\Sigma}$ on $\boldsymbol{\theta}$ (or on an arbitrary element $\omega \in \Theta$) we shall at times write these quantities as $\boldsymbol{K}(\boldsymbol{\theta})$, $\boldsymbol{k}(\boldsymbol{\theta})$, $k_0(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ [or as $\boldsymbol{K}(\omega)$, $\boldsymbol{k}(\omega)$, $k_0(\omega)$ and $\boldsymbol{\Sigma}(\omega)$], respectively. It follows from the assumptions on $\mathcal{F}$ that there exists a known $q \times 1$ vector $\boldsymbol{x}_0$ and a known $n \times q$ matrix $\boldsymbol{X}$ whose rows are $\boldsymbol{x}'_{t_1}, \ldots, \boldsymbol{x}'_{t_n}$, such that $E(y_0) = \boldsymbol{x}'_0\boldsymbol{\beta}$ and $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$. Furthermore, $\mathrm{var}(\boldsymbol{\Lambda}'\boldsymbol{y}^*) = \boldsymbol{\Lambda}'\boldsymbol{\Sigma}\boldsymbol{\Lambda}$ where, if $\mathcal{F}$ is second-order stationary [case (a)], then $\boldsymbol{\Lambda}$ is any matrix having $n+1$ columns, or if $\mathcal{F}$ is an IRF-k [case (b)], then $\boldsymbol{\Lambda}$ is any matrix satisfying $\boldsymbol{\Lambda}'\boldsymbol{x}'_0 = \boldsymbol{0}'$ and $\boldsymbol{\Lambda}'\boldsymbol{X} = \boldsymbol{0}$.

Define $\boldsymbol{z} = \boldsymbol{L}'\boldsymbol{y}$, where $\boldsymbol{L}$ is any $n \times (n-q)$ matrix of rank $n-q$ satisfying $\boldsymbol{L}'\boldsymbol{X} = \boldsymbol{0}$. The vector $\boldsymbol{z}$ is a maximal invariant with respect to transformations of the general form $T(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{X}\boldsymbol{w}$ and is of great importance in what follows.

If $\mathcal{F}$ is an IRF-k, $\boldsymbol{z}$ can be taken to be any vector of $n-q$ linearly independent generalized increments of order k (GI-k's) of $\boldsymbol{y}$. (A GI-k is defined as a linear combination $\boldsymbol{\lambda}'\boldsymbol{y}$ for which $E(\boldsymbol{\lambda}'\boldsymbol{y}) = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^q$; see, e.g., Delfiner (1976).)

It is well known (Goldberger (1962), Delfiner (1976)) that under State 1, the BLUP of y_0, i.e., the linear predictor $\boldsymbol{b}'\boldsymbol{y}$ that minimizes $\mathrm{var}(\boldsymbol{b}'\boldsymbol{y} - y_0)$ subject to $E(\boldsymbol{b}'\boldsymbol{y}) = E(y_0)$ for all $\boldsymbol{\beta} \in \mathbb{R}^q$, is $\boldsymbol{a}'\boldsymbol{y}$, where $\boldsymbol{a}$ is the first component of a solution to the equations

$$\begin{bmatrix} \boldsymbol{K} & \boldsymbol{X} \\ \boldsymbol{X}' & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{k} \\ \boldsymbol{x}_0 \end{bmatrix}. \tag{2.1}$$

Here, $\boldsymbol{\gamma}$ is a vector of Lagrange multipliers that enforces the unbiasedness condition. Assuming that $\boldsymbol{K}$ is nonsingular, the following results are easily verified:

(a) the first "component" of the unique solution to (2.1) is

$$\boldsymbol{a} = \boldsymbol{K}^{-1}\boldsymbol{k} + \boldsymbol{K}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{k}),$$

(b) the prediction error $\boldsymbol{a}'\boldsymbol{y} - y_0$ is a GI-k,

(c) $E[(\boldsymbol{a}'\boldsymbol{y} - y_0)^2] = \mathrm{var}(\boldsymbol{a}'\boldsymbol{y} - y_0) = k_0 - \boldsymbol{k}'\boldsymbol{K}^{-1}\boldsymbol{k} + (\boldsymbol{x}_0' - \boldsymbol{k}'\boldsymbol{K}^{-1}\boldsymbol{X}) \cdot (\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{k})$.

Thus, under State 1, the BLUP of y_0 is

$$\begin{aligned} p_1(\boldsymbol{y};\boldsymbol{\theta}) = {} & \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{y} \\ & + \boldsymbol{k}'\boldsymbol{K}^{-1}[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{K}^{-1}]\boldsymbol{y} \end{aligned} \tag{2.2}$$

and the MSPE $E\{[p_1(\boldsymbol{y};\boldsymbol{\theta}) - y_0]^2\}$ of $p_1(\boldsymbol{y};\boldsymbol{\theta})$ is

$$m_1(\boldsymbol{\theta}) = k_0 - \boldsymbol{k}'\boldsymbol{K}^{-1}\boldsymbol{k} + (\boldsymbol{x}_0' - \boldsymbol{k}'\boldsymbol{K}^{-1}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{k}). \tag{2.3}$$

3. The State-2 EBLUP and its MSPE

Since the quantities $p_1(\boldsymbol{y};\boldsymbol{\theta})$ and $m_1(\boldsymbol{\theta})$ may be functionally dependent on $\boldsymbol{\theta}$, under State 2 they generally are not statistics that depend only on the data. Consequently, we must look elsewhere for a predictor. It is customary to obtain a predictor and an estimator of its MSPE by substituting an estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ for $\boldsymbol{\theta}$ in the expressions for $p_1(\boldsymbol{y};\boldsymbol{\theta})$ and $m_1(\boldsymbol{\theta})$. This yields the generally nonlinear predictor $p_2(\boldsymbol{y}) = p_1(\boldsymbol{y};\hat{\boldsymbol{\theta}}(\boldsymbol{y}))$, which we call the EBLUP. Thus, the prediction error and MSPE under State 2 are $p_2(\boldsymbol{y}) - y_0$ and $m_2(\boldsymbol{\theta}) = E\{[p_2(\boldsymbol{y}) - y_0]^2\}$, respectively. Substitution of $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ for $\boldsymbol{\theta}$ in expression (2.3) yields the customary estimator $m_1(\hat{\boldsymbol{\theta}})$ of $m_2(\boldsymbol{\theta})$, which we call the EMSPE.

In the remainder of this section, we examine the relationship between $m_1(\boldsymbol{\theta})$ and $m_2(\boldsymbol{\theta})$ and the extent to which $m_1(\hat{\boldsymbol{\theta}})$ is a good estimator of $m_2(\boldsymbol{\theta})$. Our results depend on making one or more assumptions about the distribution of $\boldsymbol{y}^*$, about $\hat{\boldsymbol{\theta}}$, and about $G(\cdot\,;\boldsymbol{\theta})$.

One assumption that shall be made throughout, without further mention, is that $\hat{\boldsymbol{\theta}}$ is an even and translation invariant estimator, i.e., $\hat{\boldsymbol{\theta}}(-\boldsymbol{y}) = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ for every

$\boldsymbol{y}$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{y}+\boldsymbol{X}\boldsymbol{w}) = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ for every vector $\boldsymbol{w}$ (and every $\boldsymbol{y}$). This assumption is satisfied by all proposed estimators of covariance and generalized covariance parameters in spatial models known to the authors, including various ordinary, weighted, and generalized least squares estimators (Delfiner (1976), Cressie (1985)), maximum likelihood and restricted maximum likelihood estimators (Kitanidis (1983), Mardia and Marshall (1984)), and minimum variance and minimum norm quadratic unbiased estimators (Kitanidis (1985), Marshall and Mardia (1985)). The translation invariance of $\hat{\boldsymbol{\theta}}$ implies that $\hat{\boldsymbol{\theta}}$ depends on $\boldsymbol{y}$ only through the value of the maximal invariant $\boldsymbol{z} = \boldsymbol{L}'\boldsymbol{y}$, in which case we can alternatively represent $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ as $\hat{\boldsymbol{\theta}}(\boldsymbol{z})$.

The remaining assumptions shall be made from among the following set:

(A1) The distribution of $\boldsymbol{y}^* = (y_0, \boldsymbol{y}')'$ is symmetric about its mean.

(A2) The distribution of $\boldsymbol{y}^*$ is multivariate normal.

(B1) $\hat{\boldsymbol{\theta}}$ is unbiased.

(B2) $\hat{\boldsymbol{\theta}}(\boldsymbol{z})$ is a complete sufficient statistic for the distribution of $\boldsymbol{z}$.

(C1) $G(\cdot\,;\boldsymbol{\theta})$ is a linear function of the elements of $\boldsymbol{\theta}$.

Assumptions (A1) and (A2) (the latter of which is nested within the former) can often be satisfied (to a good approximation) by making an appropriate transformation of $\boldsymbol{y}^*$. Assumptions (B1) and (B2) are rather stringent; (B1) is satisfied in some cases by several of the aforementioned estimators of $\boldsymbol{\theta}$ but not by others, whereas (B2) is satisfied only in very special cases. Assumption (C1) is satisfied by several covariance and generalized covariance functions which have proven to be useful in practice, including the class of polynomial generalized covariance functions introduced by Delfiner (1976), the interpolating splines (Dubrule (1983)), and covariance functions with known correlation structure up to unknown additive (nugget effect) and multiplicative (partial sill) parameters.

3.1 *Relationship between $m_1(\boldsymbol{\theta})$ and $m_2(\boldsymbol{\theta})$*

The State-2 prediction error can be decomposed into two components in accordance with the identity

$$p_2(\boldsymbol{y}) - y_0 = [p_1(\boldsymbol{y};\boldsymbol{\theta}) - y_0] + [p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})]. \tag{3.1}$$

The first component is merely the State-1 prediction error; hence, the second component represents the additional error directly attributable to the lack of knowledge of $\boldsymbol{\theta}$ under State 2. It follows, by noting the unbiasedness of $p_1(\boldsymbol{y};\boldsymbol{\theta})$ and by applying Wolfe's (1973) Theorems 2.1 and 2.2 [with $\mu = 0$, $g(\boldsymbol{z}) = -\boldsymbol{z}$, $U(\boldsymbol{z}) = p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})$ and $V(\boldsymbol{z}) = \hat{\boldsymbol{\theta}}(\boldsymbol{z})$], that $p_2(\boldsymbol{y})$ is unbiased under Assumption (A1) (assuming that $E[p_2(\boldsymbol{y})]$ exists). Interestingly, $\hat{\boldsymbol{\theta}}$ need not be an unbiased estimator of $\boldsymbol{\theta}$ for the predictor of y_0 to be unbiased.

Having established relatively weak conditions under which the EBLUP is unbiased, we now characterize the relationship of the MSPE of this predictor to the MSPE of the State-1 BLUP. It follows from decomposition (3.1) that

$$\begin{aligned} m_2(\boldsymbol{\theta}) = m_1(\boldsymbol{\theta}) &+ E\{[p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})]^2\} \\ &+ 2\,\mathrm{cov}[p_1(\boldsymbol{y};\boldsymbol{\theta}) - y_0, p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})]. \end{aligned}$$

Observe that if the two components in decomposition (3.1) are uncorrelated or, more generally, if $E\{[p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})]^2\} \geq -2\ \mathrm{cov}[p_1(\boldsymbol{y};\boldsymbol{\theta}) - y_0, p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})]$, then $m_2(\boldsymbol{\theta}) \geq m_1(\boldsymbol{\theta})$. We proceed to obtain sufficient conditions for the two components to be uncorrelated.

Define

$$\begin{aligned}
\boldsymbol{P} &= \boldsymbol{P}(\boldsymbol{\theta}) = \boldsymbol{X}[\boldsymbol{X}'\boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{K}^{-1}(\boldsymbol{\theta}),\\
\hat{\boldsymbol{P}} &= \boldsymbol{P}(\hat{\boldsymbol{\theta}}) = \boldsymbol{X}[\boldsymbol{X}'\boldsymbol{K}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{K}^{-1}(\hat{\boldsymbol{\theta}}),\\
\boldsymbol{F}' &= \boldsymbol{F}'(\boldsymbol{\theta}) = \boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{L}[\boldsymbol{L}'\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{L}]^{-1},\\
\hat{\boldsymbol{F}}' &= \boldsymbol{F}'(\hat{\boldsymbol{\theta}}) = \boldsymbol{K}(\hat{\boldsymbol{\theta}})\boldsymbol{L}[\boldsymbol{L}'\boldsymbol{K}(\hat{\boldsymbol{\theta}})\boldsymbol{L}]^{-1},
\end{aligned}$$

and let $\boldsymbol{u}$ represent any vector such that $\boldsymbol{x}_0' = \boldsymbol{u}'\boldsymbol{X}$. It is easy to show that $\boldsymbol{I} - \boldsymbol{P} = \boldsymbol{F}'\boldsymbol{L}'$, $\boldsymbol{I} - \hat{\boldsymbol{P}} = \hat{\boldsymbol{F}}'\boldsymbol{L}'$ and $\hat{\boldsymbol{P}}\boldsymbol{P} = \boldsymbol{P}$. Thus,

$$\begin{aligned}
&p_2(\boldsymbol{y}) - p_1(\boldsymbol{y};\boldsymbol{\theta})\\
&\quad = \boldsymbol{u}'\hat{\boldsymbol{P}}(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y} + \boldsymbol{k}'(\hat{\boldsymbol{\theta}})\boldsymbol{K}^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{I} - \hat{\boldsymbol{P}})\boldsymbol{y} - \boldsymbol{k}'(\boldsymbol{\theta})\boldsymbol{K}^{-1}(\boldsymbol{\theta})(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y}\\
&\quad = \boldsymbol{u}'\hat{\boldsymbol{P}}\boldsymbol{F}'\boldsymbol{z} + \boldsymbol{k}'(\hat{\boldsymbol{\theta}})\boldsymbol{K}^{-1}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{F}}'\boldsymbol{z} - \boldsymbol{k}'(\boldsymbol{\theta})\boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{F}'\boldsymbol{z},
\end{aligned}$$

i.e., the second component of decomposition (3.1) depends on $\boldsymbol{y}$ only through the value of $\boldsymbol{z}$. Moreover, it is easy to show that $\mathrm{cov}[\boldsymbol{z}, p_1(\boldsymbol{y};\boldsymbol{\theta}) - y_0] = 0$. It follows that Assumption (A2) is sufficient for the two components of decomposition (3.1) to be distributed independently and hence to be uncorrelated with each other. We have therefore established the following theorem.

THEOREM 3.1. *Suppose that Assumption* (A2) *is satisfied. Then* $m_2(\boldsymbol{\theta}) \geq m_1(\boldsymbol{\theta})$, *with equality holding if and only if* $p_2(\boldsymbol{y}) = p_1(\boldsymbol{y};\boldsymbol{\theta})$ *with probability one.*

Further, if $m_1(\boldsymbol{\theta}) \geq E[m_1(\hat{\boldsymbol{\theta}})]$, then Theorem 3.1 shows, to the extent that Assumption (A2) is satisfied, that the EMSPE $m_1(\hat{\boldsymbol{\theta}})$ will tend to result in overconfidence in the EBLUP's precision. Theorem 3.1 also suggests that for $E[m_1(\hat{\boldsymbol{\theta}})] - m_1(\boldsymbol{\theta})$ to be positive could be a virtue. Conditions under which $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_2(\boldsymbol{\theta})$ will be discussed further in the next section, and situations in which $m_1(\hat{\boldsymbol{\theta}})$ tends to overestimate both $m_1(\boldsymbol{\theta})$ and $m_2(\boldsymbol{\theta})$ will be discussed in the final section.

3.2 *Performance of* $m_1(\hat{\boldsymbol{\theta}})$ *in estimating* $m_2(\boldsymbol{\theta})$

Theorem 3.1 gives a sufficient condition for the MSPE of the EBLUP to exceed the BLUP's MSPE. While this result is of interest in itself, of more interest from a practical standpoint is the performance of the EMSPE $m_1(\hat{\boldsymbol{\theta}})$ in estimating $m_2(\boldsymbol{\theta})$, the EBLUP's MSPE. We now give three theorems that shed some light on this issue.

Let $\Omega = \{\boldsymbol{\Sigma}(\boldsymbol{\omega}) : \boldsymbol{\omega} \in \Theta\}$ and $\mathcal{U} = \{\boldsymbol{u} : \boldsymbol{x}_0' = \boldsymbol{u}'\boldsymbol{X}\}$. It follows from the definition of Θ that Ω is a convex set. Clearly, any linear unbiased State-1 predictor can be written as $\boldsymbol{u}'\boldsymbol{y}$ for some $\boldsymbol{u} \in \mathcal{U}$. Define, for $\boldsymbol{u} \in \mathcal{U}$, $\psi_{\boldsymbol{u}}(\boldsymbol{\Sigma}) = \mathrm{var}(\boldsymbol{u}'\boldsymbol{y} - y_0) = k_0 - 2\boldsymbol{u}'\boldsymbol{k} + \boldsymbol{u}'\boldsymbol{K}\boldsymbol{u}$; also define $\psi(\boldsymbol{\Sigma}) = \inf_{\boldsymbol{u}\in\mathcal{U}} \psi_{\boldsymbol{u}}(\boldsymbol{\Sigma})$. It is easily verified that $\psi_{\boldsymbol{u}}(\boldsymbol{\Sigma})$

is a linear (and hence concave) function on Ω; consequently, $\psi(\mathbf{\Sigma})$ is a concave function on the convex set Ω. Therefore, for any estimator $\hat{\boldsymbol{\theta}}$, $E[\psi(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}))] \leq \psi(E[\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})])$ by Jensen's inequality. Observing that $\psi(\mathbf{\Sigma}(\boldsymbol{\omega})) = m_1(\boldsymbol{\omega})$ for any $\boldsymbol{\omega} \in \Theta$, we obtain the following theorem, which is a coordinate-space version of a result due to Eaton (1985).

THEOREM 3.2. *Suppose that* $E[\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})] = \mathbf{\Sigma}(\boldsymbol{\theta})$. *Then* $E[m_1(\hat{\boldsymbol{\theta}})] \leq m_1(\boldsymbol{\theta})$.

COROLLARY 3.1. *Under Assumptions* (B1) *and* (C1), $E[m_1(\hat{\boldsymbol{\theta}})] \leq m_1(\boldsymbol{\theta})$.

An argument similar to that used to establish Theorem 3.2 can be used to establish the following theorem.

THEOREM 3.3. *Suppose that* $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ *is negatively biased in the sense that* $E[\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})] = \mathbf{\Sigma}(\boldsymbol{\theta}) - \mathbf{\Sigma}^*$, *where* $\mathbf{\Sigma}^*$ *is positive definite. Then* $E[m_1(\hat{\boldsymbol{\theta}})] \leq m_1(\boldsymbol{\theta}) - c$, *where* $c > 0$.

The implication of Theorems 3.1, 3.2 and 3.3 is that $m_1(\hat{\boldsymbol{\theta}})$ is likely to be an accurate estimator of $m_2(\boldsymbol{\theta})$ only when $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ overestimates $\mathbf{\Sigma}(\boldsymbol{\theta})$. This is stated another way in the following corollary.

COROLLARY 3.2. *Suppose that Assumption* (A2) *is satisfied, and suppose either that Assumptions* (B1) *and* (C1) *are satisfied or that* $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ *is negatively biased in the sense defined in Theorem* 3.3. *Then* $E[m_1(\hat{\boldsymbol{\theta}})] \leq m_1(\boldsymbol{\theta}) \leq m_2(\boldsymbol{\theta})$.

Although Corollary 3.2 gives conditions under which $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_2(\boldsymbol{\theta})$, it gives no indication as to which discrepancy, $m_1(\boldsymbol{\theta}) - E[m_1(\hat{\boldsymbol{\theta}})]$ or $m_2(\boldsymbol{\theta}) - m_1(\boldsymbol{\theta})$, if either, contributes more to the bias. Under certain additional conditions, it is possible to quantify exactly the amount by which $E[m_1(\hat{\boldsymbol{\theta}})]$ tends to underestimate $m_2(\boldsymbol{\theta})$ in terms of the amount by which $m_2(\boldsymbol{\theta})$ exceeds $m_1(\boldsymbol{\theta})$. The following theorem extends a remarkable result of Harville and Jeske (1992); we omit its proof since, apart from replacing the ordinary covariance matrix by the generalized covariance matrix $\boldsymbol{K}$, it can be proved in exactly the same fashion as Harville and Jeske's result.

THEOREM 3.4. *Suppose that Assumptions* (A2), (B2) *and* (C1) *are satisfied. Then*

(a) $E[m_1(\hat{\boldsymbol{\theta}})] = m_2(\boldsymbol{\theta}) - 2[m_2(\boldsymbol{\theta}) - m_1(\boldsymbol{\theta})]$, *or equivalently,*

(b) $m_2(\boldsymbol{\theta}) = 2m_1(\boldsymbol{\theta}) - E[m_1(\hat{\boldsymbol{\theta}})]$.

Part (a) of Theorem 3.4 indicates [under Assumptions (A2), (B2) and (C1)] that $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_2(\boldsymbol{\theta})$ by an amount equal to $2[m_2(\boldsymbol{\theta}) - m_1(\boldsymbol{\theta})]$; that is, the discrepancy between $m_1(\boldsymbol{\theta})$ and $E[m_1(\hat{\boldsymbol{\theta}})]$ and the discrepancy between $m_2(\boldsymbol{\theta})$ and $m_1(\boldsymbol{\theta})$ contribute *equally* to the bias of $m_1(\hat{\boldsymbol{\theta}})$. Part (b) indicates [again under Assumptions (A2), (B2) and (C1)] that an unbiased estimator of $m_2(\boldsymbol{\theta})$ is given by $\hat{m}_2 = 2\hat{m}_1 - m_1(\hat{\boldsymbol{\theta}})$, where $\hat{m}_1$ is any unbiased estimator of $m_1(\boldsymbol{\theta})$.

3.3 *Approximations for, and approximately unbiased estimators of,* $m_2(\boldsymbol{\theta})$

Theorem 3.4 allows one to determine $m_2(\boldsymbol{\theta})$ exactly, in terms of $E[m_1(\hat{\boldsymbol{\theta}})]$, and to estimate it unbiasedly when the assumptions of the theorem are satisfied. In many cases, however, one or more of the assumptions of Theorem 3.4 may not hold, and hence neither an exact expression for, nor an unbiased estimator of, $m_2(\boldsymbol{\theta})$ is available. In such cases it may be possible to approximate $m_2(\boldsymbol{\theta})$ and then use the approximation to obtain an estimator of $m_2(\boldsymbol{\theta})$ that is less biased than the EMSPE $m_1(\hat{\boldsymbol{\theta}})$. An approximation of $m_2(\boldsymbol{\theta})$ obtained by squaring the first-order approximation of the Taylor series expansion of $p_1(\boldsymbol{y};\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}$ was proposed for use in mixed linear models with estimated variance components by Kackar and Harville (1984) and again for use in the general linear model with estimated covariance parameters by Harville and Jeske (1992). The approximation can also be used in this setting, and is given by

$$m_2(\boldsymbol{\theta}) \doteq m^*(\boldsymbol{\theta}) \equiv m_1(\boldsymbol{\theta}) + \mathrm{tr}[\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{B}(\boldsymbol{\theta})], \tag{3.2}$$

where $\boldsymbol{A}(\boldsymbol{\theta}) = \mathrm{var}[\boldsymbol{d}(\boldsymbol{y};\boldsymbol{\theta})]$, $\boldsymbol{d}(\boldsymbol{y};\boldsymbol{\theta}) = \partial p_1(\boldsymbol{y};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, and $\boldsymbol{B}(\boldsymbol{\theta})$ is a matrix that either equals or approximates the mean squared error matrix $E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})']$. Obviously, the partial derivatives of $p_1(\boldsymbol{y};\boldsymbol{\theta})$ with respect to the elements of $\boldsymbol{\theta}$ must exist if this approximation is to be used, and $\hat{\boldsymbol{\theta}}$ must be an estimator for which a corresponding $\boldsymbol{B}(\boldsymbol{\theta})$ can be calculated. With regard to the latter point, the Gaussian-based maximum likelihood (ML) or restricted maximum likelihood (REML) estimator of $\boldsymbol{\theta}$ is a convenient choice, since the inverse of the information matrix associated with the corresponding likelihood function either contains (in the case of ML) or is equal to (in the case of REML) the large-sample covariance matrix of $\hat{\boldsymbol{\theta}}$.

An estimator of $m_2(\boldsymbol{\theta})$ based on approximation (3.2) is $m^*(\hat{\boldsymbol{\theta}}) = m_1(\hat{\boldsymbol{\theta}}) + \mathrm{tr}[\boldsymbol{A}(\hat{\boldsymbol{\theta}})\boldsymbol{B}(\hat{\boldsymbol{\theta}})]$. Clearly, this estimator is more conservative than $m_1(\hat{\boldsymbol{\theta}})$, but it is equally clear that even $m^*(\hat{\boldsymbol{\theta}})$ will tend to underestimate $m_2(\boldsymbol{\theta})$ in those cases where $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_1(\boldsymbol{\theta})$ and $E\{\mathrm{tr}[\boldsymbol{A}(\hat{\boldsymbol{\theta}})\boldsymbol{B}(\hat{\boldsymbol{\theta}})]\} \doteq \mathrm{tr}[\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{B}(\boldsymbol{\theta})]$. In view of this, Prasad and Rao (1986) and Harville and Jeske (1992) proposed the estimator $m^{**}(\hat{\boldsymbol{\theta}}) = m_1(\hat{\boldsymbol{\theta}}) + 2\,\mathrm{tr}[\boldsymbol{A}(\hat{\boldsymbol{\theta}})\boldsymbol{B}(\hat{\boldsymbol{\theta}})]$; the theoretical development presented by these authors suggests that this estimator is approximately unbiased when Assumptions (A2), (B1) and (C1) are satisfied.

4. Examples

We now illustrate the results of the previous section with four examples. Each of the first three examples consists of a particular random field observed at a specially chosen spatial configuration of locations, and involves theoretical calculations. The fourth example illustrates the estimation of MSPE with actual spatial data. Throughout these examples, let r denote the Euclidean distance between two points in D, let $\mathbf{1}_n$ denote an $n \times 1$ vector of ones, and let $\boldsymbol{J}_n = \mathbf{1}_n\mathbf{1}_n'$.

4.1 *Example* 1

This example is meant to illustrate that situations exist in which the EMSPE $m_1(\hat{\boldsymbol{\theta}})$ *is* an appropriate estimator of the EBLUP's MSPE. Take $\mathcal{F}$ to be a two-dimensional isotropic IRF-0 with generalized covariance function

$$G(r;\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } r = 0, \\ -\theta_1 - \theta_2 g(r), & \text{if } r > 0, \end{cases}$$

where $\Theta = \{\boldsymbol{\theta} : \theta_1 > 0, \theta_2 \geq 0\}$ and where $g(\cdot)$ is "any" function of r [any function such that $G(\,\cdot\,;\boldsymbol{\theta})$ is a valid generalized covariance function of order 0 for all $\boldsymbol{\theta} \in \Theta$]; for example, we could take $g(r) = r$, in which case $-G(\,\cdot\,;\boldsymbol{\theta})$ would be the oft-used linear semivariogram plus nugget effect. Take $\boldsymbol{t}_1$, $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ to be points that form an equilateral triangle with sides of length one unit. We note in passing that lattices built up from such a configuration have been studied (with regard to their sampling efficiency) by Yfantis *et al.* (1987). Let $\boldsymbol{t}_0$ represent the point at the center of the triangle and take y_0 to be the value of $\mathcal{F}$ at $\boldsymbol{t}_0$.

It is easily verified that $\boldsymbol{x}_0 = 1$, $\boldsymbol{X} = \mathbf{1}_3$, $k_0 = 0$, $\boldsymbol{k} = -(\theta_1 + \theta_2 g_0)\mathbf{1}_3$, and the generalized covariance matrix (here of order zero) is $\boldsymbol{K} = (\theta_1 + \theta_2)\boldsymbol{I} - (\theta_1 + \theta_2)\boldsymbol{J}_3$, where $g_0 = g(1/\sqrt{3}\,)$. It is then a fairly simple exercise to show that $p_1(\boldsymbol{y};\boldsymbol{\theta}) = (1/3)\sum_{i=1}^{3} y_i$ and $m_1(\boldsymbol{\theta}) = (4/3)\theta_1 + (2g_0 - 2/3)\theta_2$. Observe that $p_1(\boldsymbol{y};\boldsymbol{\theta})$ does not depend on the value of $\boldsymbol{\theta}$, which implies that $p_2(\boldsymbol{y})$ is also the BLUP of y_0 and that $m_1(\boldsymbol{\theta}) = m^*(\boldsymbol{\theta}) = m_2(\boldsymbol{\theta})$. Furthermore, we see that this MSPE is a linear function of the unknown parameters; consequently, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ_1 and θ_2 (as are, e.g., the minimum variance quadratic unbiased estimators; Kitanidis (1985)), then $m_1(\hat{\boldsymbol{\theta}})$ is an unbiased estimator of the MSPE. Thus, in this example, if $\hat{\boldsymbol{\theta}}$ is unbiased, then so is $m_1(\hat{\boldsymbol{\theta}})$.

4.2 *Example* 2

Take $\mathcal{F}$ to be a two-dimensional, second-order stationary, isotropic, Gaussian random field with covariance function $C(r;\boldsymbol{\alpha}) = \alpha_1\delta(r) + \alpha_2 c(r)$, where $\alpha_1 > 0$ and $\alpha_2 \geq 0$, $\delta(\cdot)$ is the Dirac delta function, and $c(\cdot)$ is a positive definite continuous function satisfying $c(0) = 1$, $|c(r)| \leq 1$ for all r, and $c(r) = 0$ for all $r \geq r^*$. Here, r^* is assumed known. [An example of such a covariance function is the spherical covariance function with unknown nugget effect and unknown sill but known range; for a description of this and other spatial covariance functions, see Journel and Huijbregts (1978).] Take $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_8$ to be points located on the perimeter of an $R_1 \times R_2$ rectangle and take y_0 to be the value of $\mathcal{F}$ at the point $\boldsymbol{t}_0$ at the center of the rectangle, as depicted in Fig. 1. The points at which $\mathcal{F}$ is observed occur in pairs on each side of the rectangle: the distance between points within pairs is constant and is denoted by r_1, the distance from each point on the "top" and "bottom" sides of the rectangle to $\boldsymbol{t}_0$ is constant and is denoted by r_2, and the distance from each point on the left and right sides of the rectangle to $\boldsymbol{t}_0$ is constant and is denoted by r_3. It follows that the shortest distance between two points on adjacent sides of the rectangle is equal to a constant r_4; assume that $r_1 < r^* \leq \min(r_4, R_1, R_2)$. Let $c_i = c(r_i)$ for $i = 1, 2, 3$, and assume that $c_1 > 0$. For convenience, we reparameterize the covariance function by putting $\theta_1 = \alpha_1 + \alpha_2(1 - c_1)$ and $\theta_2 = \alpha_2$,

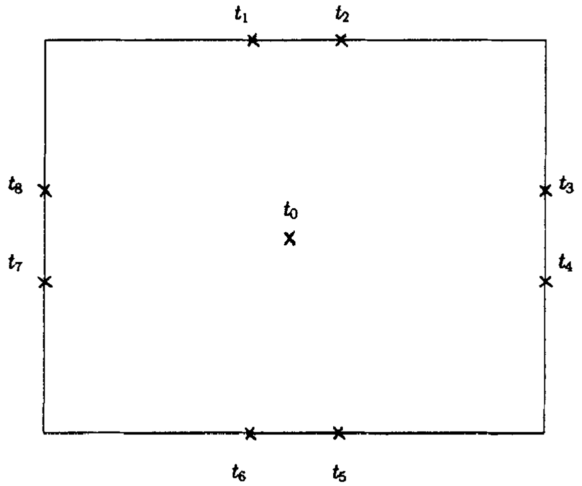

Fig. 1. Spatial configuration for Example 2.

in which case $\Theta = \{\boldsymbol{\theta} : \theta_1 > 0, \theta_2 \geq 0\}$. Note that Assumptions (A2) and (C1) are satisfied here.

It can be shown (see Zimmerman and Cressie (1989)) that the MSPE of the BLUP of y_0 is

$$(4.1) \quad m_1(\boldsymbol{\theta}) = \left(\frac{9}{8} + b\right)\theta_1 + \left(\frac{5}{4}c_1 - c_2 - c_3 - 2c_1 b\right)\theta_2 - b[\theta_1^2/(\theta_1 + 2c_1\theta_2)],$$

where $b = (c_2 - c_3)^2/2c_1^2$. More to the point, it can be shown that the estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)'$ satisfies Assumption (B2), where $\hat{\theta}_1 = S_1/4$, $\hat{\theta}_2 = [(S_2/3) - (S_1/4)]/2c_1$, $S_1 = \sum_{i=1}^8 y_i^2 - (1/2)\sum_{i=1}^4 (y_{2i-1} + y_{2i})^2$ and $S_2 = (1/2)\sum_{i=1}^4 (y_{2i-1} + y_{2i} - 2\bar{y})^2$. Hence, from Theorem 3.4,

$$(4.2) \qquad E[m_1(\hat{\boldsymbol{\theta}})] = m_1(\boldsymbol{\theta}) - \frac{7}{2}b[\theta_1^2/(\theta_1 + 2c_1\theta_2)]$$

and

$$(4.3) \qquad m_2(\boldsymbol{\theta}) = m_1(\boldsymbol{\theta}) + \frac{7}{2}b[\theta_1^2/(\theta_1 + 2c_1\theta_2)].$$

Zimmerman and Cressie (1989) exploit these results to show that an unbiased estimator of $m_2(\boldsymbol{\theta})$ is given by

$$(4.4) \qquad \hat{m}_2 = m_1(\hat{\boldsymbol{\theta}}) + \frac{14}{9}b[\hat{\theta}_1^2/(\hat{\theta}_1 + 2c_1\hat{\theta}_2)].$$

Comparison of expressions (4.2) and (4.3) reveals that $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_2(\boldsymbol{\theta})$ by an amount equal to

$$(4.5) \quad u(c_1, c_2, c_3, \alpha_1, \alpha_2) = 7b[\theta_1^2/(\theta_1 + 2c_1\theta_2)] = \frac{7(c_2 - c_3)^2[\alpha_1 + \alpha_2(1 - c_1)]^2}{2c_1^2[\alpha_1 + \alpha_2(1 + c_1)]}.$$

There are three features of interest in expression (4.5). First, for any fixed values of c_2, c_3, α_1 and α_2, $u(\cdot)$ is a decreasing function of c_1, which, if $c(\cdot)$ is monotone, implies that $u(\cdot)$ is an increasing function of the within-pair separation distance r_1. Thus, if $c(\cdot)$ is monotone, the discrepancy between $E[m_1(\hat{\boldsymbol{\theta}})]$ and $m_2(\boldsymbol{\theta})$ can be minimized by taking r_1 equal to zero, i.e., by replicating twice at the midpoints of each of the four sides of the rectangle. The resulting configuration is the basic building block of a rectangular lattice (replicated twice). Second, if $r_1 = 0$, then $u(\cdot)/\alpha_1$ is a decreasing function of α_2/α_1, the ratio of the partial sill parameter to the nugget effect, and $u(\cdot)/\alpha_1$ tends to zero as α_2/α_1 tends to infinity. Last, for any fixed values of c_1, α_1 and α_2, $u(\cdot)$ is an increasing function of the difference $|c_2 - c_3|$ and equals zero if $c_2 = c_3$.

This example is tractable enough that a simplified expression exists for the approximate MSPE $m^*(\boldsymbol{\theta})$. Zimmerman and Cressie (1989) show that $\boldsymbol{A}(\boldsymbol{\theta}) = \{a_{ij}\}$, where

$$
\begin{aligned}
a_{11} &= 2(c_2 - c_3)^2(\theta_1 + 2c_1\theta_2)(\eta_1 + 2c_1\eta_2)^2, \\
a_{12} = a_{21} &= 2(c_2 - c_3)^2(\theta_1 + 2c_1\theta_2)(\eta_1 + 2c_1\eta_2)(\eta_3 + 2c_1\eta_4), \\
a_{22} &= 2(c_2 - c_3)^2(\theta_1 + 2c_1\theta_2)(\eta_3 + 2c_1\eta_4)^2,
\end{aligned}
$$

and

$$
\eta_1 = -\frac{\theta_2}{\theta_1^2}, \quad \eta_2 = \frac{2\theta_2^2(\theta_1 + c_1\theta_2)}{\theta_1^2(\theta_1 + 2c_1\theta_2)^2}, \quad \eta_3 = \frac{1}{\theta_1}, \quad \eta_4 = -\frac{2\theta_2(\theta_1 + c_1\theta_2)}{\theta_1(\theta_1 + 2c_1\theta_2)^2};
$$

they also show that the exact mean equared error matrix of $\hat{\boldsymbol{\theta}}$ is

$$
\boldsymbol{B}(\boldsymbol{\theta}) = \frac{1}{4c_1^2}\begin{bmatrix} 2c_1^2\theta_1^2 & -\theta_1^2 c_1 \\ -\theta_1^2 c_1 & \dfrac{2(\theta_1 + 2c_1\theta_2)^2}{3} + \dfrac{\theta_1^2}{2} \end{bmatrix}.
$$

The approximate MSPE $m^*(\boldsymbol{\theta})$ can then be computed using expression (4.1) and the expressions given above for $\boldsymbol{A}(\boldsymbol{\theta})$ and $\boldsymbol{B}(\boldsymbol{\theta})$.

In Fig. 2 we compare $m_1(\boldsymbol{\theta})$, $E[m_1(\hat{\boldsymbol{\theta}})]$, and $m^*(\boldsymbol{\theta})$ to $m_2(\boldsymbol{\theta})$, where for illustration,

$$
c(r) = \begin{cases} 1 - 3r/2\sqrt{5} + r^3/[2(5^{3/2})], & 0 \le r \le \sqrt{5}, \\ 0, & \text{otherwise}, \end{cases}
$$

$r_1 = 0$, and $r_2 = (1/2)r_3 = 1$ (and necessarily $r_4 = \sqrt{5}$) are chosen. Therefore, $C(\,\cdot\,;\boldsymbol{\theta})$ is in this case the spherical covariance function (with range equal to $\sqrt{5}$) plus nugget effect. Taking r_1 to equal zero makes the spatial configuration as favorable as possible to $E[m_1(\hat{\boldsymbol{\theta}})]$ (as described above) and makes the (α_1, α_2) and (θ_1, θ_2) parametrizations equivalent. Figure 2 clearly demonstrates here that $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_2(\boldsymbol{\theta})$, particularly when $\gamma = \theta_2/\theta_1$ is small: the relative bias of $m_1(\hat{\boldsymbol{\theta}})$, i.e., $|E[m_1(\hat{\boldsymbol{\theta}})] - m_2(\boldsymbol{\theta})|/m_2(\boldsymbol{\theta})$, ranges from 33.2% when γ is near zero to 1.2% when $\gamma = 4$. Now, the correlation between observations located a distance r $(0 < r \le \sqrt{5})$ apart is $[\theta_2 c(r)]/(\theta_1 + \theta_2) = [\gamma c(r)]/(1 + \gamma)$, so the strength of the spatial correlation increases with γ. Thus, the performance of $m_1(\hat{\boldsymbol{\theta}})$ is poor when the spatial correlation is weak but improves as the spatial correlation gets stronger. Moreover, in this example at least, $m^*(\boldsymbol{\theta})$ offers a reasonable approximation to $m_2(\boldsymbol{\theta})$, which suggests that $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ would be better estimators than $m_1(\hat{\boldsymbol{\theta}})$. Of course, the estimator of choice in this example is the exactly unbiased estimator $\hat{m}_2$.

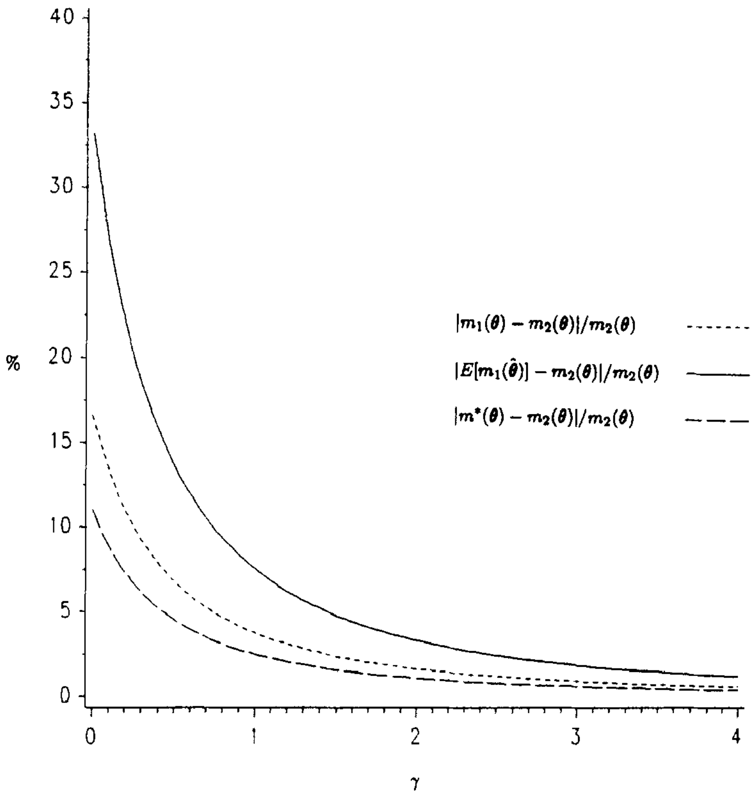

Fig. 2. Graphical comparison of $|m_1(\boldsymbol{\theta}) - m_2(\boldsymbol{\theta})|/m_2(\boldsymbol{\theta})$, $|E[m_1(\hat{\boldsymbol{\theta}})] - m_2(\boldsymbol{\theta})|/m_2(\boldsymbol{\theta})$ and $|m^*(\boldsymbol{\theta}) - m_2(\boldsymbol{\theta})|/m_2(\boldsymbol{\theta})$ versus γ over the range $0 < \gamma \leq 4$ for Example 2.

A difficulty with the estimator $\hat{\boldsymbol{\theta}}$ in this example is that it can, with nonzero probability, assume values outside Θ. One strategy for dealing with this problem is to define a modified estimator $\hat{\boldsymbol{\theta}}^+$ as follows: $\hat{\theta}_1^+ = \hat{\theta}_1$ and $\hat{\theta}_2^+ = \hat{\theta}_2$, if $\hat{\boldsymbol{\theta}} \in \Theta$, $\hat{\theta}_1^+ = (S_1 + S_2)/7$ and $\hat{\theta}_2^+ = 0$, if $\hat{\boldsymbol{\theta}} \notin \Theta$. It can be shown, by directly maximizing the likelihood function associated with $\boldsymbol{z}$ over the allowable parameter space Θ, that the estimator $\hat{\boldsymbol{\theta}}^+$ so defined is the REML estimator of $\boldsymbol{\theta}$. However, unlike $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}^+$ is biased; moreover, $(\hat{\theta}_1^+, \hat{\theta}_2^+)'$ is not a complete sufficient statistic (for the distribution of $\boldsymbol{z}$), so that one of the conditions of Theorem 3.4 is not satisfied. Nevertheless, it is our intuition that the estimator of $m_2(\boldsymbol{\theta})$ obtained by replacing

$\hat{\boldsymbol{\theta}}$ in expression (4.4) with $\hat{\boldsymbol{\theta}}^{+}$ is a better estimator than either $m_1(\hat{\boldsymbol{\theta}})$, $m_1(\hat{\boldsymbol{\theta}}^{+})$, $m^*(\hat{\boldsymbol{\theta}})$, $m^*(\hat{\boldsymbol{\theta}}^{+})$, $m^{**}(\hat{\boldsymbol{\theta}})$, or $m^{**}(\hat{\boldsymbol{\theta}}^{+})$; we think that it should retain the attractive properties of (4.4) with $\hat{\boldsymbol{\theta}}$, yet avoid the unattractive property of being functionally dependent on an estimator of $\boldsymbol{\theta}$ that can take values outside the parameter space.

4.3 *Example* 3

Take $\mathcal{F}$ to be a one-dimensional Gaussian IRF-0 with covariance function $C(s,t;\boldsymbol{\theta}) = \theta_1\delta(s-t) + \theta_2 \min\{s,t\}$, where $\delta(\cdot)$ is the Dirac delta function and $\Theta = \{\boldsymbol{\theta} : \theta_1 > 0, \theta_2 \geq 0\}$; this covariance function is a multiple of that of a Wiener process (which is not second-order stationary but is an IRF-0) plus independent white noise. Take $t_1, \ldots, t_n$ (where n is even) to be a unit-spaced regular lattice in $\mathbf{R}^1$ with t_1 at the origin. Observe, as in Example 2, that Assumptions (A2) and (C1) are satisfied here.

We consider two choices for the point t_0 at which to predict: (I) the point halfway between $t_{(n/2)}$ and $t_{(n/2)+1}$; (II) the point one unit to the right of t_n. The resulting configutations of points, hereafter called Configurations I and II, respectively, are depicted in Fig. 3.

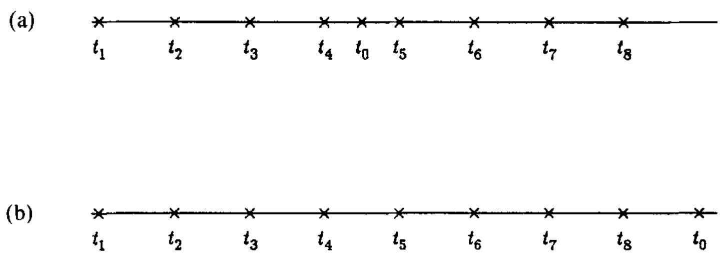

Fig. 3. Spatial configurations for Example 3, illustrated for the case $n = 8$: (a) Configuration I; (b) Configuration II.

In this example, $\boldsymbol{x}_0 = 1$, $\boldsymbol{X} = \mathbf{1}_n$, and the (ordinary) covariance matrix is $\boldsymbol{K} = \theta_1 \boldsymbol{I} + \theta_2 \boldsymbol{G}$, where

$$\boldsymbol{G} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 2 & \cdots & n-1 \end{bmatrix}.$$

The values of k_0 and $\boldsymbol{k}$ depend on the configuration: for Configuration I, $k_0 = \theta_1 + ((n-1)/2)\theta_2$ and $\boldsymbol{k} = \theta_2(0, 1, \ldots, n/2-1, (n-1)/2, (n-1)/2, \ldots, (n-1)/2)'$, whereas for Configuration II, $k_0 = \theta_1 + (n+1)\theta_2$ and $\boldsymbol{k} = \theta_2(0, 1, \ldots, n)'$. These expressions can be substituted into (2.2) and (2.3) to obtain simplified expressions for $p_1(\boldsymbol{y};\boldsymbol{\theta})$ and $m_1(\boldsymbol{\theta})$. We take $\hat{\boldsymbol{\theta}}$ to be the REML estimator of $\boldsymbol{\theta}$, i.e., the value $\hat{\boldsymbol{\theta}} \in \Theta$ that maximizes the restricted log-likelihood function $L(\boldsymbol{\theta};\boldsymbol{y}) = -(1/2)\log|\boldsymbol{L}'\boldsymbol{K}\boldsymbol{L}| - (1/2)\boldsymbol{z}'(\boldsymbol{L}'\boldsymbol{K}\boldsymbol{L})^{-1}\boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{L}'\boldsymbol{y}$ and $\boldsymbol{L}$ is an $n \times (n-1)$ matrix whose columns are the first $n-1$ columns of $\boldsymbol{I}-(1/n)\boldsymbol{J}_n$ (see, e.g., Kitanidis (1983)). The quantities $p_2(\boldsymbol{y})$, $m_1(\hat{\boldsymbol{\theta}})$, $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ can be readily computed, where in the computation of the latter two estimators we take $\boldsymbol{B}(\boldsymbol{\theta})$ to be the large-sample covariance matrix of $\hat{\boldsymbol{\theta}}$ or equivalently the inverse of the information matrix associated with the restricted log-likelihood function. The estimator $\hat{m}_2$ cannot be computed here, since $\hat{\boldsymbol{\theta}}$ does not satisfy Assumption (B2) and thus the evaluation of $E[m_1(\hat{\boldsymbol{\theta}})]$ and the determination of an unbiased estimator of $m_1(\boldsymbol{\theta})$ are intractable problems. The exact MSPE $m_2(\boldsymbol{\theta})$ is also intractable, so analytical comparisons such as those made in Example 2 cannot be made.

As an alternative to analytical comparisons, we conducted a simulation study in which $m_2(\boldsymbol{\theta})$, $m_1(\hat{\boldsymbol{\theta}})$, $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ could be compared empirically to each other and to $m_1(\boldsymbol{\theta})$. The simulation study involved (a) generating, for each of three values of θ_2 ($\theta_2 = .25, 1.0, 4.0$), 5,000 realizations of $\boldsymbol{y}^* = (\boldsymbol{y}, y_0)'$ for the special case of this example where $\theta_1 = 1$ and $n = 8$; (b) computing $\hat{\boldsymbol{\theta}}$, $[p_2(\boldsymbol{y}) - y_0]^2$, $m_1(\hat{\boldsymbol{\theta}})$, $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ for that realization; (c) averaging each of the quantities computed in part (b) over the 5,000 realizations. The averages of $[p_2(\boldsymbol{y}) - y_0]^2$, $m_1(\hat{\boldsymbol{\theta}})$, $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ so computed represent estimates of $m_2(\boldsymbol{\theta})$, $E[m_1(\hat{\boldsymbol{\theta}})]$, $E[m^*(\hat{\boldsymbol{\theta}})]$ and $E[m^{**}(\hat{\boldsymbol{\theta}})]$, respectively, and are displayed in Table 1.

Table 1. Comparison of MSPE's in Example 3. Reported values for $[p_2(\boldsymbol{y})-y_0]^2$, $m_1(\hat{\boldsymbol{\theta}})$, $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ are averages over 5,000 simulated realizations. Values in parentheses are estimated standard errors.

Configuration	θ_2/θ_1	$m_1(\boldsymbol{\theta})$	$[p_2(\boldsymbol{y}) - y_0]^2$	$m_1(\hat{\boldsymbol{\theta}})$	$m^*(\hat{\boldsymbol{\theta}})$	$m^{**}(\hat{\boldsymbol{\theta}})$
	.25	1.268	1.340 (.028)	1.209 (.012)	1.301 (.012)	1.393 (.014)
I	1.0	1.560	1.658 (.034)	1.506 (.015)	1.680 (.015)	1.854 (.018)
	4.0	2.414	2.672 (.056)	2.787 (.026)	3.253 (.028)	3.718 (.029)
	.25	1.641	1.828 (.038)	1.582 (.012)	2.251 (.018)	2.919 (.021)
II	1.0	2.618	2.983 (.062)	2.482 (.023)	3.253 (.026)	4.025 (.030)
	4.0	5.828	6.262 (.128)	6.027 (.055)	7.467 (.065)	8.907 (.074)

Table 1 clearly illustrates the result of Theorem 3.1, as $m_2(\boldsymbol{\theta})$ exceeds $m_1(\boldsymbol{\theta})$ by 5–15%. Moreover, we see that $m_1(\hat{\boldsymbol{\theta}})$ tends to underestimate $m_2(\boldsymbol{\theta})$ when θ_2/θ_1 is small but not when θ_2/θ_1 is large. Since large values of θ_2/θ_1 correspond to strong spatial correlation, the results of this example are similar to those of Example 2 in that the performance of $m_1(\hat{\boldsymbol{\theta}})$ improves as the spatial correlation gets stronger. On the other hand, the results for the alternative estimator $m^*(\hat{\boldsymbol{\theta}})$ are rather different here: it appears that $m^*(\hat{\boldsymbol{\theta}})$ tends to overestimate $m_2(\boldsymbol{\theta})$ when $\theta_2/\theta_1 = 4.0$ in the case of Configuration I and over a large range of values of θ_2/θ_1

in the case of Configuration II, and the bias of $m^{**}(\hat{\boldsymbol{\theta}})$ is even worse. Thus, while $m_1(\hat{\boldsymbol{\theta}})$ underestimates $m_2(\boldsymbol{\theta})$ under weak spatial correlation in this example, $m^*(\hat{\boldsymbol{\theta}})$ is generally not as successful an alternative to $m_1(\hat{\boldsymbol{\theta}})$ as it would appear to be in Example 2.

4.4 *Example* 4: *Iron ore data*

As our final example, we illustrate the estimation of MSPE with an actual data set consisting of iron ore (%Fe_2O_3) measurements taken from an ore body in Australia. The data, which lie on the nodes of an incomplete rectangular grid, were displayed by Cressie (1986) and have been analyzed previously by Cressie (1986) and Zimmerman and Zimmerman (1991). These authors found that the residuals from a median polish of the data were compatible with an IRF-0 model and that a suitable scaling of the coordinate axes permitted the use of an isotropic semivariogram (or equivalently, an isotropic generalized covariance function). From an examination of a plot of the estimated semivariogram, Zimmerman and Zimmerman (1991) chose to fit the class of generalized covariance function models:

$$G(r;\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } r = 0, \\ -\theta_1 - \theta_2(1-\theta_3^r), & \text{if } r > 0. \end{cases}$$

REML estimates of the parameters of $G(r;\boldsymbol{\theta})$ were found to be $\hat{\theta}_1 = 5.34$, $\hat{\theta}_2 = 6.38$ and $\hat{\theta}_3 = .895$.

Using this estimated generalized covariance function, Zimmerman and Zimmerman (1991) calculated the EBLUP of an unobserved median polish residual value at a point near the "center" of the data, obtaining $p_2(\boldsymbol{y}) = -.197\%$. The corresponding EMSPE $m_1(\hat{\boldsymbol{\theta}})$ equals $6.57(\%)^2$, whereas the more conservative MSPE estimate $m^*(\hat{\boldsymbol{\theta}})$ equals $7.57(\%)^2$. These data exhibit spatial dependence that is relatively weak [see Fig. 3 of Zimmerman and Zimmerman (1991)]; therefore, for reasons to be described in our concluding section, we would be inclined to use the latter value, $7.57(\%)^2$, as our estimate of the EBLUP's MSPE.

5. Concluding remarks

The traditional procedure for estimating the MSPE $m_2(\boldsymbol{\theta})$ of the EBLUP is to substitute an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in $m_1(\boldsymbol{\theta})$, that is, to use $m_1(\hat{\boldsymbol{\theta}})$. In addition, it is customary to use $m_1(\hat{\boldsymbol{\theta}})$ in the calculation of approximate confidence intervals for predicted values. For example, the interval $p_2(\boldsymbol{y}) \pm 2[m_1(\hat{\boldsymbol{\theta}})]^{1/2}$ is typically used as an approximate 95% confidence interval for a predicted value. Clearly, this interval may be much too optimistic if $m_1(\hat{\boldsymbol{\theta}})$ badly underestimates $m_2(\boldsymbol{\theta})$, or much too conservative if $m_1(\hat{\boldsymbol{\theta}})$ badly overestimates $m_2(\boldsymbol{\theta})$. In instances where $m_1(\hat{\boldsymbol{\theta}})$ is thought to underestimate $m_2(\boldsymbol{\theta})$ badly, we suggest that the alternative estimators $m^*(\hat{\boldsymbol{\theta}})$ or $m^{**}(\hat{\boldsymbol{\theta}})$ be used instead; this should result in more satisfactory point estimates of $m_2(\boldsymbol{\theta})$ and in more conservative (but more appropriate) confidence intervals for predicted values.

The practical issue we have attempted to address is whether $m_1(\hat{\boldsymbol{\theta}})$ is an accurate estimator of $m_2(\boldsymbol{\theta})$. It is difficult to make a general conclusion since the bias of $m_1(\hat{\boldsymbol{\theta}})$ and the usefulness of the alternative estimators $m^*(\hat{\boldsymbol{\theta}})$ and $m^{**}(\hat{\boldsymbol{\theta}})$ depend on which, if any, of Assumptions (A2), (B1), (B2) and (C1) are satisfied, and may also be affected by the number of observations n, the spatial configuration of the observations' locations, the location of the value to be predicted, the strength of the spatial correlation, and possibly other factors. One general guideline, which has emerged from the second and third examples here and from two other examples reported by Zimmerman and Zimmerman (1991), is that the performance of $m_1(\hat{\boldsymbol{\theta}})$ can often be improved upon when the spatial correlation is weak but that the performance of $m_1(\hat{\boldsymbol{\theta}})$ is adequate and sometimes superior to the alternative estimators when the spatial correlation is strong. It is interesting to note that the cases in which Zimmerman and Zimmerman (1991) found $m_1(\hat{\boldsymbol{\theta}})$ to be an approximately unbiased or positively biased estimator of $m_2(\boldsymbol{\theta})$ corresponded to cases in which the REML estimator of the nugget effect or sill value was rather badly positively biased. In view of these results, we recommend that practitioners use the alternative estimators $m^*(\hat{\boldsymbol{\theta}})$ or $m^{**}(\hat{\boldsymbol{\theta}})$ when $\hat{\boldsymbol{\theta}}$ is known to be (approximately) unbiased or $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is known to be negatively biased in the sense of Theorem 3.3, and the spatial correlation is known (or estimated) to be weak. In the absence of these conditions, we recommend the continued use of $m_1(\hat{\boldsymbol{\theta}})$ to estimate $m_2(\boldsymbol{\theta})$.

Acknowledgements

We thank the referees for helpful comments.

References

Bement, T. R. and Williams, J. S. (1969). Variance of weighted regression estimators when sampling errors are independent and heteroscedastic, *J. Amer. Statist. Assoc.*, **64**, 1369–1382.

Carroll, R. J., Wu, C. F. J. and Ruppert, D. (1988). The effect of estimating weights in weighted least squares, *J. Amer. Statist. Assoc.*, **83**, 1045–1054.

Cressie, N. (1985). Fitting variogram models by weighted least squares, *Math. Geol.*, **17**, 563–586.

Cressie, N. (1986). Kriging nonstationary data, *J. Amer. Statist. Assoc.*, **81**, 625–634.

Cressie, N. (1987). A nonparametric view of generalized covariances for kriging, *Math. Geol.*, **19**, 425–449.

Delfiner, P. (1976). Linear estimation of non-stationary spatial phenomena, *Advanced Geostatistics in the Mining Industry* (eds. M. Guarascio, M. David and C. Huijbregts), 49–68, Reidel, Dordrecht.

Dubrule, O. (1983). Two methods with different objectives: splines and kriging, *Math. Geol.*, **15**, 245–247.

Eaton, M. L. (1985). The Gauss-Markov theorem in multivariate analysis, *Multivariate Analysis — VI* (ed. P. R. Krishnaiah), 177–201, Elsevier, Amsterdam.

Fuller, W. A. and Hasza, D. P. (1981). Properties of predictors for autoregressive time series, *J. Amer. Statist. Assoc.*, **76**, 155–161.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized regression model, *J. Amer. Statist. Assoc.*, **57**, 369–375.

Harville, D. A. (1985). Decomposition of prediction error, *J. Amer. Statist. Assoc.*, **80**, 132–138.

Harville, D. A. and Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model, *J. Amer. Statist. Assoc.*, **87** (to appear).

Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*, Academic Press, London.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models, *J. Amer. Statist. Assoc.*, **79**, 853–862.

Khatri, C. G. and Shah, K. R. (1981). On the unbiased estimation of fixed effects in a mixed model for growth curves, *Comm. Statist. A—Theory Methods*, **10**, 401–406.

Kitanidis, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications, *Water Resources Research*, **19**, 909–921.

Kitanidis, P. K. (1985). Minimum variance unbiased quadratic estimation of covariances of regionalized variables, *Math. Geol.*, **17**, 195–208.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **71**, 135–146.

Marshall, R. J. and Mardia, K. V. (1985). Minimum norm quadratic estimation of components of spatial covariances, *Math. Geol.*, **17**, 517–525.

Matheron, G. (1973). The intrinsic random functions and their applications, *Adv. in Appl. Probab.*, **5**, 439–468.

Prasad, N. G. N. and Rao, J. N. K. (1986). On the estimation of mean square error of small area predictors, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 108–116, Washington, D.C.

Reinsel, G. C. (1980). Asymptotic properties of prediction errors for the multivariate autoregressive model using estimated parameters, *J. Roy. Statist. Soc. Ser. B*, **42**, 328–333.

Reinsel, G. C. (1984). Estimation and prediction in a multivariate random effects generalized linear model, *J. Amer. Statist. Assoc.*, **79**, 406–414.

Rothenberg, T. J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar, *Econometrica*, **52**, 827–842.

Starks, T. H. and Fang, J. H. (1982). The effect of drift on the experimental semivariogram, *J. Internat. Assoc. Math. Geol.*, **14**, 309–320.

Toyooka, Y. (1982). Prediction error in a linear model with estimated parameters, *Biometrika*, **69**, 453–459.

Wolfe, D. A. (1973). Some general results about uncorrelated statistics, *J. Amer. Statist. Assoc.*, **68**, 1013–1018.

Yamamoto, T. (1976). Asymptotic mean square prediction error for an autoregressive model with estimated coefficients, *Appl. Statist.*, **25**, 123–127.

Yfantis, E. A., Flatman, G. T. and Behar, J. V. (1987). Efficiency of kriging estimation for square, triangular, and hexagonal grids, *Math. Geol.*, **19**, 183–205.

Zimmerman, D. L. and Cressie, N. (1989). Improved estimation of the kriging variance, Tech. Report, No. 161, Department of Statistics, University of Iowa.

Zimmerman, D. L. and Zimmerman, M. B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors, *Technometrics*, **33**, 77–91.