Mathematical Geology, Vol. 38, No. 4, May 2006 (© 2006)
DOI: 10.1007/s11004-005-9022-8

Block Kriging for Lognormal Spatial Processes[1]

Noel Cressie[2]

Lognormal spatial data are common in mining and soil-science applications. Modeling the underlying spatial process as normal on the log scale is sensible; point kriging allows the whole region of interest to be mapped. However, mining and precision agriculture is carried out selectively and is based on block averages of the process on the original scale. Finding spatial predictions of the blocks assuming a lognormal spatial process has a long history in geostatistics. In this article, we make the case that a particular method for block prediction, overlooked in past times of low computing power, deserves to be reconsidered. In fact, for known mean, it is optimal. We also consider the predictor based on the "law" of permanence of lognormality. Mean squared prediction errors of both are derived and compared both theoretically and via simulation; the predictor based on the permanence-of-lognormality assumption is seen to be less efficient. Our methodology is applied to block kriging of phosphorus to guide precision-agriculture treatment of soil on Broom's Barn Farm, UK.

KEY WORDS: geostatistics, MSPE, permanence of lognormality, phosphorus, precision agriculture, spatial prediction.

INTRODUCTION

There have been quite a few publications in the past on geostatistics for lognormal data. The themes of these papers (Dowd, 1982; Journel, 1980; Marechal, 1974; Matheron, 1974; Rendu, 1979; Rivoirard, 1990; Roth, 1998) draw on the very best traditions of geostatistics: determine types of variogram models for lognormal data; decide whether to do inference on the original scale or the log scale; choose an optimality criterion for kriging; derive the kriging equations according to the optimality criterion; consider the cases of known or unknown mean (on the log scale); and consider whether knowing just the variogram (on the log scale) is enough to do kriging.

The purpose of this article is to take a fresh look at geostatistics for lognormal data, build on the results of the earlier papers, and develop new results in light of the statistical literature on linear models and transformations. We shall

[1]Received 10 September 2004; accepted 23 August 2005; Published online: 30 August 2006.
[2]Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210-1247; e-mail: ncressie@stat.ohio-state.edu

vigorously pursue two of the many possibilities, chosen based on the following principles:

- The original scale is for optimality criteria (including unbiasedness) but the log scale is for linear statistical analysis.
- Some form of stationarity is needed for estimation of spatial dependence but it is not needed for spatial prediction (i.e., kriging).
- Kriging is an empirical-Bayes methodology that requires efficient estimators of unknown parameters to be "plugged into" kriging equations.

Notice that "permanence of lognormality" (e.g., Rivoirard, 1990) is not one of our principles. In fact, one of the spatial predictors we consider *is* based on permanence and one is not, and in this article we give a set of recommendations (based on both geostatistical theory and a carefully designed simulation study) as to when assuming permanence is a reasonable thing to do. In a related paper, Cressie and Pavlicova (2005) find a way to correct the inherent bias in the predictor based on permanence, but still find it to be inefficient.

When discussing lognormality for geostatistical processes, two quite different issues have often arisen together. One has been the development and use of the de Wijsian variogram (Matheron, 1962) for modeling spatial dependence (the variogram has logarithmic shape), and the other has been lognormal kriging. de Wijs (1951) developed a simple model that split an orebody randomly into two halves, one with grade proportionately above average and the other with grade proportionately below average. If this is successively repeated and X_i is the random grade of one of the halves at the ith split, then after k splits, the grade of any one of the 2^k pieces is distributed as $\prod_{i=1}^{k} X_i$. As $k \to \infty$, the central limit theorem implies convergence in distribution to a lognormal random variable. This rather specialized model also gives rise to a variogram that is logarithmic in shape (Matheron, 1962), but it is clear that lognormal genesis can also happen in other ways (e.g., Brown and Sanders, 1981). It is the lognormality that we discuss in this article, and there is no requirement here that variograms be of de Wijsian form.

A very influential piece of writing on lognormal kriging has been the unpublished 43-page "note" by Matheron (1974). Matheron's approach is to look at the problem from all sides, with many calculations drawn from Matheron (1962) but no definitive conclusions. His writing touches on all the geostatistical themes given earlier. At the time it was written, the statistical influences of linear models, efficient parameter estimation, prediction theory, and workstation computing were not yet felt in geostatistics.

Some notation is needed for what follows in this article. Let the process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ denote a lognormal spatial process defined on a domain $D \subset \mathbb{R}^d$ such that it has a positive d-dimensional volume $|D|$. That is, the process:

$$Y(\mathbf{s}) \equiv \log Z(\mathbf{s}); \quad \mathbf{s} \in D, \tag{1}$$

is a Gaussian process defined by its first two moments:

$$\mu_Y(\mathbf{s}) \equiv E(Y(\mathbf{s})); \quad \mathbf{s} \in D, \tag{2}$$

$$C_Y(\mathbf{u}, \mathbf{v}) \equiv \text{cov}(Y(\mathbf{u}), Y(\mathbf{v})); \quad \mathbf{u}, \mathbf{v} \in D. \tag{3}$$

Consequently, from (1), $Z(\mathbf{s}) = \exp\{Y(\mathbf{s})\} > 0$; $\mathbf{s} \in D$, and from Aitchison and Brown (1957),

$$\mu_Z(\mathbf{s}) \equiv E(Z(\mathbf{s})) = \exp\{\mu_Y(\mathbf{s}) + (1/2)C_Y(\mathbf{s}, \mathbf{s})\}; \mathbf{s} \in D, \tag{4}$$

$$C_Z(\mathbf{u}, \mathbf{v}) \equiv \text{cov}(Z(\mathbf{u}), Z(\mathbf{v})) = \mu_Z(\mathbf{u})\mu_Z(\mathbf{v})[\exp\{C_Y(\mathbf{u}, \mathbf{v})\} - 1]. \tag{5}$$

From (4), $\mu_Z(\mathbf{s}) \geq \exp\{\mu_Y(\mathbf{s})\}$, giving rise to a potential source of bias when transforming back to the original scale. This is to be expected: Because of Jensen's inequality, $E(\exp\{Y(\mathbf{s})\}) \geq \exp\{E(Y(\mathbf{s}))\}$. The presence of the mean terms as multipliers in (5), the covariance on the original scale, is sometimes called the proportional effect.

The spatial dependence in a spatial process $Z(\cdot)$ can be characterized by the covariance function or by the *variogram*,

$$2\gamma_Z(\mathbf{u}, \mathbf{v}) \equiv \text{var}(Z(\mathbf{u}) - Z(\mathbf{v}));$$

we call $\gamma_Z(\mathbf{u}, \mathbf{v})$ the *semivariogram*. Now the variogram on the log scale can be written as,

$$2\gamma_Y(\mathbf{u}, \mathbf{v}) = C_Y(\mathbf{u}, \mathbf{u}) + C_Y(\mathbf{v}, \mathbf{v}) - 2C_Y(\mathbf{u}, \mathbf{v}),$$

and the same is true for $2\gamma_Z(\mathbf{u}, \mathbf{v})$. Hence, a formula for $2\gamma_Z(\mathbf{u}, \mathbf{v})$ can be derived in terms of $\mu_Y(\mathbf{s})$ and $C_Y(\mathbf{u}, \mathbf{v})$ via (4) and (5).

Matheron (1974) investigated whether $\mu_Y(\mathbf{s}) \equiv \mu_Y$ and $2\gamma_Y(\mathbf{u}, \mathbf{v}) \equiv 2\gamma_Y^*(\mathbf{u} - \mathbf{v})$ (i.e., intrinsic stationarity on the log scale) implied something similar on the original scale. In general, it does not, but if we assume further that $C_Y(\mathbf{s}, \mathbf{s}) \equiv \sigma_Y^2$, then

$$2\gamma_Z(\mathbf{u}, \mathbf{v}) = \exp\left(2\mu_Y + 2\sigma_Y^2\right)[1 - \exp\{-\gamma_Y^*(\mathbf{u} - \mathbf{v})\}],$$

which is a function of $\mathbf{u} - \mathbf{v}$. However, following the principle that linear statistical analysis is done on the log scale, $2\gamma_Z$ is of less interest than $2\gamma_Y$.

The spatial (lognormal) data are defined as the $(n \times 1)$ vector,

$$\mathbf{Z} \equiv (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))', \tag{6}$$

where $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ are known spatial locations. Then the transformed data,

$$\mathbf{Y} \equiv (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))' \tag{7}$$

are normally distributed and will be used to estimate unknown parameters in $\mu_Y(\mathbf{s})$ and $C_Y(\mathbf{u}, \mathbf{v})$, as well as to predict an unknown value $Y(\mathbf{s}_0)$; $\mathbf{s}_0 \in D$.

The prediction problem is sometimes called point kriging, and from Cressie (1993, Chapter 3) we see that the minimum mean squared error predictor $Y^*(\mathbf{s}_0)$ is given by,

$$Y^*(\mathbf{s}_0) \equiv E(Y(\mathbf{s}_0)|\mathbf{Y}) = \mu_Y(\mathbf{s}_0) + \mathbf{c}_Y(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y), \tag{8}$$

where $\mathbf{c}_Y(\mathbf{s}_0) \equiv (C_Y(\mathbf{s}_0, \mathbf{s}_1), \ldots, C_Y(\mathbf{s}_0, \mathbf{s}_n))'$, $\Sigma_Y \equiv \text{var}(\mathbf{Y})$, and $\boldsymbol{\mu}_Y \equiv (\mu_Y(\mathbf{s}_1), \ldots, \mu_Y(\mathbf{s}_n))'$. Notice that

$$\text{var}(Y(\mathbf{s}_0)|\mathbf{Y}) = C_Y(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}_Y(\mathbf{s}_0)'\Sigma_Y^{-1}\mathbf{c}_Y(\mathbf{s}_0), \tag{9}$$

which does not depend on $\mathbf{Y}$.

It is well known (e.g., Cressie, 1993, Section 3.4.5) that (8) is also the best heterogeneously linear predictor assuming that the mean function and the covariance function of $Y(\cdot)$ are known; this is called *simple kriging* in the geostatistics literature. In what follows in this section, we shall assume these first two moments are known unless specified otherwise. When they are unknown, as in the second section, they are estimated efficiently and plugged into the optimal predictor; that is, we take an empirical-Bayes approach to the construction of efficient spatial predictors. Cressie (1993, Section 3.4.5) shows that such an approach can in fact yield ordinary kriging and universal kriging for $Y(\cdot)$.

Scientific interest is in the process $Z(\cdot)$; hence, to predict $Z(\mathbf{s}_0)$ based on data $\mathbf{Z}$, classical prediction theory says that the optimal predictor is obtained by minimizing the mean squared prediction error,

$$E(Z(\mathbf{s}_0) - p(\mathbf{Z}; \mathbf{s}_0))^2,$$

with respect to the predictor p. The theory further tells us that the best predictor is (e.g., Cressie, 1993, p. 108):

$$Z^*(\mathbf{s}_0) \equiv E(Z(\mathbf{s}_0)|\mathbf{Z}). \tag{10}$$

Calculation of (10) is not always possible, which explains why geostatisticians compromise with the best *linear* predictor. Because $Z(\cdot)$ is lognormal, it is

unwise to use such a compromise here; in what follows, we evaluate (10):

$$\begin{aligned} Z^*(\mathbf{s}_0) &= E(\exp\{Y(\mathbf{s}_0)\}|\mathbf{Y}) \\ &= \exp\{E(Y(\mathbf{s}_0)|\mathbf{Y}) + (1/2)\mathrm{var}(Y(\mathbf{s}_0)|\mathbf{Y})\}, \end{aligned}$$

where the last equality is true because the conditional distribution of $Y(\mathbf{s}_0)$ given $\mathbf{Y}$ is normal. From (8) and (9),

$$\begin{aligned} Z^*(\mathbf{s}_0) &= \exp\{Y^*(\mathbf{s}_0) + (1/2)C_Y(\mathbf{s}_0, \mathbf{s}_0) - (1/2)\mathbf{c}_Y(\mathbf{s}_0)'\Sigma_Y^{-1}\mathbf{c}_Y(\mathbf{s}_0)\} \\ &= \exp\{Y^*(\mathbf{s}_0) + (1/2)\mathrm{var}(Y(\mathbf{s}_0)) - (1/2)\mathrm{var}(Y^*(\mathbf{s}_0))\}. \end{aligned} \tag{11}$$

Clearly, the optimal predictor (11) is loglinear in the data and unbiased. The last expression shows that another way to obtain the optimal predictor $E(Z(\mathbf{s}_0)|\mathbf{Z})$, would be to start with the optimal predictor on the log scale and transform back to the original scale, with an adjustment to preserve unbiasedness. A version of these calculations that incorporates measurement error is given in Appendix A.

An important problem in lognormal kriging is the spatial prediction of block values,

$$Z(B) \equiv \int_B Z(\mathbf{u})d\mathbf{u}/|B|, \quad B \subset D; \tag{12}$$

in the example given in the third section, where phosphorus in soil is studied, the application of a soil treatment is made on blocks (not at individual locations), and hence interest centers on predicting $Z(B)$ based on data $\mathbf{Z}$. The rest of this article is devoted to this problem. This article makes new contributions to lognormal block kriging by considering nonstationary means and covariances, by modeling trend as linear in known covariates (such as functions of spatial location), by allowing for the possibility of measurement error, and finally by deriving and comparing mean squared prediction error formulae for the kriging predictors considered.

The following section develops two spatial predictors for $Z(B)$ under the assumption that $Z(\cdot)$ is a lognormal process, and it derives their mean squared prediction errors (kriging variances). The third section applies the methodology of the second section to a data set consisting of the phosphorus content in the soil of Broom's Barn Farm, UK. A designed simulation experiment is used in the fourth section to compare the two spatial predictors (one of which relies on permanence of lognormality). Discussion and conclusions are given in the fifth section. Three technical appendices are given to allow a more concise presentation of the theory in the body of the article.

PREDICTION OF BLOCK VALUES FOR LOGNORMAL PROCESSES

Recall that $Z(\cdot) = \exp(Y(\cdot))$, where $Y(\cdot)$ is a Gaussian process, and for the moment we assume that the first two moments $\mu_Y(\cdot)$ and $C_Y(\cdot, \cdot)$ of $Y(\cdot)$ are known (section "Prediction when Parameters are Known"). This is in line with the principle that kriging is an empirical-Bayes methodology that involves parameter estimates being plugged in at the final stage of construction of the spatial predictor (section "Prediction when only Covariance Parameters are Known").

Prediction when Parameters are Known

The block value $Z(B)$ is the average of $Z(\cdot)$ over a given block $B \subset D$; see (12). The mean squared prediction error of a predictor $p(\mathbf{Z}; B)$ is

$$E(Z(B) - p(\mathbf{Z}; B))^2; \quad B \subset D, \tag{13}$$

and its minimization with respect to p yields the optimal predictor,

$$Z^*(B) = E(Z(B)|\mathbf{Z});$$

compare this to Eq. (10). Hence,

$$\begin{aligned} Z^*(B) &= \int_B E(Z(\mathbf{u})|\mathbf{Z}) d\mathbf{u}/|B| \\ &= \int_B Z^*(\mathbf{u}) d\mathbf{u}/|B|, \end{aligned} \tag{14}$$

where $Z^*(\cdot)$ is given by (10). The result (14) is general; when $Z(\cdot)$ is lognormal, we obtain from (11) the optimal predictor,

$$Z^*(B) = \int_B \exp\{Y^*(\mathbf{u}) + (1/2)C_Y(\mathbf{u}, \mathbf{u}) - (1/2)\mathbf{c}_Y(\mathbf{u})' \Sigma_Y^{-1} \mathbf{c}_Y(\mathbf{u})\} d\mathbf{u}/|B|. \tag{15}$$

The predictor $Z^*(B)$ was not used in the past because of a comment from Matheron (1974) that it "is too heavy to be used effectively in practice"; Rivoirard (1990) says it "would be possible but difficult" to compute, but Cressie (1993, p. 136) proposes it without comment about difficulties. In fact, vast increases in computing power in recent times have made quadrature of the integrand in (15) a trivial exercise.

A good deal of the literature on kriging for lognormal data has been devoted to a predictor based on an assumption of "permanence of lognormality" (e.g.,

Rivoirard, 1990). The idea is very simple: If $Z(\cdot)$ is a lognormal process, then $\log Z(B)$ is normally distributed, *provided the block is not too large*. None of the articles referred to at the beginning of the first section says what "not too large" actually means in practice. However, intuitively, the approximation of permanence makes a lot of sense.

The simple model of splitting referred to in the first section (de Wijs, 1951) gives rise to any one of the 2^k pieces at the kth split being distributed as $\prod_{i=1}^{k} X_i$. Upon aggregating rather than splitting, one obtains pieces given at the $(k-1)$th split and hence with distribution $\prod_{i=1}^{k-1} X_i$. As $k \to \infty$, both converge to a lognormal distribution. At some level of aggregation, the (approximate) lognormality will break down. In this article, we see that positive spatial dependence allows the blocks where approximate lognormality holds to be larger. Furthermore, in a small side study, where negative spatial dependence (in $\mathbb{R}^1$) was also assumed, we found that under aggregation the lognormality approximation improved for increased negative dependence, but then deteriorated rapidly as the negative dependence went from moderate to strong.

It is worth noting here that although the lognormal distribution is infinitely divisible (Thorin, 1977), it is not stable. That is, the weighted sum of two lognormals is no longer lognormal, and hence from a mathematical point of view, permanence is not possible. However, our simulations in the fourth section show that in many cases it is a very appropriate approximation.

In what follows, we present a predictor based on permanence. This allows comparison to the optimal predictor (15), both theoretically and through simulation. And, through the results of the simulation, specific recommendations as to when the permanence approximation is appropriate, are given.

Suppose that we wish to predict $Y(B) \equiv \int_B Y(\mathbf{u}) d\mathbf{u}/|B|$ based on data (on the log scale) $\mathbf{Y}$. The predictor of choice is

$$Y^*(B) \equiv E(Y(B)|\mathbf{Y}) = \mu_Y(B) + \mathbf{c}_Y(B)' \Sigma_Y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y), \tag{16}$$

where $\mu_Y(B) \equiv \int_B \mu_Y(\mathbf{u})\, d\mathbf{u}/|B|$, $\mathbf{c}_Y(B) \equiv (C_Y(B, \mathbf{s}_1), \ldots, C_Y(B, \mathbf{s}_n))'$, and $C_Y(B, \mathbf{v}) \equiv \int_B C_Y(\mathbf{u}, \mathbf{v}) d\mathbf{u}/|B|$. One possible ad hoc predictor of $Z(B)$ is $Z^+(B) \equiv \exp\{Y^*(B) + k^+\}$, where k^+ is an adjustment for bias. This approach is explored further in Cressie and Pavlicova (2005). In what follows, we give the predictor based on the permanence-of-lognormality assumption, a predictor that also depends on $Y^*(B)$.

First observe that we can write

$$Z(B) = E(Z(\mathbf{U})), \tag{17}$$

where the expectation is with respect to $\mathbf{U}$, a uniform random variable on B, independent of $Z(\cdot)$. Then $E(Z(B)) = E(Z(\mathbf{U})) = E(\exp\{Y(\mathbf{U})\})$, and assuming

permanence of lognormality,

$$\begin{aligned} E(Z(B)) &= E(\exp\{Y(\mathbf{U})\}) \\ &\simeq \exp\{E(Y(\mathbf{U})) + (1/2)\mathrm{var}(Y(\mathbf{U}))\} \\ &= \exp\left\{\int_B \mu_Y(\mathbf{u})d\mathbf{u}/|B| + (1/2)\mathrm{var}(Y(\mathbf{U}))\right\}, \end{aligned}$$

where the "$\simeq$" would be "$=$" if the permanence assumption were true. Now,

$$\begin{aligned} \mathrm{var}(Y(\mathbf{U})) &= E[\mathrm{var}(Y(\mathbf{U}) \mid Y(\cdot))] + \mathrm{var}[E(Y(\mathbf{U})|Y(\cdot))] \\ &= E\left[\int_B (Y(\mathbf{u}) - Y(B))^2 d\mathbf{u}/|B|\right] + \mathrm{var}(Y(B)) \\ &= \int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B| + \int_B (\mu_Y(\mathbf{u}) - \mu_Y(B))^2 d\mathbf{u}/|B|\,. \end{aligned}$$

Thus, an (approximately) unbiased predictor of $Z(B)$, based on the permanence assumption, is $Z^{@}(B) = \exp\{Y^*(B) + k^{@}\}$, where

$$\begin{aligned} Z^{@}(B) = \exp\{&Y^*(B) + (1/2)\int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B| \\ &+ (1/2)\int_B (\mu_Y(\mathbf{u}) - \mu_Y(B))^2 d\mathbf{u}/|B| - (1/2)\mathbf{c}_Y(B)'\Sigma_Y^{-1}\mathbf{c}_Y(B)\}\,. \end{aligned} \quad (18)$$

To our knowledge, this formula has not appeared before and, inspired by the representation (17), it offers a new way to consider the permanence of lognormality.

Comparing the predictors (15) and (18), we see that both are unbiased (actually (18) only approximately so), both require computation of kriging predictors on the log scale, and both involve quadrature of quantities involving the first two moments $\mu_Y(\cdot)$ and $C_Y(\cdot,\cdot)$. A version of these calculations that incorporates measurement error is given in Appendix A. In the next section, we invoke the empirical-Bayes principle of plugging in an efficient estimator for the unknown mean function.

Prediction when only Covariance Parameters are Known

In this presentation, we shall assume that the mean function $\mu_Y(\cdot) \equiv \mu_Y$, a constant independent of location. The case where $\mu_Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$ is considered in Appendix C. In contrast to the previous subsection, it is now assumed that μ_Y is unknown and has to be estimated.

Return to the problem of predicting $Z(\mathbf{s}_0)$ and note that $Z^*(\mathbf{s}_0)$ given by (11) depends on μ_Y through

$$Y^*(\mathbf{s}_0) = \mu_Y + \mathbf{c}_Y(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - \mu_Y\mathbf{1}),$$

where $\mathbf{1} \equiv (1, \ldots, 1)'$ is an $(n \times 1)$ vector of 1s. This simple kriging predictor becomes an ordinary kriging predictor when the generalized least squares estimator for μ_Y, $\widehat{\mu}_Y \equiv (\mathbf{1}'\Sigma_Y^{-1}\mathbf{1})^{-1}\mathbf{1}'\Sigma_Y^{-1}\mathbf{Y}$, is plugged in for the unknown μ_Y (Cressie, 1993, p. 173):

$$\widehat{Y}(\mathbf{s}_0) = \widehat{\mu}_Y + \mathbf{c}_Y(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - \widehat{\mu}_Y\mathbf{1}). \tag{19}$$

Following the principle that the predictor should be unbiased on the original scale, we obtain the unbiased predictor,

$$\begin{aligned}\check{Z}(\mathbf{s}_0) &\equiv \exp\{\widehat{Y}(\mathbf{s}_0) + (1/2)\mathrm{var}(Y(\mathbf{s}_0)) - (1/2)\mathrm{var}(\widehat{Y}(\mathbf{s}_0))\} \\ &= \exp\{\widehat{Y}(\mathbf{s}_0) + (1/2)\sigma^2_{Y,k}(\mathbf{s}_0) - m(\mathbf{s}_0)\},\end{aligned} \tag{20}$$

where $\widehat{Y}(\mathbf{s}_0) \equiv \Sigma_{i=1}^n \lambda_i(\mathbf{s}_0)\mathbf{Y}(\mathbf{s}_0)$ is the ordinary kriging predictor given by (19); $\boldsymbol{\lambda}(\mathbf{s}_0) \equiv (\lambda_1(\mathbf{s}_0), \ldots, \lambda_n(\mathbf{s}_0))'$ and $m(\mathbf{s}_0)$ solve the ordinary kriging equations,

$$\begin{aligned}\Sigma_Y\boldsymbol{\lambda}(\mathbf{s}_0) &= \mathbf{c}_Y(\mathbf{s}_0) + \mathbf{1}m(\mathbf{s}_0) \\ \mathbf{1}'\boldsymbol{\lambda}(\mathbf{s}_0) &= 1;\end{aligned}$$

and the kriging variance is

$$\sigma^2_{Y,k}(\mathbf{s}_0) = C_Y(\mathbf{s}_0, \mathbf{s}_0) - \boldsymbol{\lambda}(\mathbf{s}_0)'\mathbf{c}_Y(\mathbf{s}_0) + m(\mathbf{s}_0).$$

The predictor (20) can be found in Matheron (1974), Journel (1980), Rivoirard (1990), and Cressie (1993, p. 135).

Analogously to the definition of the optimal predictor (15), we define its empirical-Bayes version:

$$\check{Z}(B) = \int_B \check{Z}(\mathbf{u})d\mathbf{u}/|B|, \tag{21}$$

where $\check{Z}(\cdot)$ is given by (20). This will be compared with the empirical-Bayes version of the predictor based on the permanence approximation, which we now derive.

The predictor (18) takes the form $\exp\{Y^*(B) + k^@\}$, where $k^@$ is chosen to (approximately) correct for bias; recall that

$$Y^*(B) = \mu_Y + \mathbf{c}_Y(B)'\Sigma_Y^{-1}(\mathbf{Y} - \mu_Y\mathbf{1}),$$

which becomes an ordinary kriging predictor when $\widehat{\mu}_Y \equiv (\mathbf{1}'\Sigma_Y^{-1}\mathbf{1})^{-1}\mathbf{1}'\Sigma_Y^{-1}\mathbf{Y}$ is plugged in for the unknown μ_Y (Cressie, 1993, p. 173):

$$\begin{aligned}\widehat{Y}(B) &= \widehat{\mu}_Y + \mathbf{c}_Y(B)'\Sigma_Y^{-1}(\mathbf{Y} - \widehat{\mu}_Y\mathbf{1}) \\ &\equiv \Sigma_{i=1}^n \lambda_i(B)Y(\mathbf{s}_i)\,.\end{aligned} \tag{22}$$

We shall now derive a predictor of the form, $\widetilde{Z}(B) \equiv \exp\{\widehat{Y}(B) + \widetilde{k}\}$, where $\widetilde{k}$ is chosen to (approximately) correct for bias.

From the previous subsection, under the assumption of permanence of log-normality and for a constant mean μ_Y,

$$\begin{aligned}E(Z(B)) &\simeq \exp\{\mu_Y + (1/2)\mathrm{var}(Y(\mathbf{U}))\} \\ &= \exp\{\mu_Y + (1/2)\int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B|\}\,.\end{aligned}$$

Also,

$$\mathrm{var}(\widehat{Y}(B)) = \boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\,,$$

where $\boldsymbol{\lambda}(B) \equiv (\lambda_1(B), \ldots, \lambda_n(B))'$. Upon noting that $\widehat{Y}(B)$ is normally distributed and combining these last two results, we see that an approximately unbiased predictor of $Z(B)$, based on the permanence assumption, is:

$$\begin{aligned}\widetilde{Z}(B) &\equiv \exp\{\widehat{Y}(B) + (1/2)\int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B| - (1/2)\boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\} \\ &= \exp\{\widehat{Y}(B) + (1/2)\int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B| - (1/2)\int_B\int_B C_Y(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}/|B|^2 \\ &\quad + (1/2)\sigma^2_{Y,k}(B) - m(B)\}\,,\end{aligned} \tag{23}$$

where $\widehat{Y}(B) = \boldsymbol{\lambda}(B)'\mathbf{Y}$ is the ordinary kriging predictor, $\boldsymbol{\lambda}(B)$ and $m(B)$ solve the ordinary kriging equations,

$$\begin{aligned}\Sigma_Y\boldsymbol{\lambda}(B) &= \mathbf{c}_Y(B) + \mathbf{1}m(B) \\ \mathbf{1}'\boldsymbol{\lambda}(B) &= 1,\end{aligned}$$

and the kriging variance is

$$\sigma^2_{Y,k}(B) = \int_B \int_B C_Y(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}/|B|^2 - \boldsymbol{\lambda}(B)' \mathbf{c}_Y(B) + m(B)\,.$$

The equality that yields the predictor $\widetilde{Z}(B)$ given by (23) is proved in Appendix B.

The two lognormal kriging predictors (21) and (23) of $Z(B)$ can be compared through their respective mean squared prediction errors. We expect (21) to do better than (23), since it is developed from the optimal predictor. This is borne out in the simulation experiment described in the fourth section.

We now derive mean squared prediction error formulae associated with the predictors (21) and (23); we see in the fourth section that these theoretical expressions give an excellent match to the simulation-based (true) mean squared prediction errors. First consider (21), the unbiased, empirical-Bayes version of the optimal predictor; its mean squared prediction error is

$$E(Z(B) - \check{Z}(B))^2 = \int_B \int_B \operatorname{cov}(Z(\mathbf{u}) - \check{Z}(\mathbf{u}), Z(\mathbf{v}) - \check{Z}(\mathbf{v})) d\mathbf{u} d\mathbf{v}/|B|^2, \qquad (24)$$

where $\check{Z}(\cdot)$ is given by (20). From (4), (5), (19), and (20), we establish that the *integrand* of (24) is given by

$$(\exp\{\mu_Y + (1/2)C_Y(\mathbf{u}, \mathbf{u})\})(\exp\{\mu_Y + (1/2)C_Y(\mathbf{v}, \mathbf{v})\})(a - b - c + d), \qquad (25)$$

where

$$\begin{aligned} a &= \exp\{C_Y(\mathbf{u}, \mathbf{v})\} \\ b &= \exp\{(\mathbf{c}_Y(\mathbf{u}) + \mathbf{1}m(\mathbf{u}))' \Sigma_Y^{-1} \mathbf{c}_Y(\mathbf{v})\} \\ c &= \exp\{(\mathbf{c}_Y(\mathbf{v}) + \mathbf{1}m(\mathbf{v}))' \Sigma_Y^{-1} \mathbf{c}_Y(\mathbf{u})\} \\ d &= \exp\{(\mathbf{c}_Y(\mathbf{u}) + \mathbf{1}m(\mathbf{u}))' \Sigma_Y^{-1} (\mathbf{c}_Y(\mathbf{v}) + \mathbf{1}m(\mathbf{v}))\}\,. \end{aligned}$$

In practice, the integrals in (24) are approximated with finite summations.

Now consider (23), the (approximately) unbiased, empirical-Bayes version of the predictor based on the permanence assumption; its mean squared prediction error is

$$E(Z(B) - \widetilde{Z}(B))^2 \simeq \int_B \int_B \operatorname{cov}(\exp\{Y(\mathbf{u})\}, \exp\{Y(\mathbf{v})\}) d\mathbf{u} d\mathbf{v}/|B|^2$$

$$+ \operatorname{var}(\widetilde{Z}(B)) - 2\int_B \operatorname{cov}(\exp\{Y(\mathbf{u})\}, \widetilde{Z}(B))d\mathbf{u}/|B|$$
$$\equiv \left(\int_B\int_B f\, d\mathbf{u}d\mathbf{v}/|B|^2\right) + g - 2\left(\int_B h\, d\mathbf{u}/|B|\right), \quad (26)$$

where the approximation is due to the predictor having approximately zero bias. Now (23) is based on a bias adjustment of the ordinary-kriging predictor on the log scale; it is of the form $\exp\{\widehat{Y}(B) + \widetilde{k}\}$, where

$$\widetilde{k} \equiv (1/2)\int_B C_Y(\mathbf{u}, \mathbf{u})d\mathbf{u}/|B| - (1/2)\int_B\int_B C_Y(\mathbf{u}, \mathbf{v})d\mathbf{u}d\mathbf{v}/|B|^2$$
$$+ (1/2)\sigma^2_{Y,k}(B) - m(B)\,. \quad (27)$$

Thus, from (26), (27), and properties of the lognormal distribution, we obtain:

$$f = (\exp\{\mu_Y + (1/2)C_Y(\mathbf{u}, \mathbf{u})\})(\exp\{\mu_Y + (1/2)C_Y(\mathbf{v}, \mathbf{v})\})(\exp\{C_Y(\mathbf{u}, \mathbf{v})\} - 1)$$
$$g = (\exp\{2\widetilde{k}\})(\exp\{2\mu_Y + \boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\})(\exp\{\boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\} - 1)$$
$$h = (\exp\{\widetilde{k}\})(\exp\{\mu_Y + (1/2)C_Y(\mathbf{u}, \mathbf{u})\})(\exp\{\mu_Y + (1/2)\boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\})$$
$$\times (\exp\{\boldsymbol{\lambda}(B)'\mathbf{c}_Y(\mathbf{u})\} - 1)\,.$$

In practice, the integrals in (26) are approximated with finite summations. Computationally, it is not more or less intensive than the finite-summation approximation to (24).

It is difficult to compare (24) and (26) analytically. For the spatial models used in the simulation in the fourth section, we computed their values and compared them to each other (as well as to their simulation-based versions). In the fourth section, it is seen that (24) is generally smaller, leading us to choose the lognormal block kriging predictor $\check{Z}(B)$ given by (21).

BLOCK PREDICTION OF PHOSPHORUS: BROOM'S BARN FARM, UK

In the late 1950s, Broom's Barn Farm covered 77 ha near Bury St. Edmunds, Suffolk. Its crop was sugar beet, but yields were small. A soil-nutrient survey was carried out by J. A. P. Marsh and S. N. Adams to find ways to improve and even out the variability of the soil properties on the farm. Webster and McBratney (1987) describe how the data on pH, exchangeable potassium, and available phosphorus were taken, which resulted in 435 values for pH and potassium and 433 values (two missing values) for phosphorus. Data were on a square grid with grid spacing of 40 m; see Figure 1.

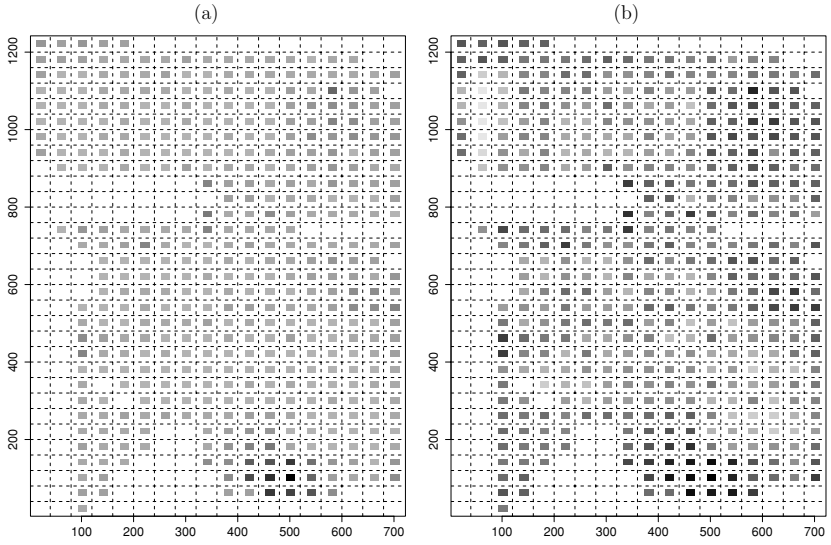

Figure 1. Broom's Barn Farm phosphorus data on a 40 m × 40 m grid. Small (large) values are colored light (dark) gray; (a) shows $\{Z(\mathbf{s}_i)\}$; and (b) shows $\{\log Z(\mathbf{s}_i)\}$.

Before we describe the geostatistical analysis and block kriging, it is worth discussing a little the possibilities of precision agriculture. In the past, most crop-production inputs (e.g., fertilizers, pesticides, water) were applied at uniform rates within a field. In recent times, agricultural practices have begun to exploit within-field heterogeneity to avoid either over-application or under-application of inputs. The result has often been higher yielding crops and more environmentally friendly fertilizer- and pesticide-application schemes. Global positioning systems (GPSs) installed on modern farm equipment can measure the application of fertilizer and insecticide down to meters, and they provide similarly precise data on crop yields. The possibility of characterizing within-field heterogeneity of a multitude of variables (e.g., here, phosphorus) has made precision agriculture an emerging area of agribusiness.

Phosphorus (P) is an important component in the development of plant reproductive parts; in particular, it is essential for seed formation. It is strongly held by soil particles so is not as easily lost as nitrogen. Soil erosion is a major factor contributing to its heterogeneity and, as a consequence, we expect it to exhibit spatial dependence.

Since the phosphorus data $\{P(\mathbf{s}_i)\}$ are highly skewed, we add 0.05 to each datum (as did Webster and McBratney, 1987):

$$Z(\mathbf{s}_i) \equiv P(\mathbf{s}_i) + 0.05; \quad i = 1, \ldots, n,$$

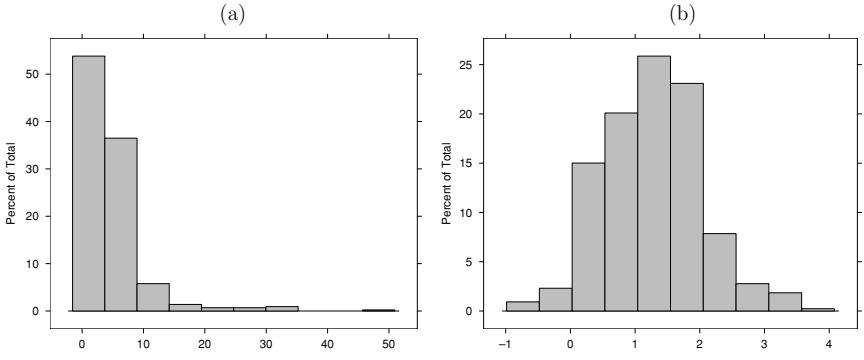

Figure 2. Histograms of 433 values of: (a) $\{Z(\mathbf{s}_i)\}$; and (b) $\{\log Z(\mathbf{s}_i)\}$.

and then we take logs (to the base e):

$$Y(\mathbf{s}_i) \equiv \log(Z(\mathbf{s}_i)); \quad i = 1, \ldots, n.$$

Figure 2 shows the histograms of $\{Z(\mathbf{s}_i)\}$ and $\{Y(\mathbf{s}_i)\}$.

Variography indicated no preferred direction of spatial dependence, nor was any rescaling necessary. Figure 3 shows the isotropic empirical semivariogram based on the data $\mathbf{Y}$ and a superimposed fitted spherical semivariogram, $\gamma_Y^{(0)}(h; \widehat{\boldsymbol{\theta}})$, where

$$\gamma_Y^{(0)}(h; \boldsymbol{\theta}) \equiv \begin{cases} 0; & h = 0 \\ \sigma_Y^2[\nu + (1-\nu)((3/2)(h/R) - (1/2)(h/R)^3)]; & 0 < h < R \\ \sigma_Y^2; & h \geq R, \end{cases}$$

$\sigma_Y^2 \geq 0$, $0 \leq \nu \leq 1$, and $R \geq 0$. The parameters $\boldsymbol{\theta} \equiv (\sigma_Y^2, \nu, R)'$ were estimated by weighted least squares (Cressie, 1985) to be $\widehat{\boldsymbol{\theta}} = (0.5717, 5.58\%, 246.81\,\text{m})'$. Note that in terms of the factors of the simulation experiment described in the fourth section, $\sigma_Y^2 \in (0.1, 0.7)$, $\nu \in (0\%, 10\%)$, and $D_R \in (\frac{1}{8}, \frac{1}{4})$.

Application of fertilizer is made in units of the spreader dimension and the equipment's response to the farmer's spreading instructions. In this application, we set the blocks of interest to start at 16 m × 16 m, whose linear dimension is roughly that of the width of a spreader. We then increased this to 24 m × 24 m, 32 m × 32 m, ..., 80 m × 80 m, to gauge the effect of block size on the block predictions. This leads to supports S_R (in units of range; see the fourth section) of 0.4, 0.6, ..., 2.

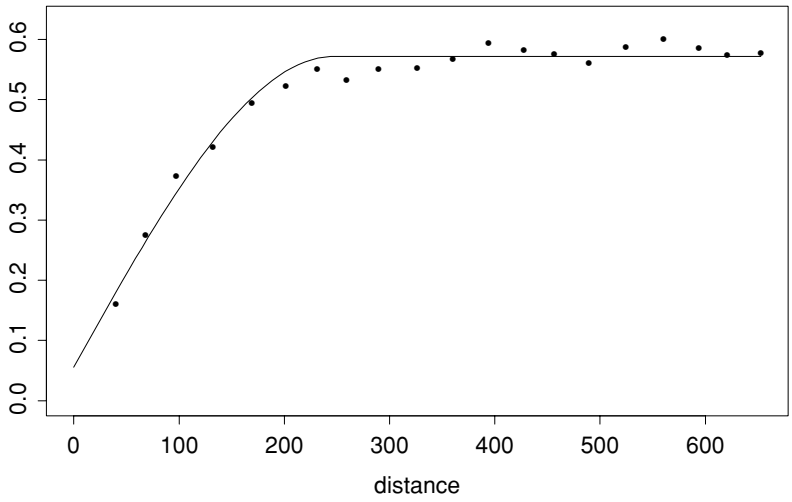

Figure 3. Empirical semivariogram estimates (dots) $\widehat{\gamma}_Y^{(0)}(h)$ as a function of lag distance h (in meters). Superimposed is the weighted-least-squares-fitted spherical semivariogram model (solid line) $\gamma_Y^{(0)}(h;\widehat{\boldsymbol{\theta}})$.

Based on the covariance function,

$$C^{(0)}(h;\widehat{\boldsymbol{\theta}}) \equiv \widehat{\sigma}_Y^2 - \gamma^{(0)}(h;\widehat{\boldsymbol{\theta}}),$$

we performed block kriging using $\check{Z}(B)$ given by (21) and $\widetilde{Z}(B)$ given by (23). The predictors for 40 m $\times$ 40 m blocks are shown in Figure 4(a) and (b). Figure 4(c) shows the difference between the two predictors and Figure 4(d) shows the efficiency ratio (24)/(26). Observe from Figure 4(d) that the efficiency of $\widetilde{Z}(B)$ is always less than or equal to 1.

To see the effect of block support on the efficiency ratio, we chose nonoverlapping blocks of given sizes within the region of interest and plotted the ratio as a function of block size (support). The result is given in Figure 5 and provides a graphic illustration of the worsening efficiency of $\widetilde{Z}(B)$ as block size increases.

Our conclusion from this case study is that the predictors $\check{Z}(B)$ and $\widetilde{Z}(B)$ are often quite similar, but their mean squared prediction errors, $\check{M}SPE$ and $\widetilde{M}SPE$, can be quite different. The superiority of $\check{Z}(B)$, as quantified by its smaller $\check{M}SPE$, is apparent. In the next section, we use simulation to arrive at the same conclusion, and to show that the theoretical quantities $\check{M}SPE$ and $\widetilde{M}SPE$ match their empirical (true) versions very closely.

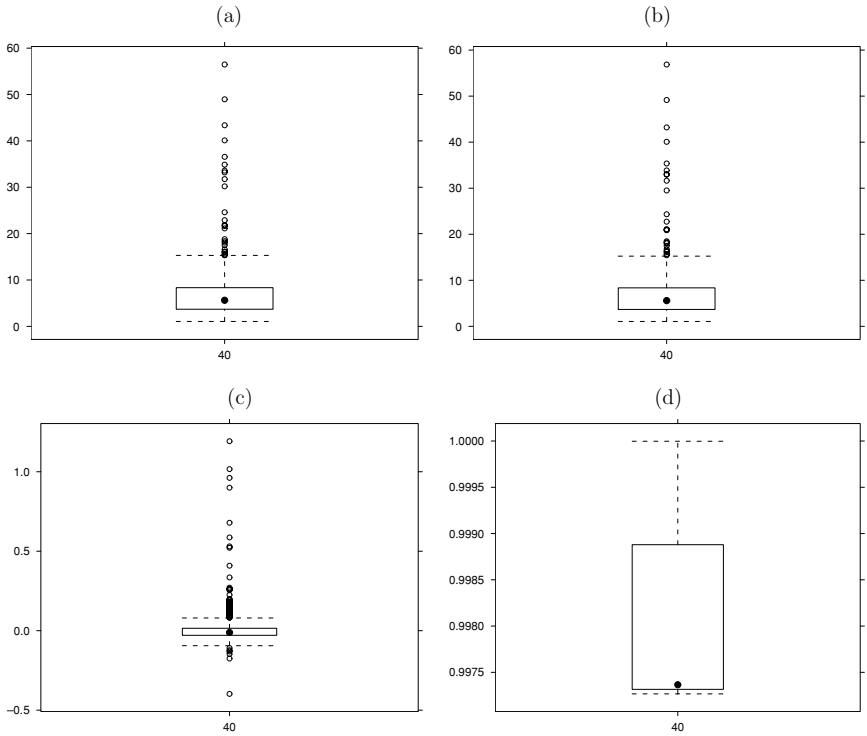

Figure 4. Boxplots of lognormal kriging predictors: (a) $\check{Z}(B)$; and (b) $\widetilde{Z}(B)$, where B is a 40 m $\times$ 40 m block varying over the Broom's Barn Farm region illustrated in Figure 1; (c) shows the differences, $\check{Z}(B) - \widetilde{Z}(B)$; (d) shows the efficiencies of $\widetilde{Z}(B)$ relative to $\check{Z}(B)$, as measured by the ratio (24)/(26).

BLOCK PREDICTION: A SIMULATION EXPERIMENT

A simulation experiment was conducted in order to compare the two lognormal kriging predictors (21) and (23); we use the terminology of experimental design, implementing a factorial design with a number of factors and varying levels within factors. We expect (21) to have better performance than (23), since it is developed from the optimal predictor, but a quantitative verification is important. And the question remains how close the two predictors are in practice, since (23) has been used a lot in past applications of lognormal kriging.

We generated a spatially dependent Gaussian process $Y(\cdot)$ on a 32×32 square with 33×33 nodes, each one unit apart; see Figure 6. The Gaussian process had $\mu_Y = 0$, and an isotropic covariance function $C_Y(\mathbf{u}, \mathbf{v}) = C_Y^{(0)}(\|\mathbf{u} - \mathbf{v}\|) \geq 0$, given by the spatial moving average described in Cressie and Pavlicova (2002). Different parameters were varied.

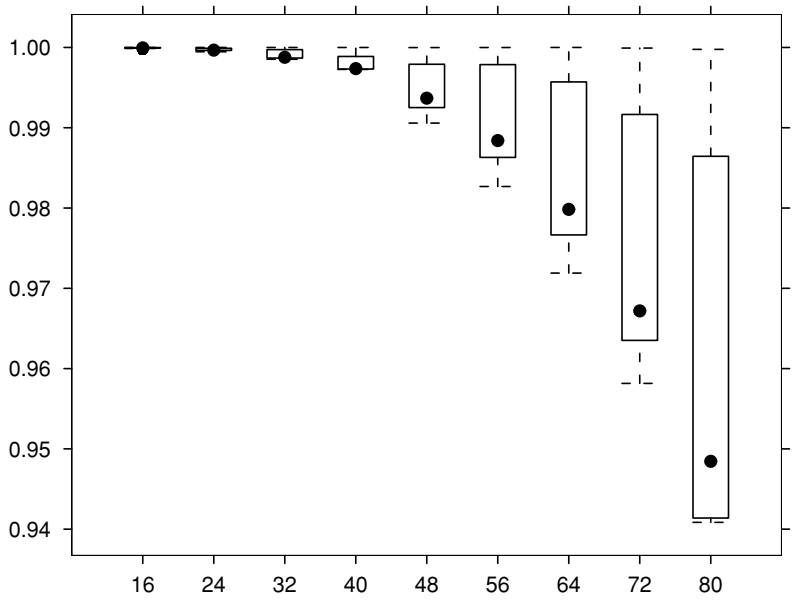

Figure 5. Boxplots of the efficiencies (24)/(26) of $\widetilde{Z}(B)$ relative to $\check{Z}(B)$ as a function of B, where $B \in \{16 \times 16, \ldots, 80 \times 80\}$ varies over the Broom's Barn Farm region. Units on the horizontal axis are in the linear dimension of B.

- Sill: $C_Y^{(0)}(0) \equiv \sigma_Y^2 \in \{0.1, 0.7, 2.5\}$.
- Nugget effect: $\lim_{h \to 0}\{C_Y^{(0)}(0) - C_Y^{(0)}(h)\}/C_Y^{(0)}(0) \equiv \nu \in \{0\%, 10\%, 30\%, 50\%\}$.
- Range: $R \equiv \arg\inf\{h: C_Y(h') = 0, h' \geq h\} \in \{8, 16, 32, 64\}$.

The sill values were chosen to give a representative range of coefficients of variation, $\mathrm{CV} = (\exp\{\sigma_Y^2\} - 1)^{1/2} \in \{0.32, 1.01, 3.34\}$. The nugget effect is anywhere up to 50% of the sill, and the linear dimension of the 32×32 square is anywhere between four times (weak spatial dependence) and half (very strong spatial dependence) the range.

Although observations on $Y(\cdot)$ were simulated at each grid node $\{\mathbf{u}_i : i = 1, \ldots, 33 \times 33\}$, only a subset $\{\mathbf{s}_i : i = 1, \ldots, n\}$ were used to generate data for the experiment.

- Data: $Z(\mathbf{s}_i) \equiv \exp\{Y(\mathbf{s}_i)\}$; $i = 1, \ldots, n$, where $n \in \{4, 25, 81\}$; the data were nested according to Figure 6. Notice that for $n = 4$, the data are 16 units apart, for $n = 25$, the data are 8 units apart and for $n = 81$, the data are 4 units apart.

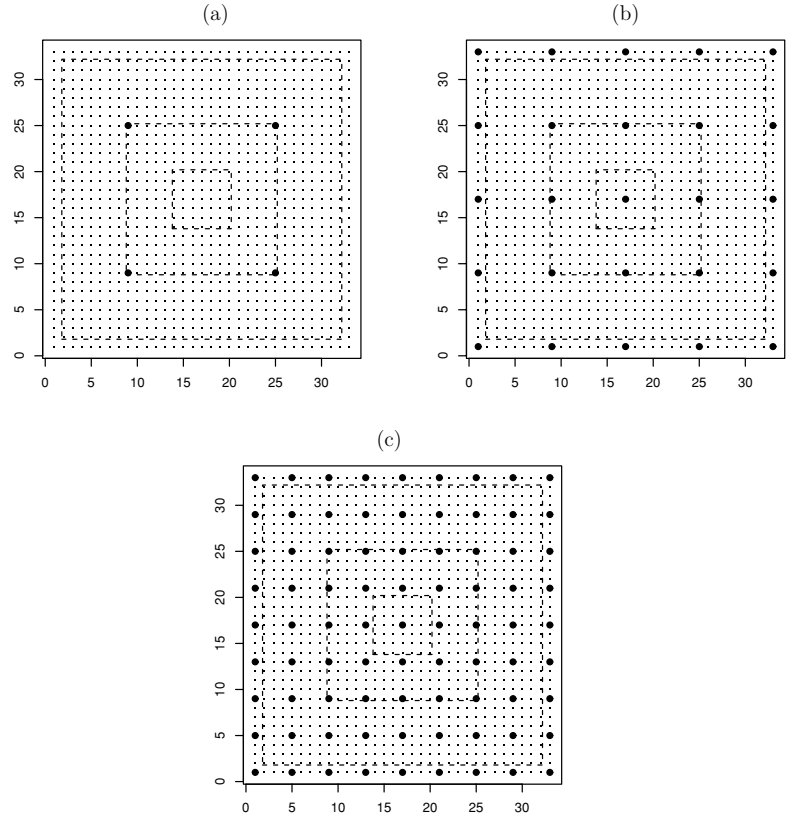

Figure 6. Spatial configuration of the simulation experiment. Small dots show the simulation grid, large dots show the data locations and dashed lines show the small (6×6), medium (16×16), and large (30×30) blocks.

Lognormal kriging is carried out on blocks of varying supports since the permanence assumption is likely to be better for smaller blocks.

- Support: Predict $Z(B)$ on blocks $B \in \{2 \times 2, 4 \times 4, 6 \times 6, \ldots, 32 \times 32\}$, centered on the center of the 33×33 grid $\{\mathbf{u}_i\}$ and nested. Small support (6×6), medium support (16×16), and large support (30×30) are featured.

Finally, the responses of the factorial experiment are based on the two fundamental properties of a predictor: bias and mean squared prediction error. The responses are then used to compare (21) and (23) at the specified levels of the factors: Sill, Nugget effect, Range, Data, and Support. Consider a generic predictor $Z^{\dagger}(B)$. Let

$Z^{(\ell)}(\cdot)$ denote the ℓth simulation of the log Gaussian process with specified σ_Y^2, ν, and R; $\ell = 1, \ldots, L$, and $Z^{(\ell)\dagger}(B)$ is the generic predictor of $Z^{(\ell)}(B)$. Then the (estimated) bias of the generic predictor $Z^{\dagger}(B)$ is:

$$\text{BIAS} \equiv (1/L) \sum_{\ell=1}^{L} \{Z^{(\ell)\dagger}(B) - Z^{(\ell)}(B)\}, \tag{28}$$

and the (estimated) mean squared prediction error of $Z^{\dagger}(B)$ is:

$$\text{MSPE} \equiv (1/L) \sum_{\ell=1}^{L} \{Z^{(\ell)\dagger}(B) - Z^{(\ell)}(B)\}^2, \tag{29}$$

where any integrals in $Z^{(\ell)\dagger}(B)$ or $Z^{(\ell)}(B)$ are approximated as sums based on the finest grid spacing. The value $L = 6400$ was chosen to guarantee accuracy of results to the second decimal place, where that digit is conservatively plus or minus 2; see Aldworth and Cressie (1999) for the relevant calculations that determine L.

When $Z^{\dagger}(B)$ is $\check{Z}(B)$ given by (21), theory tells us that the BIAS should be zero and the MSPE is given by (24). When $Z^{\dagger}(B)$ is $\widetilde{Z}(B)$ given by (23) and based on the permanence assumption, theory tells us the BIAS is *approximately* zero and the MSPE is *approximately* given by (26). Thus, the simulation experiment not only allows direct comparison of the true moments of $\check{Z}(B)$ and $\widetilde{Z}(B)$, it also tells us when the permanence assumption is appropriate.

A basic analysis of the data coming from the simulation experiment allows us to conclude:

- $\widetilde{Z}(B)$ is approximately unbiased for all B; see Figure 7. The slight negative bias for some of the factor combinations and the 6×6 block is due entirely to the large-sill case of $\sigma_Y^2 = 2.5$.
- The first and second theoretical moments of both $\check{Z}(B)$ and $\widetilde{Z}(B)$ are approximately equal to their true moments [estimated from the simulation according to (28) and (29)]. Hence, their theoretical and true mean squared prediction errors are approximately equal; see Figure 8.
- The quantity $(\text{BIAS})^2$ obtained from (28) is only a fraction (no more than 0.20%) of MSPE obtained from (29). This is true for both $\check{Z}(B)$ and $\widetilde{Z}(B)$; see Figure 9.
- The MSPE for $\check{Z}(B)$ is smaller than (and sometimes equal to) the MSPE for $\widetilde{Z}(B)$. That is, the spatial predictor $\check{Z}(B)$ is dominant over $\widetilde{Z}(B)$.

It is this latter result that we would like to explore in greater depth, since the improvement in efficiency obtained by using $\check{Z}(B)$ is not uniform over all

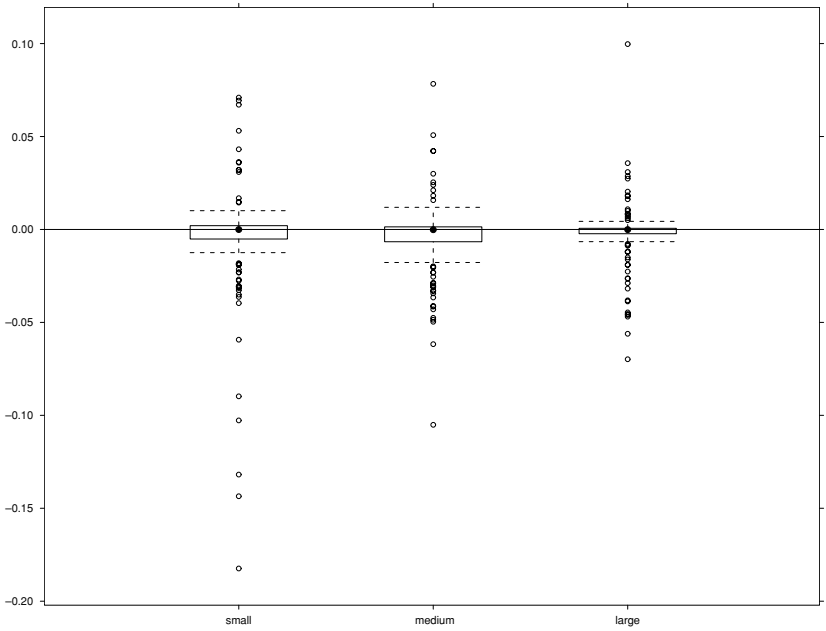

Figure 7. Boxplots of bias of $\widetilde{Z}(B)$, where B is small (6×6), medium (16×16), and large (30×30). The box plots are formed from the various factor-level combinations.

combinations of the factors of the simulation experiment. The efficiency of $\widetilde{Z}(B)$ relative to $\check{Z}(B)$ is defined as

$$E \equiv \check{M}SPE / \widetilde{M}SPE\,, \tag{30}$$

where $\check{M}SPE$ ($\widetilde{M}SPE$) is given by (29) with $Z^\dagger \equiv \check{Z}$ ($Z^\dagger \equiv \widetilde{Z}$). We now analyze these efficiencies according to different combinations of the factors of the simulation experiment.

Some preliminary analysis showed it was useful to modify certain factors and their levels:

- Data: Now measure this factor in terms of the smallest distance between two data in units of the range R. Denote it by D_R, and hence $D_R \in \{\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2\}$.
- Support: Now measure this factor in terms of the side of the square block in units of the range R. Denote it by S_R, and hence $S_R \in \{\frac{6}{64}, \frac{6}{32}, \frac{1}{4}, \frac{6}{16}, \frac{30}{64}, \frac{1}{2}, \frac{6}{8}, \frac{30}{32}, 1, \frac{30}{16}, 2, \frac{30}{8}, \}$.

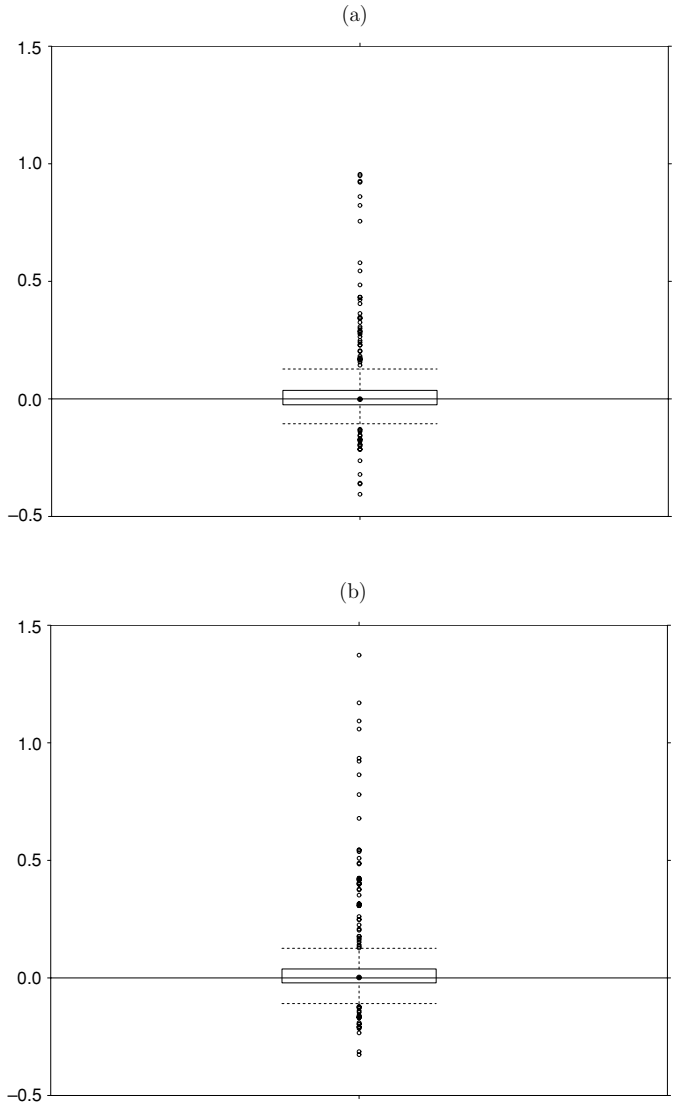

Figure 8. Boxplots of: (a) $[(24)/\check{M}SPE] - 1$; and (b) $[(26)/\widetilde{M}SPE] - 1$. The boxplots are formed from all factor-level combinations.

Figure 10(a) shows E for all factor combinations, and the dominance of $\check{Z}(B)$ over $\widetilde{Z}(B)$ is striking. Figure 10(b) shows that the efficiency of $\widetilde{Z}(B)$ decreases as the sill σ_Y^2 increases, showing that the more skewed the lognormal distribution, the greater the potential gains in efficiency using $\check{Z}(B)$. Figure 10(c)

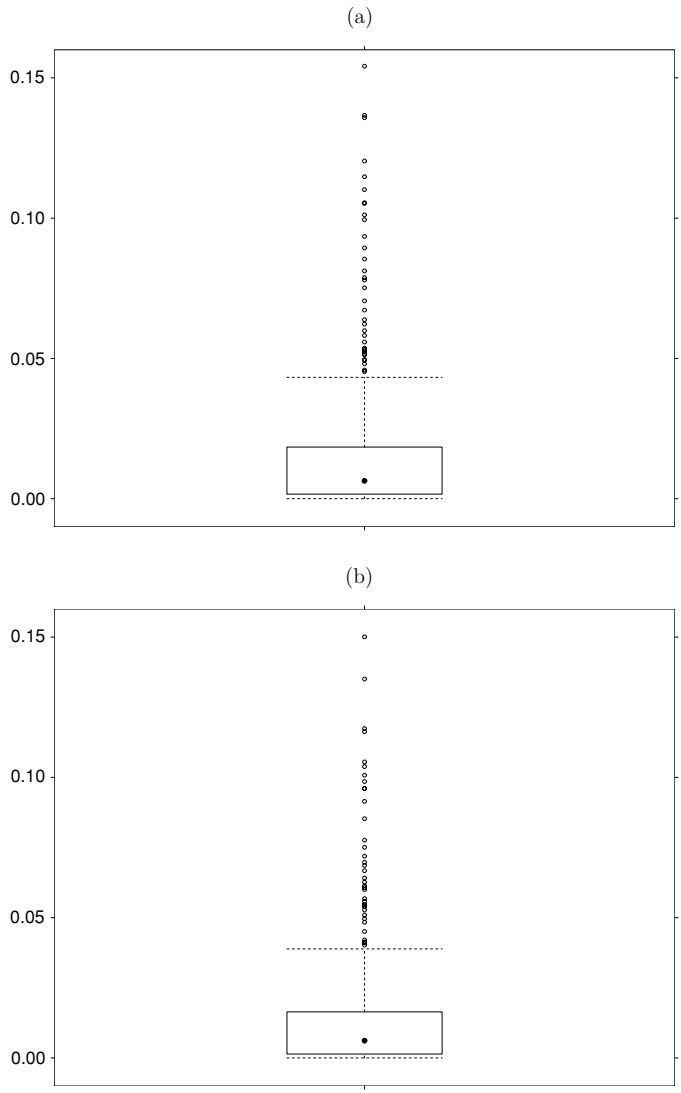

Figure 9. Boxplots of $(BIAS)^2/MSPE$ for: (a) $\check{Z}(B)$; and (b) $\widetilde{Z}(B)$. The boxplots are formed from all factor-level combinations.

shows that the efficiency of $\widetilde{Z}(B)$ increases as the nugget effect ν increases. That is, as the spatial dependence gets weaker, $\check{Z}(B)$ is not as dominant over $\widetilde{Z}(B)$. A plot of E broken down by range (not shown here) reinforces this observation; for small R (weaker spatial dependence), $\check{Z}(B)$ is not as dominant over $\widetilde{Z}(B)$. From Figure 10(d), we see that $\check{Z}(B)$ dominates over $\widetilde{Z}(B)$ when data are

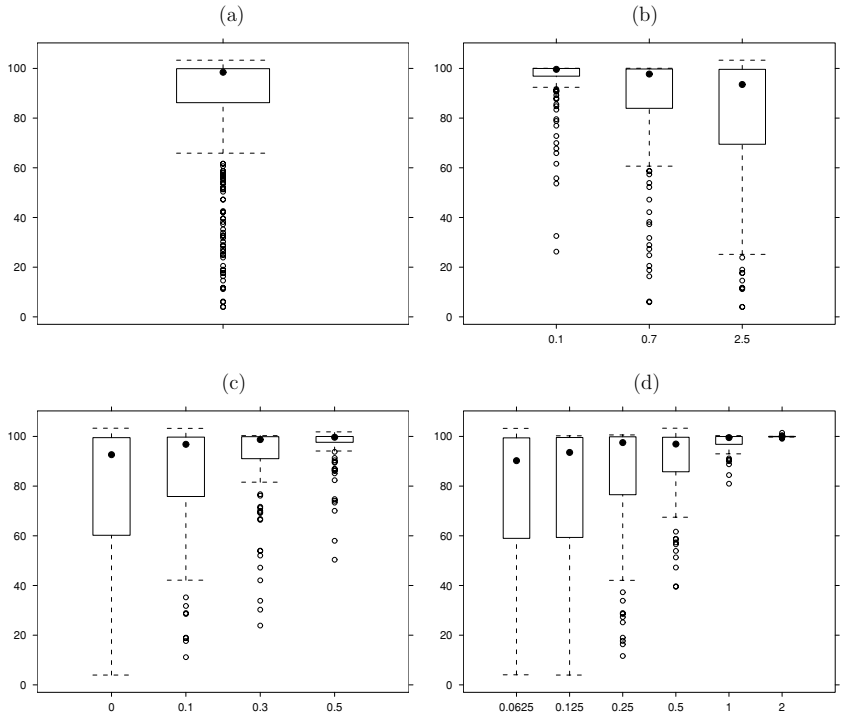

Figure 10. Boxplots of efficiencies E, of $\widetilde{Z}(B)$ relative to $\check{Z}(B)$: (a) boxplot is formed as in Figure 8; (b) boxplots are a function of the sill σ_Y^2 and are formed from all remaining factor-level combinations; (c) boxplots are a function of the nugget effect ν and are formed from all remaining factor-level combinations; (d) boxplots are a function of the smallest distance between data D_R and are formed from all remaining factor-level combinations.

closer together (in units of range) and is less dominant when they are far apart. That is, the more nearby the spatial data are, the better $\check{Z}(B)$ is able to use that information.

The data and block factors, as measured by D_R and S_R, respectively, show interesting interaction. What looks like no pattern at all in Figure 11(a), in terms of S_R, takes on remarkable structure when it is further broken down by D_R. Figure 11(b) shows that for a given D_R, E decreases as S_R increases. That is, it is on large supports where $\check{Z}(B)$ is particularly dominant over $\widetilde{Z}(B)$.

Finally, it is interesting to look at how E varies as a function of block size, $|B|^{1/2}$; $B \in \{2 \times 2, 4 \times 4, 6 \times 6, \ldots, 32 \times 32\}$. Figure 12 shows four such plots for four different factor-level combinations where D_R is held fixed at its smallest value of $1/16$ (i.e., $R = 64$ and $n = 81$). All plots show a decrease in efficiency as block size increases, as expected from other results in this section. The biggest

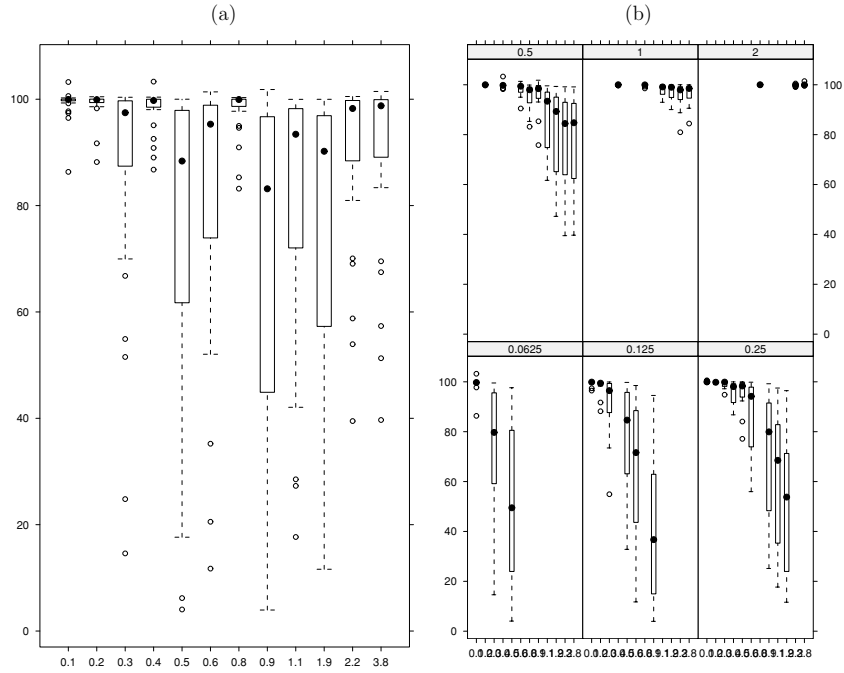

Figure 11. Boxplots of efficiencies E, of $\widetilde{Z}(B)$ relative to $\check{Z}(B)$: (a) boxplots are a function of the support (block size) S_R and are formed from all remaining factor-level combinations; (b) boxplots are broken down by the smallest distance between data D_R; they are functions of the support S_R and are formed from all remaining factor-level combinations.

drop in efficiency occurs when ν is 0% (high spatial dependence) and σ_Y^2 is 2.5 (large skewness).

DISCUSSION AND CONCLUSIONS

Data that exhibit non-normality could be handled by a normal-score transform, but this approach does not account for the randomness inherent in the transform. Data that exhibit skewness may be successfully modeled as coming from a lognormal spatial process; such data are common in soil science (e.g., Webster, 2001). Under lognormality and a known mean and covariance function, an optimal spatial predictor can be derived that predicts at both point and block supports. The optimal predictor $\check{Z}(B)$ at block support B requires quadrature of the optimal predictor at point support, something that is trivial to do in current computing environments. When the mean function is unknown, the problem is easily solved by invoking the principle that efficient mean estimation yields ef-

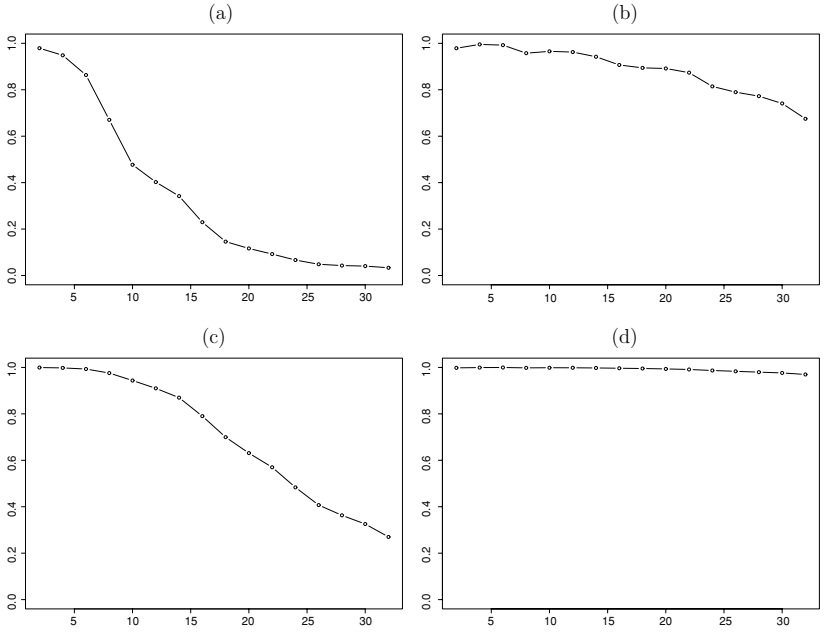

Figure 12. Plot of efficiency E as a function of the block size $|B|^{1/2}$, where $R = 64$ and $n = 81$: (a) $\nu = 0\%$ and $\sigma_Y^2 = 2.5$; (b) $\nu = 50\%$ and $\sigma_Y^2 = 2.5$; (c) $\nu = 0\%$ and $\sigma_Y^2 = 0.1$; (d) $\nu = 50\%$ and $\sigma_Y^2 = 0.1$.

ficient spatial prediction. The efficiency of $\check{Z}(B)$ can be quantified by its mean squared prediction error, for which a formula is given. This involves quadrature, but again it is trivial to evaluate in current computing environments.

Lognormal kriging has traditionally used another principle, that of permanence of lognormality. Mathematically, permanence is impossible, but it can lead to simple predictors. The permanence-based predictor $\widetilde{Z}(B)$ is only approximately unbiased and the formula for its mean squared prediction error is only approximately correct, in contrast to the optimality-based predictor $\check{Z}(B)$. The question that this paper addresses is which predictor is the most efficient.

Although formulae are given for both constant and non-constant mean, the case of constant mean is featured in the example and the simulation experiment. Further experiments based on non-constant mean are needed to fully complement the empirical results obtained here. Because the predictor $\check{Z}(B)$ is based on optimality and efficient parameter estimation, we expect the results to remain unchanged.

Through simulation and practical application, it is shown that the optimality-based predictor $\check{Z}(B)$ is superior, particularly in situations where the spatial

dependence is strong, the lognormal data are more skewed, and the spatial data are closer together. In current computing environments, the two predictors have the same computational burden, and hence we conclude unequivocably that $\check{Z}(B)$ is the superior predictor.

ACKNOWLEDGMENTS

This research was funded in part by the Office of Naval Research under grant N00014-02-1-0052 and N00014-05-1-0133. The main theoretical results came out of early discussions with Isobel Clark. Martina Pavlicova gave invaluable assistance with the computational aspects and gave generous feedback on various drafts of the manuscript, and the referees helped clarify the presentation. Many thanks go to Richard Webster for making the Broom's Barn Farm data available and to Kathy Flanders for helpful discussions on precision agriculture.

REFERENCES

Aitchison, J., and Brown, J. A. C., 1957, The lognormal distribution: Cambridge University Press, Cambridge, 176 p.

Aldworth, J., and Cressie, N., 1999, Sampling designs and prediction methods for Gaussian spatial processes, *in* Ghosh, S., ed., Multivariate Design and Sampling: Marcel Dekker, New York, p. 1–54.

Brown, G., and Sanders, J. W., 1981, Lognormal genesis: J. Appl. Probabil., v. 18, p. 542–547.

Cressie, N., 1985, Fitting variogram models by weighted least squares: Math. Geol., v. 17, no. 5, p. 563–586.

Cressie, N., 1993, Statistics for spatial data (revised edition): Wiley, New York, 900 p.

Cressie, N., and Pavlicova, M., 2002, Calibrated spatial moving average simulations: Stat. Model., v. 2, p. 267–279.

Cressie, N., and Pavlicova, M., 2005, Permanence of lognormality: Bias adjustment and kriging variances, *in* Leuangthong, O., and Deutsch, C. V., eds., Geostatistics Banff 2004: Springer, Dordrecht, p. 1027–1036.

de Wijs, H. J., 1951, Statistics of ore distribution. I: Frequency distribution of assay values: Geologie en Mijmbouw, v. 13, p. 365–375.

Dowd, P. A., 1982, Lognormal kriging—The general case: Math. Geol., v. 14, no. 5, p. 475–499.

Journel, A. G., 1980, The lognormal approach to predicting local distributions of selective mining unit grades: Math. Geol., v. 12, no. 4, p. 285–303.

Marechal, A., 1974, Krigeage normal et lognormal. Centre de Morphologie Mathematique, Ecole des Mines de Paris, Publication N-376, 10 p.

Matheron, G., 1962, Traite de geostatistique appliquee: Memoires du Bureau de Recherche Geologiques et Minieres, no. 14, Ed. Technip, Tome 1, 333 p.

Matheron G., 1974, Effet proportionnel et lognormalite ou: Le retour du serpent de mer. Centre de Morphologie Mathematique, Ecole des Mines de Paris, Publication N-374, 43 p.

Rendu, J.-M., 1979, Normal and lognormal estimation: Math. Geol., v. 11, no. 4, p. 407–422.

Rivoirard, J., 1990, A review of lognormal estimators for in situ reserves: Math. Geol., v. 22, no. 2, p. 213–221.

Roth, C., 1988, Is lognormal kriging suitable for local estimation? Math. Geol., v. 30, no. 8, p. 999–1009.
Thorin, O., 1977, On the infinite divisibility of the lognormal distribution: Scand. Actuar. J., v. 1977, p. 121–148.
Webster, R., 2001, Statistics to support soil research and their presentation: Eur. J. Soil Sci., v. 52, 331–340.
Webster, R., and McBratney, A. B., 1987, Mapping soil fertility at Broom's Barn by simple kriging: J. Sci. Food Agric., v. 38, p. 97–115.

APPENDIX A: LOGNORMAL KRIGING IN THE PRESENCE OF MEASUREMENT ERROR: FIRST TWO NORMAL MOMENTS KNOWN

Recall that $Z(\cdot) = \exp(Y(\cdot))$, where $Y(\cdot)$ is a normal process. Assume that $Y(\cdot) = S(\cdot) + \varepsilon(\cdot)$, where $\varepsilon(\cdot)$ is a measurement-error process: zero-mean, white noise, with variance σ_ε^2. Then

$$Z(\cdot) = W(\cdot)\exp(\varepsilon(\cdot))\,,$$

where $W(\cdot) \equiv \exp(S(\cdot))$ is a lognormal process with multiplicative measurement error. Scientific interest is often in $W(\cdot)$.

To predict $W(\mathbf{s}_0)$, we minimize the mean squared predictor error,

$$E(W(\mathbf{s}_0) - p(\mathbf{Z};\mathbf{s}_0))^2\,,$$

with respect to p, yielding the optimal predictor,

$$\begin{aligned} W^*(\mathbf{s}_0) &\equiv E(W(\mathbf{s}_0)|\mathbf{Z}) \\ &= \exp\{S^*(\mathbf{s}_0) + (1/2)\mathrm{var}(S(\mathbf{s}_0)) - (1/2)\mathrm{var}(S^*(\mathbf{s}_0))\}\,, \end{aligned}$$

where $S^*(\mathbf{s}_0) = \mu_Y(\mathbf{s}_0) + \mathbf{c}_S(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y)$ and $\mathbf{c}_S(\mathbf{s}_0) \equiv (\mathrm{cov}(S(\mathbf{s}_0), S(\mathbf{s}_1)), \ldots, \mathrm{cov}(S(\mathbf{s}_0), S(\mathbf{s}_n)))'$. Hence, $\mathrm{var}(S^*(\mathbf{s}_0)) = \mathbf{c}_S(\mathbf{s}_0)'\Sigma_Y^{-1}\mathbf{c}_S(\mathbf{s}_0)$. Notice that when $\mathbf{s}_0 \notin \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, $\mathbf{c}_S(\mathbf{s}_0) = \mathbf{c}_Y(\mathbf{s}_0)$; otherwise, when $\mathbf{s}_0 = \mathbf{s}_1$ for example, $\mathrm{cov}(S(\mathbf{s}_0), S(\mathbf{s}_1)) = C_Y(\mathbf{s}_1, \mathbf{s}_1) - \sigma_\varepsilon^2$, and $\mathrm{cov}(S(\mathbf{s}_0), S(\mathbf{s}_i)) = C_Y(\mathbf{s}_0, \mathbf{s}_i)$, $i = 2, \ldots, n$. Finally, $\mathrm{var}(S(\mathbf{s}_0)) = \mathrm{var}(Y(\mathbf{s}_0)) - \sigma_\varepsilon^2$. Substitution of these values into the expression for $W^*(\mathbf{s}_0)$ yields the optimal predictor for $W(\mathbf{s}_0)$ with mean squared prediction error, $E(W(\mathbf{s}_0) - W^*(\mathbf{s}_0))^2$. The optimal predictor for $W(B) \equiv \int_B W(\mathbf{u})d\mathbf{u}/|B|$ is:

$$W^*(B) \equiv \int_B W^*(\mathbf{u})d\mathbf{u}/|B|\,,$$

with mean squared prediction error, $E(W(B) - W^*(B))^2$.

Recall that the predictor (18) is of the form $\exp\{Y^*(B) + k^@\}$, where $k^@$ is chosen to (approximately) correct for bias. In the presence of measurement error, $S(B) = Y(B)$ and $S^*(B) = Y^*(B)$, for $|B| > 0$. Now the predictor based on the permanence approximation is of the form $\exp\{S^*(B) + k'\}$, where k' is an adjustment that gives an (approximate) expectation equal to $E(W(B)) = E(\exp\{S(\mathbf{U})\})$. Using the same reasoning that led to (18), it is straightforward to show that

$$\begin{aligned} W^@(B) \equiv \exp\{S^*(B) &+ (1/2)\int_B (C_Y(\mathbf{u},\mathbf{u}) - \sigma_\varepsilon^2)d\mathbf{u}/|B| \\ &+ (1/2)\int_B (\mu_Y(\mathbf{u}) - \mu_Y(B))^2 d\mathbf{u}/|B| \\ &- (1/2)\mathbf{c}_Y(B)'\Sigma_Y^{-1}\mathbf{c}_Y(B)\}\,; \quad |B| > 0\,, \end{aligned}$$

where the only difference between this and $Z^@(B)$ in (18) is the presence of the measurement-error variance σ_ε^2 in the bias adjustment. The mean squared prediction error is, $E(W(B) - W^@(B))^2$.

APPENDIX B: LOGNORMAL KRIGING WHEN THE NORMAL MEAN IS CONSTANT AND UNKNOWN

The predictor based on the permanence approximation is,

$$\begin{aligned} \widetilde{Z}(B) &= \exp\{\widehat{Y}(B) + (1/2)\mathrm{var}(Y(\mathbf{U})) - (1/2)\mathrm{var}(\widehat{Y}(B))\} \\ &= \exp\{\widehat{Y}(B) + (1/2)\int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B| - (1/2)\boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\}\,, \end{aligned}$$

which we now show is equal to (23). Because $\widehat{Y}(B)$ is unbiased for $Y(B)$, the mean squared prediction error satisfies,

$$\begin{aligned} \sigma^2_{Y,k}(B) &= \mathrm{var}(Y(B) - \widehat{Y}(B)) \\ &= \mathrm{var}(Y(B)) + \mathrm{var}(\widehat{Y}(B)) - 2\mathrm{cov}(\widehat{Y}(B), Y(B)) \\ &= \int_B\int_B C_Y(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}/|B|^2 + \boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B) - 2\boldsymbol{\lambda}(B)'\mathbf{c}_Y(B)\,, \end{aligned}$$

where $\widehat{Y}(B) = \boldsymbol{\lambda}(B)'\mathbf{Y}$ and $\boldsymbol{\lambda}(B)$ and $m(B)$ solve the ordinary kriging equations,

$$\begin{aligned} \Sigma_Y\boldsymbol{\lambda}(B) &= \mathbf{c}_Y(B) + \mathbf{1}m(B) \\ \mathbf{1}'\boldsymbol{\lambda}(B) &= 1\,. \end{aligned}$$

Hence, $2\boldsymbol{\lambda}(B)'\mathbf{c}_Y(B) = 2\boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B) - 2m(B)$, which leads to:

$$\sigma^2_{Y,k}(B) = \int_B\int_B C_Y(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}/|B|^2 - \boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B) + 2m(B)\,.$$

Substituting this relation into the expression for $\widetilde{Z}(B)$ yields (23).

In the case where there is measurement error, we wish to predict $W(B) = \int_B \exp\{S(\mathbf{u})\}d\mathbf{u}/|B|$, where $Z(\cdot) = \exp\{S(\cdot) + \varepsilon(\cdot)\}$; see Appendix A. Then the predictor based on the permanence approximation is, $\widetilde{W}(B) \equiv \exp\{\widehat{S}(B) + k''\}$, where k'' is an adjustment that gives an (approximate) expectation equal to $E(W(B)) = E(\exp\{S(\mathbf{U})\})$. We use the same derivations as those leading to (23) and note that $S(B) = Y(B)$ and $\widehat{S}(B) = \widehat{Y}(B)$, for $|B| > 0$. Then

$$\begin{aligned}\widetilde{W}(B) = \exp\{\widehat{S}(B) &+ (1/2)\int_B C_Y(\mathbf{u},\mathbf{u})d\mathbf{u}/|B| - (1/2)\sigma^2_\varepsilon \\ &- (1/2)\int_B\int_B C_Y(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}/|B|^2 \\ &+ (1/2)\sigma^2_{Y,k}(B) - m(B)\}\,; \quad |B| > 0\,,\end{aligned}$$

where the only difference between this and $\widetilde{Z}(B)$ in (23) is the presence of the measurement-error variance σ^2_ε in the bias adjustment. The mean squared prediction error is, $E(W(B) - \widetilde{W}(B))^2$.

APPENDIX C: LOGNORMAL KRIGING WHEN THE NORMAL MEAN IS A LINEAR REGRESSION WITH UNKNOWN REGRESSION COEFFICIENTS

Suppose that $\mu_Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$, a linear regression on known explanatory variables, $\mathbf{x}(\mathbf{s}) \equiv (x_1(\mathbf{s}), \ldots, x_p(\mathbf{s}))'$. The only case of interest is $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_p)'$ unknown; when $\boldsymbol{\beta}$ is known, we can use the predictors (14) and (18) in section "Prediction when Parameters are Known".

When $\boldsymbol{\beta}$ is unknown, ordinary kriging in section "Prediction when only Covariance Parameters are Known" is replaced with universal kriging. Then (19) is replaced with

$$\widehat{Y}(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\widehat{\boldsymbol{\beta}} + \mathbf{c}_Y(\mathbf{s}_0)'\Sigma_Y^{-1}(\mathbf{Y} - X\widehat{\boldsymbol{\beta}})\,,$$

where $\widehat{\boldsymbol{\beta}} \equiv (X'\Sigma_Y^{-1}X)^{-1}X'\Sigma_Y^{-1}\mathbf{Y}$ is the generalized least squares estimator of $\boldsymbol{\beta}$, and X is the $(n \times p)$ matrix $(\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$. Furthermore, (20) is replaced with

$$\check{Z}(\mathbf{s}_0) = \exp\{\widehat{Y}(\mathbf{s}_0) + (1/2)\sigma^2_{Y,k}(\mathbf{s}_0) - \mathbf{m}(\mathbf{s}_0)'\mathbf{x}(\mathbf{s}_0)\}\,,$$

where $\widehat{Y}(\mathbf{s}_0) \equiv \boldsymbol{\lambda}(\mathbf{s}_0)'\mathbf{Y}$ is the universal kriging predictor, $\boldsymbol{\lambda}(\mathbf{s}_0)$ and $\mathbf{m}(\mathbf{s}_0)$ solve

$$\begin{aligned}\Sigma_Y \boldsymbol{\lambda}(\mathbf{s}_0) &= \mathbf{c}_Y(\mathbf{s}_0) + X\mathbf{m}(\mathbf{s}_0)\\ X'\boldsymbol{\lambda}(\mathbf{s}_0) &= \mathbf{x}(\mathbf{s}_0)\,,\end{aligned}$$

and the kriging variance is

$$\sigma^2_{Y,k}(\mathbf{s}_0) = C_Y(\mathbf{s}_0, \mathbf{s}_0) - \boldsymbol{\lambda}(\mathbf{s}_0)'\mathbf{c}_Y(\mathbf{s}_0) + \mathbf{m}(\mathbf{s}_0)'\mathbf{x}(\mathbf{s}_0)\,.$$

Finally, the empirical-Bayes version of the optimal predictor,

$$\check{Z}(B) \equiv \int_B \check{Z}(\mathbf{u}) d\mathbf{u}/|B|\,,$$

is calculated along with its mean squared prediction error, $E(Z(B) - \check{Z}(B))^2$.

In a like manner, we derive an empirical-Bayes version of the predictor based on the permanence approximation, $\widetilde{Z}(B) \equiv \exp\{\widehat{Y}(B) + \widetilde{k}\}$, where $\widetilde{k}$ is chosen to (approximately) correct for bias. The universal kriging predictor is $\widehat{Y}(B) \equiv \boldsymbol{\lambda}(B)'\mathbf{Y}$, where $\boldsymbol{\lambda}(B)$ and $\mathbf{m}(B)$ solve

$$\begin{aligned}\Sigma_Y \boldsymbol{\lambda}(B) &= \mathbf{c}_Y(B) + X\mathbf{m}(B)\\ X'\boldsymbol{\lambda}(B) &= \mathbf{x}(B)\,,\end{aligned}$$

with $\mathbf{x}(B) \equiv (\int_B x_1(\mathbf{u})d\mathbf{u}/|B|, \ldots, \int_B x_p(\mathbf{u})d\mathbf{u}/|B|)'$. The predictor is the empirical-Bayes version of (18) and is given by,

$$\begin{aligned}\widetilde{Z}(B) \equiv \exp\{\widehat{Y}(B) &+ (1/2)\int_B C_Y(\mathbf{u}, \mathbf{u}) d\mathbf{u}/|B| - (1/2)\boldsymbol{\lambda}(B)'\Sigma_Y\boldsymbol{\lambda}(B)\\ &+ (1/2)\boldsymbol{\beta}' M(B)\boldsymbol{\beta}\}\,,\end{aligned}$$

where $M(B) \equiv \int_B (\mathbf{x}(\mathbf{u}) - \mathbf{x}(B))(\mathbf{x}(\mathbf{u}) - x(B))' d\mathbf{u}/|B|$ is a $(p \times p)$ matrix. Then, in a like manner to the derivation in Appendix B, we obtain,

$$\begin{aligned}\widetilde{Z}(B) = \exp\{\widehat{Y}(B) &+ (1/2)\int_B C_Y(\mathbf{u}, \mathbf{u}) d\mathbf{u}/|B| - (1/2)\int_B\int_B C_Y(\mathbf{u}, \mathbf{v}) d\mathbf{u}d\mathbf{v}/|B|^2\\ &+ (1/2)\sigma^2_{Y,k}(B) - \mathbf{m}(B)'\mathbf{x}(B) + \boldsymbol{\beta}' M(B)\boldsymbol{\beta}\}\,,\end{aligned}$$

where the kriging variance is

$$\sigma^2_{Y,k}(B) = \int_B \int_B C_Y(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} / |B|^2 - \boldsymbol{\lambda}(B)' \mathbf{c}_Y(B) + \mathbf{m}(B)' \mathbf{x}(B) .$$

However, there is a problem with predictor $\widetilde{Z}(B)$ given just above: it depends on $\boldsymbol{\beta}$ unless $M(B)$ is the zero matrix. Now if $\mathbf{x}(\mathbf{s}) \equiv 1$, $M(B) = 0$, which is the case we presented in Appendix B. Otherwise, $M(B)$ is nonzero. The generalized least squares estimator $\widehat{\boldsymbol{\beta}}$ could be substituted in for $\boldsymbol{\beta}$, but that would lead to a deterioration in bias and an increase in mean squared prediction error. In contrast, $\check{Z}(B)$ does not depend on $\boldsymbol{\beta}$ and is the empirical-Bayes version of the optimal predictor.