Inference from spatial processes

Carol Anne Gotway

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Gotway, Carol Anne, "Inference from spatial processes " (1989). Retrospective Theses and Dissertations. 9041. https://lib.dr.iastate.edu/rtd/9041

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17” x 23” black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
Inference from spatial processes

Gotway, Carol Anne, Ph.D.

Iowa State University, 1989
Inference from spatial processes

by

Carol Anne Gotway

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

in Charge of Major Work

For the Major Department

For the Graduate College

Iowa State University
Ames, Iowa

1989
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>SECTION I. IMPROVED PREDICTION WITH THE GENERALIZED LINEAR MODEL</td>
<td>9</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>11</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>12</td>
</tr>
<tr>
<td>2. LINEAR PREDICTION</td>
<td>16</td>
</tr>
<tr>
<td>2.1. Known mean</td>
<td>16</td>
</tr>
<tr>
<td>2.2. Unknown mean</td>
<td>19</td>
</tr>
<tr>
<td>3. NONLINEAR PREDICTION</td>
<td>25</td>
</tr>
<tr>
<td>4. A GENERAL CLASS OF PREDICTORS</td>
<td>27</td>
</tr>
<tr>
<td>4.1. Decomposition of risk</td>
<td>28</td>
</tr>
<tr>
<td>4.2. Improved prediction</td>
<td>32</td>
</tr>
<tr>
<td>5. SUMMARY</td>
<td>43</td>
</tr>
<tr>
<td>6. ACKNOWLEDGMENT</td>
<td>44</td>
</tr>
<tr>
<td>7. REFERENCES CITED</td>
<td>45</td>
</tr>
<tr>
<td>SECTION II. A SPATIAL ANALYSIS OF VARIANCE</td>
<td>48</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>50</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>51</td>
</tr>
<tr>
<td>2. GEOSTATISTICS AND KRIGING</td>
<td>52</td>
</tr>
<tr>
<td>3. ESTIMATING SPATIAL DEPENDENCE: VARIOGRAM ESTIMATION AND MODEL FITTING</td>
<td>56</td>
</tr>
<tr>
<td>4. THE USE OF KRIGING AND CROSS-VALIDATION IN OUTLIER DETECTION AND DATA EDITING</td>
<td>64</td>
</tr>
</tbody>
</table>
### 5. INFERENCE ON MEAN EFFECTS

5.1. Estimation of main effects 69

5.2. A spatial analysis of variance 72

5.3. Testing the hypothesis of equality among columns within treatments 74

5.4. Testing the hypothesis of equality of average treatment effects 75

5.5. Pairwise contrasts 77

### 6. SUMMARY

Page 80

### 7. ACKNOWLEDGMENT

Page 81

### 8. REFERENCES CITED

Page 82

#### SUMMARY

Page 104

#### LITERATURE CITED

Page 106

#### ACKNOWLEDGMENTS

Page 109
GENERAL INTRODUCTION

Nearly everyone is familiar with the notion of a time series: a sequence of observations that vary with time. However, there are many naturally occurring phenomena that also vary in space. Examples include the amount and grade of ore in an ore body, or the potentiometric surface that influences the direction of groundwater flow. Consequently, an understanding of such processes must include an understanding of their spatial variability. The theory of regionalized variables (Matheron 1971) is a sophisticated and elegant interpretation of spatial processes in probabilistic terms.

Consider the stochastic process

$$\{Z(s): s \in D \subseteq \mathbb{R}^d\}, \quad (1)$$

where $D$ is a fixed open set in $d$-dimensional Euclidean space. Such a stochastic process is also referred to as a random function. A realization $\{z(s): s \in D \subseteq \mathbb{R}^d\}$ is just a function of the spatial index $s$, and is sometimes called a regionalized variable (Matheron 1971). In fact, only a partial realization of this random function is observed as data. As random variables, these data are

$$Z(s_1), \ldots, Z(s_n), \quad (2)$$

where $s_1, \ldots, s_n$ are known fixed locations in $D$. Clearly, to make any
Inferences about the process based upon this small fragmentary sample we need to make some assumptions.

If

\[ E(Z(s)) = \mu, \text{ for all } s \in D, \]  

and

\[ \text{var}(Z(s+h) - Z(s)) = 2\gamma(h), \text{ for all } s, s+h \in D, \]

then the process in (1) is said to be intrinsically stationary. The function \( 2\gamma(h) \) characterizes the "small-scale" spatial variability of the process and is called the variogram (Matheron 1963). If a stationary covariance function also exists, viz.,

\[ C(h) = \text{cov}(Z(s+h), Z(s)), \text{ for all } s, s+h \in D, \]

then

\[ C(h) = C(0) - \gamma(h). \]

These assumptions of stationarity can be relaxed somewhat; the mean could be expressed as an unknown linear combination of known explanatory variables in place of (3), and higher-order differences could be used to define a generalized covariance in place of (4).

Prediction

Spatial prediction is the prediction of a functional \( f(Z(s)) \) (or a related noiseless signal) from data \( Z(s_1), \ldots, Z(s_n) \). Assuming that
The mean of the process is known, optimal linear spatial prediction refers to predicting \( f(Z(s)) \) using a linear predictor

\[
\hat{f}(Z(s)) = \lambda_0 + \sum_{i=1}^{n} \lambda_i Z(s_i).
\]  

(7)

The weights \( \{\lambda_i: i = 0, \ldots, n\} \) are chosen so that

\[
E(\hat{f}(Z(s)) - f(Z(s)))^2
\]

is minimized, and so they depend on the underlying covariance structure of the process. When the mean is unknown, the optimal linear unbiased spatial predictor has the form of (7), where now the optimal weights are chosen to minimize (8) subject to unbiasedness constraints.

Since this optimal linear unbiased predictor minimizes (8), it necessarily has a smaller prediction-mean-squared error than that of any other linear unbiased predictor (in particular, the best linear unbiased predictor) obtained by assuming uncorrelated data. Consequently, when spatial correlation is present, but overlooked or ignored, the resulting linear unbiased predictors have standard errors that are larger than necessary.

Kriging is the name given by Matheron (1963) to optimal spatial prediction, named after a South African mining engineer D. G. Krige for his contributions to the estimation of mineral deposits (see Cressie 1989 for the details of the origins of kriging).
Although it is largely terminology from the mining industry that as prevailed in this part of the spatial statistics literature, Gandin (1963) independently and simultaneously developed his theory of optimum interpolation for applications in meteorology. Gandin's goal was the same as Matheron's; namely, to develop the theory of optimal linear unbiased prediction for spatial processes. Although both Matheron and Gandin were interested in spatial applications, the concept of using the covariance structure of a process to construct weights for linear prediction has roots in statistics (mixed linear models, time series), econometrics, and animal breeding.

In a discrete time series context, assuming a known mean, Wold (1938) developed a predictor of the form of (7) using the autocorrelation function of the process to construct optimal weights. Kolmogorov (1941) and Wiener (1949) independently extended Wold's results to more general settings. In a paper concerned with smoothing noisy maps, Thompson (1956) extended Wiener's results to spatial processes.

Goldberger (1962) in econometrics, and Whittle (1963) in mathematical statistics were among the first to consider optimal linear unbiased prediction, using generalized least squares to estimate the unknown mean efficiently. Whittle also mentions the possible application of his results to spatial processes.

A similar development also occurred in the areas of plant and animal breeding. In plant breeding, Fairfield Smith (1936) considers linear prediction of a random variable from explanatory variables.
Using the sample mean to estimate the unknown mean of the data, he uses the covariance structure of the mean-corrected data, as well as the cross-covariance between the mean-corrected data and the unobserved random variable, to construct optimal weights. In a similar manner, Hazel (1943) considers linear prediction in the area of animal breeding. Henderson (1963) extended their results by considering efficient mean-corrections that led to optimal linear unbiased prediction; he uses generalized least squares to estimate the unknown mean, and so arrives at a predictor very similar to that obtained by Goldberger and Whittle.

Thus, it seems that the theory of optimal linear prediction of a random quantity was developed in many different areas almost simultaneously. Judging from the acknowledgment in Fairfield Smith's paper, R. A. Fisher himself may have contemplated kriging! But in a spatial context, it has been Matheron's theory of regionalized variables that has found the largest audience and the broadest applications.

**Estimation and hypothesis testing**

Consider the following general linear model for the data

\[ z = (z(s_1), \ldots, z(s_n))' \]

\[ z = X\beta + \xi, \quad (9) \]

where

- \( X \) is an \( n \times p \) matrix of explanatory variables;
- \( \beta \) is a \( p \times 1 \) vector of fixed unknown parameters;
and
\[ \xi \text{ is an } n \times 1 \text{ random vector with mean zero and covariance matrix } \sigma^2 \Sigma_{ZZ}. \]
Assume \( \Sigma_{ZZ} \) is a known positive definite matrix and \( \sigma^2 \) is an unknown constant.

Since \( \beta \) is unknown, inference on the mean vector \( X\beta \) requires estimation of \( \beta \). Aitken (1935) extended the Gauss-Markov theory of estimation to allow for a very general covariance structure among the data. His generalized least squares estimator is given by

\[ \hat{\beta}_{GLS} = (X' \Sigma_{ZZ}^{-1} X)^{-1} X' \Sigma_{ZZ}^{-1} x, \quad (10) \]

where \( A^- \) denotes a generalized inverse of the matrix \( A \). The estimator \( \hat{\beta}_{GLS} \) is called the best linear unbiased estimator of \( X\beta \) since it has the smallest variance among all other linear unbiased estimators of \( X\beta \).

In particular, it has a smaller variance than that of the ordinary least squares estimator \( \hat{\beta}_{OLS} \), where

\[ \hat{\beta}_{OLS} = (X' X)^{-1} X' Z. \quad (11) \]

Thus, in general, inference procedures for testing hypotheses involving \( \beta \) should be based on (10) rather than (11).

Suppose we wish to test

\[ H_0: \ \beta = \phi, \quad (12) \]
versus the general alternative, (9). Assuming $\mathbf{Z}$ is Gaussian with $\Sigma_{ZZ} = I_n$, the $n \times n$ identity matrix, the appropriate test statistic is (Rao 1973):

$$F = \frac{(\hat{C}_{OLS})'(C(X'X)^{-1}C')^{-1}(\hat{C}_{OLS})/\text{rank}(C)}{\hat{s}^2_{OLS}},$$

(13)

where $\hat{s}^2_{OLS} = (n-p)^{-1}(\mathbf{Z} - \hat{X}_{OLS})'(\mathbf{Z} - \hat{X}_{OLS})$. When $\Sigma_{ZZ} = I_n$, the numerator of (13) has a chi-squared distribution with $k$ (= rank $(C)$) degrees of freedom, and the denominator has a chi-squared distribution with $n-p$ degrees of freedom. The independence of numerator and denominator then implies that the ratio (13) has an $F$-distribution with $k$ and $n-p$ degrees of freedom. However, when $\Sigma_{ZZ}$ is not the identity, the numerator of (13) no longer has a chi-squared distribution (Searle 1971, p. 58). In this case, Johnson and Kotz (1970), page 152 give the exact distribution function of the numerator, although it is cumbersome and computationally prohibitive. However, using Fourier inversion formulas, Imhof (1961) obtains an integral expression for this distribution function that may be approximated using standard numerical integration techniques. Gabler and Wolff (1987) use the technique of matching moments to construct another very accurate approximation that is much easier to use than that of Imhof’s.

From the discussion above, in the case of a general covariance matrix $\Sigma_{ZZ}$, a test for the hypotheses (12) should be based on the generalized least squares estimator (10). Thus, an appropriate test
statistic is

\[ F^* = \frac{(\hat{C}^\text{GLS})'(C(X' \Sigma^{-1}X)^{-1}C')^{-1}(\hat{C}^\text{GLS})/k}{s^2_{\text{GLS}}} \]  \hspace{1cm} (14)

where \( s^2_{\text{GLS}} = (n-p)^{-1}(Z - \hat{X}^\text{GLS})'(\Sigma Z \Sigma)^{-1}(Z - \hat{X}^\text{GLS}) \). It can easily be shown that \( F^* \) follows an F distribution with \( k = \text{rank (C)} \), and \( n-p \) degrees of freedom.

In general, the statistics (13) and (14) are different, and if spatial correlation is present only (14) follows an F distribution. Therefore, valid and efficient inferences for spatial processes must consider the spatial correlation between different parts of the \( Z \)-processes. Failure to do so can lead to the use of inappropriate inference procedures and result in erroneous conclusions.
SECTION I. IMPROVED PREDICTION WITH THE GENERALIZED LINEAR MODEL
IMPROVED PREDICTION WITH THE GENERALIZED LINEAR MODEL

by

Carol A. Gotway and Noel Cressie

Department of Statistics, Iowa State University, Ames, IA 50011

Supported in part by the National Science Foundation under grant number DMS-8902812.
ABSTRACT

A fundamental relationship between prediction and estimation is exploited to obtain biased predictors with smaller risk than that of the usual best linear unbiased predictor. Assuming a general covariance structure, several such predictors are derived and their properties are discussed. Simultaneous multiple prediction is emphasized and applications to spatial statistics are featured.
1. INTRODUCTION

The theory of univariate prediction has largely been limited to the problem of predicting a single random variable or a single linear combination of random variables. Little attention has been given to multiple univariate prediction; in practice such predictions are carried out by predicting each random variable one by one using techniques optimal for predicting just one random variable. When the goal is the prediction of many points, predictors and the criteria used to judge them should reflect this goal. In this paper we extend the idea of prediction to include simultaneous multiple prediction, and consider a risk function that measures the global performance of the prediction technique.

Assume the following generalized linear model for $Z$, and unobservables $Z_0$:

$$Z = X\hat{\theta} + \epsilon$$

(1.1)

$$Z_0 = X_0\hat{\theta} + \epsilon_0,$$

where

$Z$ is the $n$-dimensional data vector;

$Z_0$ is a $k$-dimensional vector of unobserved values that will be predicted from the data;
and \(X\) are matrices of explanatory variables (\(\text{rank}(X) = p\)), the rows of which are often polynomials of spatial locations;

\(\beta\) is a px1 vector of unknown fixed parameters;

and

\(\varepsilon\) and \(\varepsilon_0\) are random errors with zero mean and covariance

\[
\text{var}((\varepsilon', \varepsilon_0')) = \sigma^2 \begin{bmatrix} \Sigma_{Z\varepsilon} & \Sigma_{Z0} \\ \Sigma_{0Z} & \Sigma_{00} \end{bmatrix} = \sigma^2 \Sigma, \tag{1.2}
\]

where \(\Sigma_{Z\varepsilon}\), \(\Sigma_{00}\), and \(\Sigma\) are known positive definite matrices and \(\sigma^2\) is an unknown constant.

Based on this model, the problem is to predict the \(k\)-dimensional vector \(Z_0\) using some optimal function of the data \(Z\). Let the predictor be

\[
p(Z) = \begin{bmatrix} p_1(Z) \\ \vdots \\ p_k(Z) \end{bmatrix}, \tag{1.3}
\]

and write
Following the ideas of James and Stein (1961) in the realm of multivariate estimation, an appropriate measure (loss function) of overall prediction performance is

\[ l(p, Z_0) = \sigma^{-2} \sum_{i=1}^{k} (p_i(Z) - Z_{0,i})^2, \]  

which may be considered as an index of accuracy for either stochastic or nonstochastic predictors. For stochastic predictors, expectations may be taken, yielding the risk function:

\[ r(p, Z_0; \beta, \sigma^2) = \sigma^{-2} E\left( \sum_{i=1}^{k} (p_i(Z) - Z_{0,i})^2 \right) \]

\[ - \sigma^{-2} E\left( (p(Z) - Z_0)'(p(Z) - Z_0) \right). \]

This risk function reflects, in statistical terms, the desire to predict well at all k points collectively and not just at each point individually. This can be important for spatial processes where the goal is to produce a map of a spatial variable of interest, or in time series where
we need to make multiple-step-ahead forecasts. Then (1.5) and (1.6) are measures of map or forecasting accuracy.

In the following sections, we present and compare several stochastic predictors using the measure (1.6). Section 2 gives a concise overview of linear predictors, and Section 3 extends these ideas to nonlinear prediction. In Section 4, we present a general class of predictors based on a fundamental relationship between prediction and estimation and go on to develop predictors with smaller risk than that of the best linear unbiased predictor.
2. LINEAR PREDICTION

2.1. Known mean

Initially, to motivate the ideas behind linear prediction, we shall assume that \( \beta \) in (1.1) is known. Since this is often an unrealistic assumption, it will be relaxed in Section 2.2, where linear prediction with \( \beta \) unknown is considered.

i) Best homogeneous linear prediction (Toutenburg 1982).

Consider the linear predictor

\[
p(Z) = AZ,
\]

where \( A \) is a \( k \times n \) matrix. Its risk (from (1.7)) is given by

\[
s^{-2}(\beta'(AX - X_0)'(AX - X_0)\beta) + \text{tr}(\Sigma_0 - 2\Sigma_2) + \Sigma_0 - 2\Sigma_2. \tag{2.2}
\]

The best homogeneous linear predictor is constructed by finding the matrix \( A \) that minimizes (2.2). Differentiating with respect to \( A \), equating to zero and solving gives the optimal value of \( A \):

\[
A_{\text{opt}} = (\sigma^2\Sigma_0 + (X_0\beta)(X\beta)' + \sigma^2\Sigma_2 + (X_\beta)(X\beta)')^{-1}. \tag{2.3}
\]


It is straightforward to verify that $A_{\text{opt}}$ minimizes (2.2) and hence the best homogeneous linear predictor $g(Z) = A_{\text{opt}}Z$, is:

$$P_1(Z) = \langle \sigma^2\Sigma_{0Z} + (X_0\beta)(X\beta)' \rangle \langle \sigma^2\Sigma_{ZZ} + (X\beta)(X\beta)' \rangle^{-1}Z. \quad (2.4)$$

which is a well-defined predictor if $\sigma^2$ is assumed known.

The predictor $P_1(Z)$ has bias,

$$E(P_1(Z) - Z_0) = \langle \sigma^2\Sigma_{0Z} + (X_0\beta)(X\beta)' \rangle \langle \sigma^2\Sigma_{ZZ} + (X\beta)(X\beta)' \rangle^{-1}X - X_0\beta. \quad (2.5)$$

Its risk is given by (2.2) with $A$ replaced with $A_{\text{opt}}$.

ii) Best heterogeneous linear prediction.

Consider the heterogeneous linear predictor

$$P(Z) = BZ + \zeta, \quad (2.5)$$

where $B$ is a $k \times n$ matrix and $\zeta$ is a $k \times 1$ vector. From (1.7) its risk is

$$\sigma^{-2}((BZ - X_0\beta + \zeta)'((BZ - X_0\beta + \zeta)) + \text{tr}(B\Sigma_{ZZ}B' + \Sigma_{00} - 2B\Sigma_{Z0}). \quad (2.6)$$

The best heterogeneous linear predictor is based on finding $B$ and $\zeta$ that minimize (2.6). Differentiating once with respect to $B$, and again with respect to $\zeta$, and equating to zero gives the optimal values for $B$ and $\zeta$. 
that these values minimize (2.6) is shown in Toutenburg (1982), p. 140. Hence the best heterogeneous linear predictor \( p_2(Z) = B_{\text{opt}} Z + \zeta_{\text{opt}} \), is

\[
  p_2(Z) = \Sigma_{0Z} \Sigma_{ZZ}^{-1} Z + X_0 \theta - \Sigma_{0Z} \Sigma_{ZZ}^{-1} x_0. \tag{2.8}
\]

This predictor \( p_2(Z) \) is well known in time series (see Wold 1938, p. 77, or Fuller 1975, p. 75), and in geostatistics it is called simple kriging.

The best heterogeneous linear predictor (2.8) is unbiased for \( Z_0 \), and the minimized risk is

\[
  r(p_2, Z_0; \theta, \sigma^2) = \text{tr}(\Sigma_{00} - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \Sigma_{Z0}). \tag{2.9}
\]

Consequently,

\[
  r(p_2, Z_0; \theta, \sigma^2) \leq r(p_1, Z_0; \theta, \sigma^2)
\]

(Toutenburg 1982), and so the best heterogeneous linear predictor (2.8) is preferred to the best homogeneous linear predictor (2.4) when the mean is known.
2.2. Unknown mean

1) Best linear unbiased prediction (Goldberger 1962).

Consider the heterogeneous linear predictor

\[ \mathbf{p}(\mathbf{Z}) = \mathbf{CZ} + \mathbf{d} \tag{2.10} \]

where \( \mathbf{C} \) is a \( k \times n \) matrix and \( \mathbf{d} \) is a \( k \times 1 \) vector. Its risk (using (1.7)) is

\[ \sigma^2 \{ (\mathbf{CX} - \mathbf{X} \theta)^\top (\mathbf{CX} - \mathbf{X} \theta) + \mathbf{d}^\top \mathbf{d} \} + \text{tr}(\mathbf{B} \Sigma_{\mathbf{ZZ}} \mathbf{B}^\top + \Sigma_{00} - 2 \Sigma_{0Z}) \tag{2.11} \]

The best linear unbiased predictor is determined by finding \( \mathbf{C} \) and \( \mathbf{d} \) that minimize (2.11) subject to the uniform unbiasedness constraint \( \mathbf{CX} \theta + \mathbf{d} = \mathbf{X} \theta \) for all \( \theta \). Using the method of Lagrange multipliers gives

\[ \mathbf{d}_{\text{opt}} = \mathbf{0}, \]

\[ \mathbf{C}_{\text{opt}} = \Sigma_{0Z} \Sigma_{ZZ}^{-1} - (\mathbf{X}_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \mathbf{X})(\mathbf{X}' \Sigma_{ZZ}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{ZZ}^{-1}; \tag{2.12} \]

which can be shown to minimize (2.11) subject to the constraint. Hence the best linear unbiased predictor \( \hat{\mathbf{p}}_3(\mathbf{Z}) = \mathbf{C}_{\text{opt}} \mathbf{Z} + \mathbf{d}_{\text{opt}} \), is

\[ \hat{\mathbf{p}}_3(\mathbf{Z}) = \Sigma_{0Z} \Sigma_{ZZ}^{-1} \mathbf{Z} + (\mathbf{X}_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \mathbf{X}) \hat{\theta}_{\text{GLS}}, \tag{2.13} \]

where \( \hat{\theta}_{\text{GLS}} \) is the generalized least squares estimator of \( \theta \).
In the geostatistics literature, the predictor $p_3(Z)$ is called the universal kriging predictor.

The relationship (2.13) gives a hint of a link between optimal estimators and optimal predictors. This relationship may also be seen from another point of view. Consider writing the model (1.1) as

$$Y = W\beta + u,$$  \hspace{1cm} (2.15)

where

$$\bar{Y} = \begin{bmatrix} Z \\ \phi \end{bmatrix}, \quad \bar{w} = \begin{bmatrix} X & \phi \\ X_0 & -I_k \end{bmatrix}, \quad \gamma = \begin{bmatrix} \beta \\ z_0 \end{bmatrix}, \quad \text{and } u = \begin{bmatrix} \varepsilon \\ \varepsilon_0 \end{bmatrix};$$

hence $\text{var}(u) = \sigma^2\Sigma$, as in (1.2). Then under model (2.15), the generalized least squares estimator of $\gamma$ is

$$\hat{\gamma}_{GLS} = (\bar{w}'\Sigma^{-1}\bar{w})^{-1}\bar{w}'\Sigma^{-1}\bar{Y},$$

which after simplification gives

$$\hat{\gamma}_{GLS} = \left(\begin{array}{c}
(X'\Sigma_{zz}^{-1}X)^{-1}X'\Sigma_{zz}^{-1}Z \\
\Sigma_{oz}\Sigma_{zz}^{-1}Z + (X_0 - \Sigma_{oz}\Sigma_{zz}^{-1}X)(X'\Sigma_{zz}^{-1}X)^{-1}X'\Sigma_{zz}^{-1}Z
\end{array}\right)$$ \hspace{1cm} (2.16)
Thus, a generalized least squares estimation procedure applied using model (2.15) produces the best linear unbiased predictor (2.13).

The predictor \( \hat{p}_3^*(Z) \) is unbiased by construction and its risk is

\[
r(\hat{p}_3^*, Z_0; \beta, \sigma^2) = \text{tr}(\Sigma_{00} - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \Sigma_{Z0}) \\
+ \text{tr}((X' \Sigma_{ZZ}^{-1} X)^{-1} (X_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} X)' (X_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} X)).
\]  

(2.17)

It is also equivariant to location and scale changes in that

\[
p_3(b(Z+Xk)) = b(p_3(Z) + X_0 k), \tag{2.18}
\]

for all pxl vectors \( k \) and all scalars \( b \). Using \( p_3(Z) = C_{opt}^Z \) in (2.18) implies

\[
bc_{opt}^Z + bc_{opt}^X k = bc_{opt}^Z + bX_0 k,
\]

for all \( b \) and \( k \). This gives the condition \( C_{opt}^X k = X_0 k \) for all \( k \), which it can be shown is exactly the uniform unbiasedness condition given earlier. Thus, minimizing (2.11) subject to equivariance (see (2.18)) is
equivalent to minimizing (2.11) subject to unbiasedness. Hence $p_3(Z)$ is also the best linear equivariant predictor (Kaminsky et al. 1975).

Because its first-order and second-order moments are straightforward to evaluate, the best linear unbiased predictor $p_3(Z)$ is the predictor most often used in cases when the mean is unknown. The problem with nonlinear and biased predictors, as we shall see in the next section, is that properties such as bias and risk are very difficult to ascertain. However, in a later section we shall develop nonlinear biased predictors whose first-order and second-order moments are relatively easy to evaluate, and whose risk is smaller than that of the best linear unbiased predictor.

ii) Shrinkage prediction.

Following an eloquent interpretation of shrinkage estimation by Srivastava and Giles (1987), consider

$$p'(Z) = (1 - \theta)(p_3(Z) - X_0\theta) + X_0\theta,$$  \hspace{1cm} (2.19)

obtained by shrinking the best linear unbiased predictor towards the mean of $Z_0$. From (1.7), its risk is

$$\sigma^{-2}(1 - \theta)^2 \text{tr}(\text{var}(p_3(Z))) - 2(1 - \theta)E((p_3(Z) - X_0\theta)'Z) + \text{tr}(\Sigma_0).$$ \hspace{1cm} (2.20)
The parameter $\theta$ in (2.19) will be chosen so that the risk (2.20) is minimized. Differentiating (2.20) with respect to $\theta$, equating to zero and solving gives

$$
\theta_{\text{opt}} = \frac{\text{tr}(\text{var}(p_3(Z))) - \text{tr}(\text{cov}(p_3(Z), Z_0))}{\text{tr}(\text{var}(p_3(Z)))},
$$

(2.21)

where

$$
\text{var}(p_3(Z)) = \sigma^2 \Sigma_0 \Sigma_{ZZ}^{-1} \Sigma_{Z0}
$$

$$
+ (X_0 - \Sigma_0 \Sigma_{ZZ}^{-1} X) (X' \Sigma_{ZZ}^{-1} X)^{-1} (X_0 - \Sigma_0 \Sigma_{ZZ}^{-1} X)',
$$

and

$$
\text{cov}(p_3(Z), Z_0) = \sigma^2 \Sigma_0 \Sigma_{ZZ}^{-1} \Sigma_{Z0} + (X_0 - \Sigma_0 \Sigma_{ZZ}^{-1} X) (X' \Sigma_{ZZ}^{-1} X)^{-1} X' \Sigma_{ZZ}^{-1} X_0 \Sigma_{Z0}'
$$

It is easily seen that $\theta_{\text{opt}}$ minimizes (2.20). The resulting predictor is then

$$
p_4(Z) = (1 - \theta_{\text{opt}})(p_3(Z) - X_0 \beta) + X_0 \beta.
$$

(2.22)

However, $p_4(Z)$ is not of any practical use since it depends on $\beta$. Then estimating $\beta$ by $\hat{\beta}_{\text{GLS}}$ in (2.14) gives a feasible predictor of $Z_0$,

$$
p_5(Z) = (1 - \theta_{\text{opt}})(p_3(Z) - X_0 \hat{\beta}_{\text{GLS}}) + X_0 \hat{\beta}_{\text{GLS}}.
$$

(2.23)
It is easy to verify that the predictor (2.23) is unbiased for $Z_0$, and its risk is

\[
\begin{align*}
    r(\mathcal{P}_5, Z_0; \theta, \sigma^2) &= \text{tr}(\Sigma_{00}) + \sigma^{-2} \theta_{\text{opt}}^2 \text{tr}(\text{var}(\mathcal{P}_3(Z))) \\
    &\quad - 2\sigma^{-2} \theta_{\text{opt}} \text{tr}(\text{cov}(\mathcal{P}_3(Z), Z_0)) \\
    &\quad + (1 - \theta_{\text{opt}})(\theta_{\text{opt}}) \text{tr}(X_0 (X' \Sigma_{ZZ}^{-1}X)^{-1}X'_0) \\
    &\quad - 2(1 - \theta_{\text{opt}}) \text{tr}(X_0 (X' \Sigma_{ZZ}^{-1}X)^{-1}X'_0 \Sigma_{ZZ}^{-1}X_0').
\end{align*}
\]
3. NONLINEAR PREDICTION

i) Homogeneous linear prediction with estimated parameters.

The best homogeneous linear predictor

\[ P_1(Z) = (\sigma^2 \Sigma_{ZZ} + (X_0 \hat{\beta})(X_0 \hat{\beta})') (\sigma^2 \Sigma_{ZZ} + (X \hat{\beta})(X \hat{\beta})')^{-1} Z \]

was derived for the case when the mean is known, and is not a feasible predictor when the mean is unknown since it depends on unknown parameters. Estimating \( \hat{\theta} \) by \( \hat{\theta}_{GLS} \) in (2.14), and \( \sigma^2 \) by \( s^2 = (Z-X \hat{\beta}_{GLS})' \Sigma_{ZZ}^{-1} (Z-X \hat{\beta}_{GLS}) \), gives a practical version of (2.4), namely

\[ P_6(Z) = (s^2 \Sigma_{ZZ} + (X_0 \hat{\beta}_{GLS})(X_0 \hat{\beta}_{GLS})') (s^2 \Sigma_{ZZ} + (X \hat{\beta}_{GLS})(X \hat{\beta}_{GLS})')^{-1} Z. \]

Although the optimality of the original estimator is now lost due to this substitution, it is not unreasonable to believe that this resulting predictor may have desirable properties of its own. However, it is difficult to obtain exact expressions for the bias and risk of \( P_6(Z) \) because of its highly nonlinear nature. Approximations may be possible using the results of Hosmane (1988) on small disturbance methodology.
ii) James-Stein prediction via regression.

For $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, $\text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}$, the James-Stein estimator of $\boldsymbol{\beta}$ is given by (James and Stein 1961, Baranchik 1964, and Judge et al. 1985):

$$\hat{\boldsymbol{\beta}}_{JS} = \left(1 - \frac{(p - 2)(n - p)s^2}{(n - p + 2)\hat{\mathbf{e}}_{GLS}'\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}}\right)\hat{\mathbf{e}}_{GLS}, \quad (3.2)$$

where

$$s^2 = (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS})'\mathbf{Z}\mathbf{Z}_{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}).$$

Now, applying this procedure to $\mathbf{y} = \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{Z}_0 \end{bmatrix}$ in model (2.15) and simplifying gives

$$\hat{\mathbf{y}}_{JS} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{y}} \\ \mathbf{p}_3(\mathbf{Z}) \end{bmatrix} = \left(1 - \frac{(p + k - 2)(n - p)s^2}{(n - p + 2)\hat{\mathbf{e}}_{GLS}'\mathbf{X}'\mathbf{Z}\mathbf{Z}_{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_{GLS} + \mathbf{Z}'\mathbf{V}\mathbf{Z}}\right) \begin{bmatrix} \hat{\mathbf{e}}_{GLS} \\ \mathbf{p}_3(\mathbf{Z}) \end{bmatrix},$$

where $\mathbf{V} = \mathbf{Z}\mathbf{Z}_{-1}\mathbf{Z}_0\mathbf{Z}_0'\mathbf{Z}\mathbf{Z}_{-1}\mathbf{Z}_0'\mathbf{Z}\mathbf{Z}_{-1}\mathbf{Z}_0$ and $\mathbf{p}_3(\mathbf{Z})$ is the best linear unbiased predictor of $\mathbf{Z}_0$ given in (2.13).

By construction, $\hat{\mathbf{y}}_{JS}$ has uniformly smaller risk than $\hat{\mathbf{y}}_{GLS}$ in (2.16). However, it may not be the case that $\mathbf{p}_7(\mathbf{Z})$ has uniformly smaller risk than that of $\mathbf{p}_3(\mathbf{Z})$. In Section 4.2, we define a predictor that is similar in form to $\mathbf{p}_7(\mathbf{Z})$, but always has smaller risk than $\mathbf{p}_3(\mathbf{Z})$ (recall that $\mathbf{p}_3(\mathbf{Z})$ is the universal kriging predictor).
4. A GENERAL CLASS OF PREDICTORS

In this section, we exploit a fundamental relationship between prediction and estimation to construct a general class of predictors (assuming an unknown mean) that includes all of the predictors mentioned previously. Moreover, the risk of these predictors is composed of two parts: one corresponding to pure prediction error and one corresponding to the estimation of the unknown mean parameter.

Consider the best heterogeneous predictor of $Z_0$ from Section 2.1:

$$P_2(Z) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}Z + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\beta}. \quad (4.1)$$

In Section (2.2), the best linear unbiased predictor of $Z_0$ was obtained by using the generalized least squares estimator to estimate $\beta$ in (4.1). Using this same idea, consider the general class of predictors

$$P(Z, \hat{\beta}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}Z + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\beta}, \quad (4.2)$$

where $\hat{\beta}$ is any estimator of $\beta$. If $(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)$ is of full column rank, then any predictor may be written in this form, for if $p(Z)$ is any predictor, just take $\hat{\beta}$ to be

$$(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)'(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)^{-1}(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)'(p(Z) - \Sigma_{0Z}\Sigma_{ZZ}^{-1}Z).$$

Moreover, $P(Z, \hat{\beta})$ inherits its first-order and second-order moment properties from the estimator $\hat{\beta}$: If $\hat{\beta}$ is unbiased for $\beta$, then $P(Z, \hat{\beta})$ is
unbiased for \( Z_0 \), and it is shown below that if \( \hat{\theta} \) "improves upon" \( \hat{\theta}_{GLS} \), then the predictor \( p(Z, \hat{\theta}) \) "improves upon" \( p_3(Z) \).

4.1. Decomposition of risk

Harville (1985) presents a general decomposition of prediction error that can be specialized to our situation. We present here a direct proof of the decomposition of risk since it illuminates our basic approach to improved prediction.

**Lemma 4.1.**

Assume the model (1.1). If \( \hat{\theta} = BZ \) for some pxn matrix B, and

\[
p(Z, \hat{\theta}) = \begin{bmatrix} p_1(Z, \hat{\theta}) \\ \vdots \\ p_k(Z, \hat{\theta}) \end{bmatrix}
\]
given by equation (4.2), is a predictor of

\[
Z_0 = \begin{bmatrix} Z_{0,1} \\ \vdots \\ Z_{0,k} \end{bmatrix}, \text{ then}
\]

\[
E(p_1(Z, \hat{\theta}) - Z_{0,1})^2 = E(p_1(Z, \hat{\theta}) - Z_{0,1})^2 + E(p_1(Z, \hat{\theta}) - p_1(Z, \hat{\theta}))^2,
\]

which upon expansion is equal to
\[ \sigma^2(\Sigma_{00,(1,1)} - \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}(\Sigma_{0Z,i}')) + \mathbb{E}((\hat{\beta} - \beta)')(X_{0,i} - \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}X)'(X_{0,i} - \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}X)(\hat{\beta} - \beta) \]

where \( \Sigma_{00,(1,1)} \) is the \((1,1)\)-th element of \( \Sigma_{00} \) and \( X_{0,i} \) and \( \Sigma_{0Z,i} \) are the \( i \)-th rows of \( X_0 \) and \( \Sigma_{0Z} \), respectively.

**PROOF:**

\[
\mathbb{E}(p_i(Z,\hat{\beta}) - Z_{0,i})^2 = \mathbb{E}(p_i(Z,\hat{\beta}) - p_i(Z,\beta))^2
\]

\[+ 2\mathbb{E}(p_i(Z,\hat{\beta}) - p_i(Z,\beta))(p_i(Z,\beta) - Z_{0,i})
\]

\[+ \mathbb{E}(p_i(Z,\beta) - Z_{0,i})^2. \quad (4.3)\]

If \( \hat{\beta} = BZ \), then \( \text{cov}(\hat{\beta}, p_i(Z,\beta) - Z_{0,i}) = \phi \) and the cross product term of (4.3) is zero. Thus,

\[
\mathbb{E}(p_i(Z,\hat{\beta}) - Z_{0,i})^2 = \mathbb{E}(p_i(Z,\hat{\beta}) - p_i(Z,\beta))^2 + \mathbb{E}(p_i(Z,\beta) - Z_{0,i})^2.
\]

Substituting \( p_i(Z,\hat{\beta}) = \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}Z + (X_{0,i} - \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}X)\hat{\beta} \), and

\[p_i(Z,\beta) = \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}Z + (X_{0,i} - \Sigma_{0Z,i} \Sigma_{ZZ}^{-1}X)\beta\]

into (4.3) and taking expectations gives the result. \( \blacksquare \)
Lemma 4.2

Assume the model (1.1), and that

\[ E(Z_{0,i} | Z) = \Sigma_{0,i} \Sigma_{Z}^{-1} Z + (X_{0,i} - \Sigma_{0,i} \Sigma_{Z}^{-1} X) \beta \]

\[ = p_i(Z, \beta). \]

Then Lemma 4.1 holds for any \( \hat{\beta} \), linear or not.

PROOF:

Let \( \delta(Z) \) be any function of the data \( Z \). Then

\[ \text{cov}(\delta(Z), p_i(Z, \beta) - Z_{0,i}) = \text{cov}(\delta(Z), E(Z_{0,i} | Z) - Z_{0,i}) \]

\[ = E(\delta(Z) (E(Z_{0,i} | Z) - Z_{0,i})) \]

\[ = E(\delta(Z) (E(Z_{0,i} | Z) - Z_{0,i} | Z)) \]

\[ = 0. \]

Thus, since \( \hat{\beta} - \delta(Z) \), for some function \( \delta \), \( \text{cov}(\hat{\beta}, p_i(Z, \beta) - Z_{0,i}) = \phi \), and the cross product term in (4.3) is zero. Hence the decomposition of Lemma 4.1 holds.
Note that when \( Z \) and \( Z_0,1 \) are jointly normal, the assumption of Lemma 4.2 is satisfied. Thus, for linear predictors under model (1.1), or for any predictors under (1.1) that satisfy the assumptions of Lemma 4.2, or when \( Z \) and \( Z_0 \) are jointly normal under (1.1), the risk function (1.6) may be decomposed into the sum of two parts.

\[
r(p, Z_0; \beta, \sigma^2) = \text{tr}(\Sigma_{00} - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \Sigma_{Z0}) \\
+ \sigma^{-2} E(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta),
\]

(4.4)

where

\[
Q = (X_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} X)'(X_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} X).
\]

(4.5)

The first term of the risk (4.4) is inherent in prediction of \( Z_0 \) with \( \beta \) known (it is equal to \( r(p_2, Z_0; \hat{\beta}, \sigma^2) \)); the second is due to estimation of \( \beta \).

The decomposition (4.4) has important implications for the comparison of two predictors.

**Definition.** A predictor \( p^*(Z) = p(Z, \hat{\beta}_1) \) is preferred to another predictor \( p^{**}(Z) = p(Z, \hat{\beta}_2) \) if \( r(p^*, Z_0; \hat{\beta}, \sigma^2) \leq r(p^{**}, Z_0; \hat{\beta}, \sigma^2) \), for all \( \hat{\beta} \) and \( \sigma^2 \).
From (4.4), and the definition, \( \hat{p}^\ast \) is preferable to \( \hat{p}^{**} \) if and only if

\[
\sigma^{-2}E((\hat{\beta}_1 - \beta)'Q(\hat{\beta}_1 - \beta)) - \sigma^{-2}E((\hat{\beta}_2 - \beta)'Q(\hat{\beta}_2 - \beta)) \leq 0, \quad (4.6)
\]

for all \( \beta \) and \( \sigma^2 \), i.e., if and only if \( \hat{\beta}_1 \) has uniformly smaller (or equal) weighted mean squared error than that of \( \hat{\beta}_2 \). Thus, the problem of prediction and the comparison of predictors has been reduced to the problem of estimation and the comparison of estimators.

### 4.2. Improved prediction

In this section we construct predictors, based on the general class (4.2), that have risk that is less than or equal to that of the best linear unbiased predictor, over regions of \( (\beta, \sigma^2) \in \mathbb{R}_+^2 \times (0, \infty) \). The minimax predictor to be defined, is shown to achieve uniform improvement, making it the predictor of choice among the many presented in this article.

1) Bayesian prediction.

In model (1.1), suppose

\[
Z|\beta \sim N(X\beta, \sigma^2 \Sigma_{zz}) \quad (4.7)
\]

and

\[
\beta \sim N(\theta, \Sigma_{\beta\theta}).
\]
The distribution of $\beta$ reflects the prior information about the unknown parameter $\beta$. Here $\sigma^2$ is a nuisance parameter assumed fixed but unknown; a prior for $\sigma^2$ could also have been chosen, but this will not change the estimator of $\beta$. Straightforward algebra (see, e.g., Berger 1985) gives

$$\hat{\beta}_B = (X'\Sigma_{ZZ}^{-1}X + \Sigma_{\beta\beta}^{-1})^{-1}(\Sigma_{\beta\beta}^{-1} \mu + X'\Sigma_{ZZ}^{-1} \pi)$$

(4.8)

as the Bayes estimator of $\beta$. Substituting (4.8) into (4.2) gives the Bayes predictor of $Z_0$:

$$p_\beta(Z) = p(Z, \hat{\beta}_B)$$

(4.9)

$$= \Sigma_0^{-1}Z + (X_0'\Sigma_{0Z}^{-1}X)X'\Sigma_{ZZ}^{-1}X + \Sigma_{\beta\beta}^{-1})^{-1}(\Sigma_{\beta\beta}^{-1} \mu + X'\Sigma_{ZZ}^{-1} \pi).$$

This is also the Bayes predictor derived independently by Kitanidis (1986). Notice that when $\Sigma_{\beta\beta}^{-1} = \phi$, corresponding to no information on $\beta$, (4.9) reduces to $p_3(Z)$ of (2.13), the best linear unbiased (kriging) predictor. Also, taking $\beta = \phi$, and $\Sigma_{\beta\beta} = (1/k)I_p$ for some constant $k$ in (4.8) gives the ridge regression estimator

$$\hat{\beta}_R = (X'\Sigma_{ZZ}^{-1}X + kI_p)^{-1}(X'\Sigma_{ZZ}^{-1}Z)$$

(4.10)

(Hoerl and Kennard 1970), and hence a ridge predictor.
The Bayes predictor (4.9) is biased;

\[ \mathbb{E}(p_B(Z) - Z_0) = (\Sigma_0 Z \Sigma_Z^{-1} X - X_0)^T \theta \]

\[ + (X_0 - \Sigma_0 Z \Sigma_Z^{-1} X)(X' \Sigma_Z^{-1} X + \Sigma_{\beta \beta}^{-1})^{-1} \theta + X' \Sigma^{-1} X \theta), \]

where the expectation is taken with respect to the distribution of \( Z | \theta \).

Also, since \( \hat{\theta}_B \) is linear in the data, Lemma 4.1 applies and from (4.4) its risk is:

\[ r(p_B; Z_0; \theta, \sigma^2) = \text{tr}(\Sigma_{00} - \Sigma_0 Z \Sigma_Z^{-1} Z_0) \]

\[ + \text{tr}(T'QT(X' \Sigma_Z^{-1} X)^{-1}) + \sigma^{-2}(\theta - \hat{\theta})'W'QW(\theta - \hat{\theta}), \]

where \( T = (X' \Sigma_Z^{-1} X + \Sigma_{\beta \beta}^{-1})^{-1}(X' \Sigma_Z^{-1} X), W = I - T, \) and \( Q \) is given by (4.5). From this we can see that \( r(p_B; Z_0; \theta, \sigma^2) \leq r(p_3; Z_0; \theta, \sigma^2), \) i.e., the Bayes predictor (4.9) has smaller or equal risk than that of the best linear unbiased (kriging) predictor (2.13), if and only if

\[ \sigma^{-2}(\theta - \hat{\theta})'W'QW(\theta - \hat{\theta}) \leq \text{tr}((X' \Sigma_Z^{-1} X)^{-1}) - \text{tr}(T'QT(X' \Sigma_Z^{-1} X)^{-1}). \]
It is possible to satisfy (4.13) by choosing $\beta$ close to $\theta$. Of course, the problem is that $\theta$ is unknown. Hoerl and Kennard (1970) give some discussion on this problem in the context of ridge regression. For further discussion, see Smith (1973) and Giles and Rayner (1979).

ii) Minimax prediction.

Judge and Bock (1978) consider the family of estimators of $\beta$:

$$\hat{\beta}_{JB} = (I_p - h(\hat{\beta}_{GLS}^TB\hat{\beta}_{GLS}/s^2)C)\hat{\beta}_{GLS},$$

(4.14)

where

$B$ and $C$ are pxp real matrices;

$h(\cdot)$ is a real-valued function;

$$s^2 = (z - X\hat{\beta}_{GLS})'(ZZ)^{-1}(z - X\hat{\beta}_{GLS});$$

(4.15)

$\hat{\beta}_{GLS}$ is the generalized least squares estimator of $\beta$ given in equation (2.14);

and

$I_p$ is the pxp identity matrix.

Using $\hat{\beta}_{JB}$ as an estimator of $\beta$ in (4.2), we obtain the corresponding family of predictors

$$p_{10}(Z) = p(Z, \hat{\beta}_{JB}) = \Sigma_0Z_{ZZ}^{-1}z + (x_0 - \Sigma_0Z_{ZZ}^{-1}x)\hat{\beta}_{JB}.$$
The conditions under which this family of predictors has risk (given by (4.4)) less than or equal to that of the best linear unbiased predictor are given in the following theorem.

**Theorem 4.1**

Assume $Z$ and $Z_0$ are jointly Gaussian with mean and covariance given by (1.1) and (1.2), respectively. Furthermore, assume that $Q$ given by (4.5) has rank $p$, and matrices $C$ and $B$ of (4.14) are chosen so that $Q^{1/2}CQ^{1/2}$ and $Q^{-1/2}BQ^{-1/2}$ are positive definite matrices that commute with each other and that also commute with $Q^{1/2}(X'\Sigma_{ZZ}^{-1}X)^{-1}Q^{1/2}$. Let $\lambda_1(D)$ and $tr(D)$ denote the maximum eigenvalue and the trace of any square matrix $D$.

If

\[
0 < c = \frac{2(tr(C(X'\Sigma_{ZZ}^{-1}X)^{-1}Q - 2\lambda_1(C(X'\Sigma_{ZZ}^{-1}X)^{-1}Q)))}{(n - p + 2) \lambda_1(C'QCB^{-1})};
\]  

(4.17)

ii) $0 \leq h(u) \leq c/u$, for all $u > 0$, and $h$ is differentiable for all $u > 0$;

iii) $\psi(u) = u^q(c/u - h(u))^{-(1+f)}h(u)$

is nondecreasing in $u$ if $h(u) \leq c/u$, where
\[ q = \left( c(n - p - 2)/4 \right) \left( \lambda_1 (\Sigma CB^{-1}) / \lambda_1 (\Sigma C (C' \Sigma C)^{-1} C') \right) \]

and \( f = \left( 4/(n - p - 2) \right) q \);

then

\[ r(P_{10}; Z_0; \beta, \sigma^2) \leq r(P_3; Z_0; \beta, \sigma^2), \] for all \( (\beta, \sigma^2) \in \mathbb{R}^p \times (0, \infty) \).

Here \( r(\cdot, Z_0; \beta, \sigma^2) \) is given in (4.4), \( P_3(Z) \) is the best linear unbiased predictor of \( Z_0 \) given in (2.13), and \( P_{10} \) is given by (4.16).

**PROOF:**

Since \( Z \) and \( Z_0 \) are jointly normal, Lemma 4.2 applies so that from (4.4),

\[ r(P_{10}; Z_0; \beta, \sigma^2) = \text{tr}(\Sigma_{00} - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \Sigma_{Z0}) \]

\[ + \sigma^2 E((\hat{\beta} - \beta)'Q(\hat{\beta} - \beta)). \]

Substituting in for \( P_{10}(Z) \), \( P_3(Z) \), \( \hat{\beta}_{JB} \) and \( \hat{\beta}_{GLS} \) using (4.16), (2.13), (4.14) and (2.14) respectively gives

\[ r(P_{10}; Z_0; \beta, \sigma^2) - r(P_3; Z_0; \beta, \sigma^2) \]

\[ = \sigma^2 (E((\hat{\beta}_{JB} - \beta)'Q(\hat{\beta}_{JB} - \beta)) - E((\hat{\beta}_{GLS} - \beta)'Q(\hat{\beta}_{GLS} - \beta))). \]
Conditions (i), (ii) and (iii) ensure that $\hat{\beta}_{JB}$ is minimax (Judge and Bock 1978, p. 234) under loss function $\sigma^{-2}(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta)$. Hence

\[
\mathcal{R}_1(\mathbb{P}_0, Z_0; \beta, \sigma^2) \leq \mathcal{R}_3(\mathbb{P}_3, Z_0, \beta, \sigma^2), \text{ for all } (\beta, \sigma^2) \in \mathbb{R}^p \times (0, \infty).
\]

Theorem 4.1 shows that a large class of (nonlinear) predictors can be found that have uniformly smaller or equal risk than the best linear unbiased (kriging) predictor. We now specialize this class further to give predictors that we would recommend using in practice.

iii) Improved nonlinear prediction.

**Corollary 4.1.**

Under the assumptions and notation of theorem 4.1, if

\[
\mathbb{P}_{11}(Z) = \mathbb{P}(Z, \hat{\beta}_{JS}) = \Sigma_{ZZ}^{-1}Z + (X_0 - \Sigma_{ZZ}^{-1}X)\hat{\beta}_{JS},
\]

where

\[
\hat{\beta}_{JS} = (1 - \frac{as^2}{\hat{\beta}_{GLS}'\Sigma_{ZZ}^{-1}X\hat{\beta}_{GLS}})\hat{\beta}_{GLS},
\]

and

\[
0 \leq a \leq \frac{2[tr((X'\Sigma_{ZZ}^{-1}X)^{-1}Q)\lambda_1^{-1} - 2]}{(n - p + 2)},
\]
where $\lambda_1$ is the largest eigenvalue of $(X'\Sigma_{zz}^{-1}X)^{-1}Q$; then

$$r(p_{11}, \frac{X}{0}, \frac{X}{0}, \sigma^2) \leq r(p_3, \frac{X}{0}, \frac{X}{0}, \sigma^2),$$

for all $(\beta, \sigma^2) \in \mathbb{R}^{p \times (0, \infty)}$.

**PROOF:**

Take $h(x) = a/x$, $a \in \mathbb{R}$, $B = (X'\Sigma_{zz}^{-1}X)$, and $C = I_p$ and apply Theorem 4.1.

Expressions for the bias and the risk may be derived for any member of the class $p_{10}(Z)$ (and hence for $p_{11}(Z)$), once $h(\cdot)$ has been specified. Straightforward extensions of theorems in Appendix B of Judge and Bock (1978) give

$$E(p_{11}(Z) - Z_0) = -a(X_0 - \Sigma_0^2\Sigma_{zz}^{-1}X)E(\chi^2_{n-p}/\chi^2_{p+2, \lambda})\hat{\beta}, \quad (4.21)$$

and

$$r(p_{11}, \frac{X}{0}, \frac{X}{0}, \sigma^2) = tr(\Sigma_{00} - \Sigma_0^2\Sigma_{zz}^{-1}\Sigma_0 Z_0)
+ \text{tr}(Q(X'\Sigma_{zz}^{-1}X)^{-1})E(1 - aX^2_{n-p}/X^2_{p+4, \lambda})^2
+ \beta'Q\beta[E(1 - aX^2_{n-p}/X^2_{p-4, \lambda})^2
+ 2E(aX^2_{n-p}/X^2_{p+2, \lambda}) - 1], \quad (4.22)$$
where \( \chi^2_{(*, \lambda)} \) is a non central chi-squared random variable with noncentrality \( \lambda = \beta' X \Sigma_{ZZ}^{-1} X \beta / 2\sigma^2 \), independent of the central chi-squared random variable \( \chi^2_{(n-p)} \).

The expressions (4.21) and (4.22) may be evaluated using the computational formulas for inverse moments of a noncentral chi-squared distribution given in Bock et al. (1984) and Xie (1988). Of course, both expressions depend on the unknowns \( \beta \) and \( \sigma^2 \), but estimators based on \( \hat{\beta}_{\text{GLS}} \) and \( s^2 \) can be used.

It is known that the positive part of \( \hat{\beta}_{\text{JS}} \) (Baranchik 1964),

\[
\hat{\beta}^+_{\text{JS}} = \left[ 1 - \frac{s^2}{\hat{\beta}_{\text{GLS}}' X' \Sigma_{ZZ}^{-1} X \hat{\beta}_{\text{GLS}}} \right] \hat{\beta}_{\text{GLS}}, \tag{4.23}
\]

where \( g^+ = \max(0, g) \), dominates \( \hat{\beta}_{\text{JS}} \), i.e.,

\[
E((\hat{\beta}^+_{\text{JS}} - \beta)' W (\hat{\beta}^+_{\text{JS}} - \beta)) \leq E((\hat{\beta}_{\text{JS}} - \beta)' W (\hat{\beta}_{\text{JS}} - \beta)), \tag{4.24}
\]

for all \((\beta, \sigma) \in \mathbb{R} \times (0, \infty)\), and any positive definite weight matrix \( W \); for a proof of this see Judge and Bock (1978), p. 239. Consequently, with \( Q \) playing the role of \( W \) in (4.24), the predictor

\[
P_{12}(Z) = P(Z, \hat{\beta}^+_{\text{JS}}) = \Sigma_0 Z^{'} \Sigma_{ZZ}^{-1} Z + (X_0 - \Sigma_0 Z^{'} \Sigma_{ZZ}^{-1} X) \hat{\beta}^+_{\text{JS}} \tag{4.25}
\]
must necessarily have smaller risk (as given in (1.6)) than that of \( P_{11}(Z) \) in (4.18).

Expressions for the risk of the estimator \( \hat{\beta}^+_{JS} \), and hence for the risk of the predictor \( P_{12}(Z) \), are intractable. However, (4.22) could be used instead, since it would give a conservative value of the risk of \( P_{12}(Z) \).

When the conditions of Theorem 4.1 are satisfied, the family of predictors \( P_{10}(Z) \) in (4.16) has uniformly smaller risk than that of the best linear unbiased predictor. This result could have important implications for spatial processes: when the goal is the prediction of many points simultaneously, the family of predictors (4.16) improves uniformly upon the universal kriging predictor.

The predictors \( P_{10} \) have a nice interpretation as shrinkage predictors. Since \( \hat{\beta}_{JB} \) shrinks the generalized least squares estimator \( \hat{\beta}_{GLS} \) towards zero, the corresponding predictors in (4.16) shrink the best linear unbiased predictor \( P_3(Z) \) to the best linear predictor under the assumption that \( \beta \) is known and equal to \( \phi \); i.e.,

\[
P_{10}(Z) = (I_p - \mathbf{h}(\hat{\beta}_{GLS}^r \hat{\beta}_{GLS}/s^2)') \hat{\beta}_{GLS} P_3(Z) - \Sigma_{0Z} \Sigma_{ZZ}^{-1} Z)
+ \Sigma_{0Z} \Sigma_{ZZ}^{-1} Z.
\]

Throughout this development of minimax prediction we have chosen to focus on the family of minimax estimators developed by Judge and Bock (1978) simply because the conditions for minimaxity are concise and easy to use, and because the familiar James-Stein type estimators are members of this class. However, our results are not limited to the consideration
of this one class. Other minimax predictors can be constructed using the minimax estimators developed by, e.g., Berger (1976) and Strawderman (1978). In practical situations, the choice of the estimator (and hence the predictor) is a difficult one; see Berger (1982) for some interesting ideas on minimax estimator selection.

If the condition on the constant $a$ in (4.20) is not met, we recommend using the universal kriging predictor given in (2.13). However, if this condition is satisfied, then the predictor $\mathcal{E}_{12}(Z)$ has risk which is uniformly smaller than that of the best linear unbiased predictor and is the predictor of choice among those presented in this paper.
5. SUMMARY

Assuming a generalized linear model with known covariance matrix, several linear and nonlinear predictors are presented and their properties are discussed. In the context of simultaneous multiple prediction, a total sum of squared errors is suggested as a loss function for comparing predictors. Based on a fundamental relationship between prediction and estimation, a very general class of predictors is developed from which predictors with uniformly smaller risk than that of the classical best linear unbiased (i.e., universal kriging) predictor can be constructed.
6. ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under grant number DMS-8902812.
7. REFERENCES CITED


SECTION II. A SPATIAL ANALYSIS OF VARIANCE
A SPATIAL ANALYSIS OF VARIANCE

by

Carol A. Gotway and Noel Cressie

Department of Statistics, Iowa State University, Ames, IA 50011

Supported in part by The National Science Foundation under grant number DMS-8703083.
ABSTRACT

A spatial analysis of variance uses the spatial dependence among the observations to modify the usual inference procedures associated with a statistical linear model. When spatial correlation is present, the usual tests for presence of treatment effects may no longer be valid, and erroneous conclusions may result from assuming that the usual F-ratios are F-distributed. This is demonstrated by extending the spatial analysis of soil-water infiltration data presented in Cressie and Horton (1987). Emphasis is placed on modeling the spatial dependence structure using geostatistical techniques. This spatial dependence structure is then used to test hypotheses about fixed effects in a nested linear model.
1. INTRODUCTION

At a given location in the field, the ability of water to infiltrate soil depends upon the existing soil-water distribution with depth, the rate of water application to the soil surface, and the soil-pore-structure distribution with depth. As the location varies across the field, this ability will vary spatially so that locations nearby are more alike with regard to infiltration than those far apart. This spatial dependence among the infiltration measurements may be used to enhance any statistical analysis of soil-water infiltration. Moreover, failure to account for spatial correlation, in general, can lead to erroneous inference procedures that could result in incorrect scientific conclusions.

In what is to follow, we summarize the data, methodology, and results from the robust-resistant spatial analysis of soil-water infiltration data presented in Cressie and Horton (1987). The spatial correlations among the soil-water infiltration measurements are modeled using geostatistical methods; kriging and cross-validation techniques are implemented to check and adjust for outliers. Finally, using a nested linear model with covariances determined by the modeled spatial correlations, various statistical hypotheses of interest are tested, and the consequences of overlooked spatial dependence are demonstrated.
2. GEOSTATISTICS AND KRIGING

In this section we present a brief overview of geostatistical theory and ordinary kriging. For complete details, the reader is referred to the reference books of Journel and Huijbregts (1978), or Clark (1979).

First, consider the stochastic process

\[ \{Z(s): s \in D \subset \mathbb{R}^d\}, \quad (2.1) \]

which varies continuously over some domain of interest D (usually D is a subset of two-or-three dimensional Euclidean space). Assume that the quantity

\[ 2\gamma(h) = \text{var}(Z(s + h) - Z(s)) \quad (2.2) \]

is a function only of h, the separation between s + h and h;

\( (2\gamma(h): h \in \mathbb{R}^d) \) is called the variogram (Matheron 1963), and

\( (\gamma(h): h \in \mathbb{R}^d) \) is called the semivariogram. For a summary of the properties of the variogram, see Cressie (1988). If we assume further that \( Z(s) \) has a constant mean:

\[ E(Z(s)) = \mu \quad (2.3) \]

(where \( E \) denotes the expectation operator), then the process given in (2.1) is called **intrinsically stationary**.
Under the framework of intrinsic stationarity, several variogram estimators have been proposed. For completeness and clarification, we present two such estimators here.

Suppose that we sample the process at \( n \) spatial locations \( \{s_i; i = 1, \ldots, n\} \) and thus obtain data \( \{Z(s_i); i = 1, \ldots, n\} \). Then the method-of-moments estimator of the variogram (Matheron 1963) is

\[
2\gamma(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2 ; \ h \in \mathbb{R}^d , \tag{2.4}
\]

where the sum is taken over \( N(h) = \{(s_i, s_j); s_i - s_j = h\} \) and \(|N(h)|\) is the number of distinct elements in the set \( N(h) \). A robust version of (2.4) is given in Cressie and Hawkins (1980):

\[
2\gamma(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left|\frac{Z(s_i) - Z(s_j)}{|Z(s_i) - Z(s_j)|} \right|^{1/2^4} ; \ h \in \mathbb{R}^d . \tag{2.5}
\]

This estimator automatically downweights contaminated data and may be preferable to that of (2.4) if the data are very noisy or contain extreme observations.

From the empirical variogram, a smoothed version is obtained by fitting a model to the estimated points. Journel and Huijbregts (1978) give an extensive list of valid semivariogram models. Once the functional form of the model is chosen (e.g., exponential, spherical, generalized linear), its parameters must be estimated. Zimmerman and
Zimmerman (1989) summarize and compare several methods of semivariogram parameter estimation. They find that no method dominates, but their simulation results indicate that a weighted least squares approach due to Cressie (1985) usually performs very well and never does poorly.

Often in geostatistics, the goal is to predict an observation, say $Z(s_0)$, at some fixed, known, but unsampled spatial location $s_0$ using data $\{Z(s_i): i = 1, \ldots, n\}$. The (ordinary) kriging procedure assumes that the predictor $\hat{Z}(s_0)$ is a linear function of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i). \quad (2.6)$$

The weights $\{\lambda_i: i = 1, \ldots, n\}$ are chosen so that $\hat{Z}(s_0)$ is unbiased (i.e., $E(\hat{Z}(s_0)) = E(Z(s_0)) = \mu$), and so that the mean-squared-prediction error, $E(\hat{Z}(s_0) - Z(s_0))^2$ is minimized. For a complete discussion of kriging and the properties of the kriging predictor, the interested reader is referred to Journel and Huijbregts (1978). In particular, the kriging weights satisfy the kriging equations

$$\Gamma \lambda = \gamma, \quad (2.7)$$

where $\lambda = (\lambda_1, \ldots, \lambda_n)^\prime$, $\gamma = (\gamma(s_1 - s_0), \ldots, \gamma(s_n - s_0), 1)^\prime$, $\Gamma$ is an $(n+1) \times (n+1)$ symmetric matrix whose $(i,j)$-th entry is

$$\gamma(s_i - s_j); \quad i = 1, \ldots, n, \quad j = 1, \ldots, n$$

$$1; \quad i = n + 1, \quad j = 1, \ldots, n$$

$$0; \quad i = n + 1, \quad j = n + 1$$
and \( m \) is a Lagrange multiplier to ensure that \( \sum_{i=1}^{n} \lambda_i = 1 \). The minimized root-mean-squared-prediction error, called the kriging standard error, is given by

\[
\sigma(s_0) = \left( \sum_{i=1}^{n} \lambda_i (s_i - s_0) + m \right)^{1/2}.
\]  

(2.8)

The geostatistical techniques described above will be used to compare the effect of tillage on soil-water infiltration. A description of the experiment, the preliminary robust-resistant analysis of Cressie and Horton (1987), and the subsequent variogram model fitting are given in the following section.
3. ESTIMATING SPATIAL DEPENDENCE: VARIOGRAM ESTIMATION AND MODEL FITTING

The variable of interest in this study is soil-water infiltration, as computed from a double-ring infiltrometer apparatus. The double-ring infiltrometer is a device consisting of two concentric rings: the outer ring is used to stop the horizontal spread of the water so that only the vertical subsidence is measured, and the other is used to pond the water so that the infiltration rate can be measured. An experiment was conducted in the summer of 1983 to determine the effects of tillage treatment on soil-water infiltration. Several plots of the experiment were set aside for a more detailed spatial analysis; the plots were plowed in the fall of 1982 using the following tillage treatments: moldboard-plow (15-20 cm), paraplow (25-30 cm), chisel-plow (15-20 cm), and no-tillage. For more details see Mukhtar et al. (1985) and Cressie and Horton (1987).

Water stage recorders were used to record soil-water infiltration as a function of time (Mukhtar et al. 1985). For the part of the experiment of interest to us here, thirty-minute cumulative infiltration measurements (in cm) were made at twenty-four locations (on a 3x8 grid arrangement) within each plot. Two sets of infiltration measurements were obtained, one in May and one in July, but we will analyze only the July data here. Figure 1 (from Cressie and Horton 1987) illustrates the arrangement of the spatial sites and

Figure 1 here
and the tillage treatments. Due to limited resources, only one block of a randomized block design was used for the spatial experiment. This poor design makes any conclusions tentative since strictly speaking treatment and plot location are confounded; further details are given below. However, there are a number of instances in science where, even with unlimited resources, replicated designs are an impossibility, and comparison of properties among different units may still be desired (e.g., comparison of lithological characteristics among rock units in a formation). At the very least, the analysis that follows provides an illustration of how to carry out a spatial analysis of variance.

To begin the spatial analysis of soil-water infiltration, write the data in Figure 1 as

\[(y_{ijk} : i = 1,2,3,4, j = 1,2,3, k = 1,\ldots,8) , \quad (3.1)\]

so that \(y_{ijk}\) is the \(k\)-th observation in the \(j\)-th column of treatment \(i\). Arbitrarily, set \(i = 1\) for the moldboard tillage treatment, \(i = 2\) for paraplow, \(i = 3\) for chisel, and \(i = 4\) for no-till. Using robust-resistant exploratory spatial data-analysis procedures, Cressie and Horton (1987) show that in order to estimate the spatial dependence in the data, a symmetrizing and variance stabilizing square-root transformation is needed, followed by subtraction of column medians to remove trend. Specifically, define
where $\tilde{z}_{ij} = \text{med}(z_{ijk}; k = 1, \ldots, 8)$. The square-root transformation was applied so that on this scale, data can be written as a mean effect (made up of additive components of spatial location and treatment effects) plus homoskedastic (within plots) Gaussian error. Figure 2 shows stem-and-leaf plots of the median-based residuals $r_{ijk}$.

These residuals now appear to have come from a trend-free process (Cressie and Horton 1987 demonstrate lack of trend in the east-west direction) but are not homoskedastic between plots even after the square-root transformation. (As we shall see below, treatments $i = 1$ and $i = 2$ show much more error variation than treatments $i = 3$ and $i = 4$.) However, we may use the median-based residuals of (3.3) to estimate a stationary covariance structure within each treatment.

From the discussion above we can write

$$z_{ijk} = (y_{ijk})^{1/2} \quad (3.2)$$

$$r_{ijk} = z_{ijk} - \tilde{z}_{ij} \quad (3.3)$$
where $\mu_{ijk}$ is the mean of the $k$-th observation in the $j$-th column using treatment $i$, and $\delta = (\delta_{1,1,1}, \delta_{1,1,2}, \ldots, \delta_{4,3,8})'$ is a realization of a 96x1 vector of random errors with mean zero and covariance matrix $\sigma^2 \Sigma$. Thus the spatial correlation among the soil-water infiltration measurements is portrayed through $\Sigma$, and will be estimated and modeled through the variogram (2.2).

Since the spacing in the east-west direction is 3 meters, and that in the north-south direction is 1.5 meters, it is convenient to write the classical variogram estimator of (2.4) in the east-west direction as:

$$2\gamma_i(2ah) = \frac{3}{|N(h)|} \sum_{j=1}^{8-h} \sum_{k=1}^{8-h} (r_{i,j,k+h} - r_{i,j,k})^2 / |N(h)| ; h = 1, \ldots, 7, \quad (3.5)$$

where $|N(h)| = 3(8-h)$ and $a = 1.5m$. The robust version corresponding to (3.5) is

$$2\overline{\gamma}_i(2ah) = \frac{\left\{ \left\{ \sum_{j=1}^{8-h} \sum_{k=1}^{8-h} |r_{i,j,k+h} - r_{i,j,k}|^{1/2} / |N(h)| \right\}^4 \right\}^{1/4}}{0.457 + 0.494/|N(h)|} ; h = 1, \ldots, 7. \quad (3.6)$$

Both (3.5) and (3.6) were computed for the soil-water infiltration data, but for robustness reasons (see Section 4) only those for $2\overline{\gamma}_i(\cdot)$ are presented here. Figure 3 gives plots for each of the treatments up to a lag distance of 15 meters.
Spatial dependence, as summarized by the estimated variograms, clearly changes with treatment. It is most pronounced for moldboard; among the four treatments, moldboard is the plowing technique that causes the greatest soil disturbance.

Semivariogram models were fit to the robust empirical variograms using a two-stage fitting based on the weighted least squares algorithm described by Cressie (1985). After choosing an appropriate class of models, one first fits a variogram slope through the estimates near the origin; the rest of the variogram model is then fit fixing that estimated slope (see Cressie et al. 1989). Since chisel and no-till semivariogram estimators were strikingly similar, they were pooled, and a semivariogram model was fit to the combined data. The following semivariogram models for the east-west direction were fit:

\[
\gamma_1(h) = \begin{cases} 
3.0308 \left( \frac{3}{2} \frac{h}{17.2980} \right) - \left( \frac{1}{2} \frac{h}{17.2980} \right)^3 & ; 0 \leq h \leq 17.29 \\
3.0308 & ; h \geq 17.29 
\end{cases} 
\] (3.7)

(moldboard)

\[
\gamma_2(h) = \begin{cases} 
0 & ; h = 0 \\
1.662 & ; h > 0 
\end{cases} 
\] (3.8)

(paraplow)
\[
\gamma_3(h) = \begin{cases} 
0 & ; h = 0 \\
0.2881 & ; h > 0
\end{cases} 
\] (3.9) (chisel)

\[
\gamma_4(h) = \begin{cases} 
0 & ; h = 0 \\
0.2881 & ; h > 0
\end{cases} 
\] (3.10) (no-till).

Figure 4 illustrates each of these models superimposed on the empirical semivariograms. Since each model has a sill \( \gamma_1(\infty) \), the spatial dependence may equally be described through a stationary covariance function

\[
G(h) = \sigma_i^2 - \gamma_1(h); \ h \geq 0 ,
\] (3.11)

where \( \sigma_i^2 = \gamma_1(\infty) \).

Very few lags were available in the north-south directions from which to estimate the error variograms. Those that were, justified an isotropy assumption for the spatial dependence within each plot. To examine the spatial dependence between neighboring plots, sample correlation coefficients were computed. They indicated a lack of
dependence, leading to an (estimated) covariance model for $\Sigma$ in (3.4),
given by (3.12) below. Let the data be a realization of
$Z' = (Z_1', Z_2', Z_3', Z_4')$, where $Z_1' = (Z_{i, 1, 1}, Z_{i, 1, 2}, \ldots, Z_{i, 1, 8})$. Then, our
(estimated) model for the covariance structure of $Z$ is

$$\text{var}(Z) = \sigma^2 \begin{bmatrix} \Sigma_1 & \phi & \phi & \phi \\ \phi & \Sigma_2 & \phi & \phi \\ \phi & \phi & \Sigma_3 & \phi \\ \phi & \phi & \phi & \Sigma_4 \end{bmatrix} = \sigma^2 \Sigma, \quad (3.12)$$

a 96x96 block diagonal matrix where each block is 24x24. Matrices $\Sigma_2$, $\Sigma_3$, $\Sigma_4$ are proportional to the identity matrix $I_{24}$: $\Sigma_2 = 1.662 \, I_{24}$, corresponding to paraplow, and $\Sigma_3 = \Sigma_4 = 0.2881 \, I_{24}$, corresponding to chisel and no-till. Only the matrix $\Sigma_1$, corresponding to the moldboard tillage treatment shows spatial dependence: $\Sigma_1$ is made up of elements obtained from the stationary covariance function

$$C(h) = 3.0308 - \gamma_1(h), \quad (3.13)$$

where $\gamma_1(h)$ is given by (3.7). The notation $\phi$ is used to represent a matrix (of any order) with zero entries.

In a later section, we shall proceed with inference on the mean effects $(\mu_{ijk})$ of model (3.4). To do this we shall assume the model for $\text{var}(Z)$ given by (3.12), however, one should not forget that it has in fact been obtained by model fitting to empirical variograms. Moreover,
the data set is not a large one. That is why we have included in (3.2) a proportionality constant $\sigma^2$, to be determined by the data. Clearly, more research is needed to determine the effects of using fitted variance matrix parameters as if they were known; this remains an open problem for determining such an effect on kriging standard errors as well.
Having modeled a semivariogram for each of the four treatments, it is necessary to check for model lack-of-fit. We confine ourselves here to an evaluation of the semivariogram model for the moldboard data (see (3.7)).

In geostatistics, a technique frequently used in checking the fit of a semivariogram model is the cross-validation criterion (Stone 1976) applied to the kriging predictor.

The basic idea is as follows:

**Step 1:** Delete an observation, \(Z(s_i)\), and then predict it from the remaining data points \(\{Z(s_j): j \neq i\}\) using the kriging predictor (2.6). Thus obtain \(\hat{Z}_{-i}(s_i)\).

**Step 2:** Compute

\[
Z^o_i = \frac{\hat{Z}_{-i}(s_i) - Z(s_i)}{\sigma_{-i}(s_i)},
\]

(4.1)

where \(\sigma_{-i}(s_i)\) is the kriging standard error of \(\hat{Z}_{-i}(s_i)\) (see (2.8)).

**Step 3:** Repeat steps 1 and 2, systematically deleting and then predicting each of \(\{Z(s_i): i = 1, \ldots, n\}\) in turn, and computing the corresponding standardized statistics \(Z^o_i: i=1, \ldots, n\).
Then the sample mean and sample variance of \(Z_i^0: i = 1, \ldots, n\) should be approximately zero and one, respectively. Deviation from this indicates lack-of-fit of the semivariogram model. Moreover, if the underlying process is Gaussian, \(Z_i^0\) is also Gaussian so that "spatial outliers" could be detected by comparing the histogram of \(\{Z_i^0: i = 1, \ldots, n\}\) to that of a standard normal density.

Using the twenty-four (median-swept) observations from the moldboard plot given by (3.3), cross-validation was performed using the semivariogram model of (3.7). In this case, write the cross-validation statistic as,

\[
\hat{r}_{1jk} = \frac{\hat{r}_{1jk} - r_{1jk}}{\sigma_{1jk}},
\]

where \(\hat{r}_{1jk}\) is the kriging predictor (see (2.6)) of \(r_{1jk}\) using the other twenty-three detrended observations, and \(\sigma_{1jk}\) is its associated kriging standard error (see (2.8)). A stem-and-leaf plot of the resulting \(\{r_{1jk}^0: j = 1, 2, 3, k = 1, \ldots, 8\}\) is presented in Figure 5, and a normal probability plot of \(r_{1jk}^0\) (see, e.g., Barnett 1975) is shown in Figure 6.
Note from these diagrams that the fit of the semivariogram model is adequate, and would be quite good if it were not for the extreme point $r^o_{1,3,7} = 3.70$. This value is large because of the large difference between the observed value and the predicted value (the kriging standard errors are all very similar). Since the kriged value gives more weight to nearest neighbors, one reason for this large difference is that $r^o_{1,3,7}$ (and $z^o_{1,3,7}$, from (3.2)) is large relative to its nearest neighbors. These "spatial outliers" are hard to detect as extreme or unusual observations in a histogram of the data since it ignores the (relative) spatial locations of the observations; i.e., the histogram is insensitive to spatial information in the data.

As with any outlier, spatial outliers should never be deleted without good reason (Anscombe 1960), although extreme observations will destroy even the most robust statistical analysis. A compromise is Winsorization (see Huber 1979, and Hawkins and Cressie 1984).

Winsorization is a data-editing technique where unusual observations are not deleted, but are replaced by a less extreme version. That is, replace $z_{ijk}$ with

$$z^{(s)}_{ijk} = \text{med}(z_{ijk}: k = 1, \ldots, 8) + \begin{cases} \hat{r}_{ijk} + c\sigma_{ijk} & \text{if } r^o_{ijk} > c \\ r_{ijk} & \text{if } |r^o_{ijk}| \leq c \\ \hat{r}_{ijk} - c\sigma_{ijk} & \text{if } r^o_{ijk} < -c \end{cases} \quad (4.3)$$

where $\hat{r}_{ijk}$ is the kriging predictor with kriging standard error $\sigma_{ijk}$, $r^o_{ijk}$ is the cross-validation statistic (see (4.2)) and $c$ is a tuning parameter.
constant controlled by the user. Common values for $c$ lie in the range of 1.5 to 3.0; the smaller the value of $c$, the more the data tend to be edited ($c = \infty$ corresponds to no editing at all).

Actually, if the data are Gaussian, a normal probability plot gives a nice way of obtaining $c$ adaptively: From the probability plot, choose $c$ to be the X-coordinate of the unusual point in question, which moves the point to the target 45° line. Applying this to the spatial outlier of the moldboard data, we obtain $c = 2.04$. However, since there is some deviation about the normal line, we chose a less severe $c = 2.5$. Then $z_{1,3,7} = 6.19$ is replaced by $z_{1,3,7}^{(s)} = 3.99 + (- 0.5106 + (2.5)(.7314))$

$$= 5.31 \text{ cm}^{1/2},$$

which back on the original scale gives $y_{1,3,7}^{(s)} = 28.22 \text{ cm}$ (see (3.1)). This can be compared with $y_{1,3,7} = 38.31 \text{ cm}$. From checking the experimental records, no reason for this outlier was apparent; a large subsurface crack might account for the higher-than-expected infiltration rate.

In the analysis to follow the data will be edited according to the computations above, and for notational convenience we now drop the superscript "(s)." Consequences of working with unedited data will be illustrated briefly in the next section.
5. INFERENCE ON MEAN EFFECTS

One of the goals of this paper is to formulate valid and efficient inference procedures for the mean function, \( \mu_{ijk} \) of the model in (3.4). Based on the exploratory spatial data analysis of Cressie and Horton (1987), the following additive model seems appropriate:

\[
Z_{ijk} = \mu + t_i + \beta_{ij} + \epsilon_{ijk},
\]

where

- \( Z_{ijk} \) is the square-root of the \( k \)-th datum in the \( j \)-th column of treatment \( i \),

- \( \mu \) represents the overall average soil-water infiltration,

- \( t_i \) is an effect due to treatment \( i \),

- \( \beta_{ij} \) is an effect associated with the \( j \)-th column of treatment \( i \),

and

- \( \epsilon_{ijk} \) is random error with zero mean and covariance matrix \( \sigma^2 \Sigma \) given by (3.12).

This model is general enough to account for north-south trend and differential treatment effects; east-west trends are assumed negligible (see Section 3).
It is convenient to write the model in (5.1) using matrix notation:

\[ Z = X\beta + \epsilon, \quad (5.2) \]

where

\[ Z \text{ is the } 96 \times 1 \text{ data vector,} \]
\[ Z = (Z_{1,1}, Z_{1,2}, \ldots, Z_{1,1,8}, Z_{1,2,1}, \ldots, Z_{4,3,8})', \]
\[ \beta = (\mu, t_1, \ldots, t_4, \beta_{1,1}, \beta_{1,2}, \beta_{1,3}, \beta_{2,1}, \ldots, \beta_{4,3})', \]
\[ X \text{ is an incidence matrix of 0's and 1's (see Searle 1971, p. 145), that specifies } E(Z_{ijk}) = \mu + t_i + \beta_{ij}, \]
\[ \epsilon \text{ is the associated vector of random errors with variance} \]
\[ \text{matrix, } \text{var}(\epsilon) = \sigma^2 \Sigma \text{ given by (3.12).} \]

5.1. Estimation of main effects

The first step towards inference on \(X\beta\), is to specify an estimation procedure. If spatial dependence is ignored, or overlooked (as is often the case), the ordinary least squares estimator of \(X\beta\), namely

\[ \hat{X}_\text{OLS}^\beta = X(X'X)^{-1}X'Z, \quad (5.3) \]

might be used. A stem-and-leaf plot of the ordinary-least-squares residuals from fitting the model (5.2), (with the one outlier Winsorized; see Section 4), is presented in Figure 7a.
Although the shape of this stem-and-leaf plot appears to be Gaussian, a corresponding residual plot (Figure 7b) shows that the variability of the residuals increases with increasing mean, and thus suggests that a weighted estimation procedure is necessary.

In the case of the soil-water infiltration data, because of the heteroskedasticity and the spatial dependence, a generalized least squares estimator of \( \hat{X\beta} \), namely

\[
\hat{X\beta}_{\text{GLS}} = X(XX'X)^{-1}X'Z',
\]

is more appropriate (Rao 1973, Chapter 4). Now the stem-and-leaf plot of the residuals from the generalized least squares fit (Figure 8a) looks to be roughly Gaussian (although somewhat more granular than Figure 7a), and the associated residual plot (Figure 8b) does not suggest carrying out any further transformation or weighting. Comparing the residual plots in Figures 7b and 8b we see that without the weighting, differences in the estimated means for each treatment are masked, whereas after the weighting the treatment divisions are more clearly defined. This is
particularly true for expected values corresponding to the moldboard data (recall that plots receiving the moldboard treatment are the only ones with significant spatial dependence). In Figure 8b, the residuals from the moldboard plots are those with the smallest expected values.

It is interesting to see how the Winsorization of the data has improved the weighted estimation of $X\hat{\beta}$. Figure 9 shows a stem-and-leaf plot of residuals obtained from the generalized least squares estimator (5.4) with the original (unedited) square-root data given in equation (3.2). The one very large residual of $3.23 \, \text{cm}^{1/2}$ can be traced back to $y_{1,3,7} = 38.31 \, \text{cm}$, the same value that gave a large value for the cross-validation statistic in Section 4. Residuals obtained from fitting with the unusual observation deleted (not shown here) behave very much like the residuals obtained using the edited values. However, since Winsorization uses some of the information contained in the original data point, it offers a satisfactory compromise.
5.2. A spatial analysis of variance

From the generalized least squares approach to estimation of the mean parameters in (5.1), we may write a general analysis of variance:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>IJ - 1</td>
<td>SS (model)</td>
</tr>
<tr>
<td>treatments</td>
<td>I - 1</td>
<td>SS (treatments)</td>
</tr>
<tr>
<td>columns in treatments</td>
<td>I(J - 1)</td>
<td>SS (columns in treatment)</td>
</tr>
<tr>
<td>residual</td>
<td>N - IJ</td>
<td>SS (residual)</td>
</tr>
<tr>
<td>corrected total</td>
<td>N - 1</td>
<td>SS (corrected total)</td>
</tr>
</tbody>
</table>

where

\[ SS_V(\text{model}) = \hat{\beta}_{\text{GLS}} X' V^{-1} Z - m_V, \] (5.5)

\[ SS_V(\text{columns in treatments}) = SS_V(\text{model}) - SS_V(\text{treatments}), \] (5.6)

\[ SS_V(\text{residual}) = Z' V^{-1} Z - \hat{\beta}_{\text{GLS}} X' V^{-1} Z, \] (5.7)

\[ SS_V(\text{corrected total}) = Z' V^{-1} Z - m_V, \] (5.8)

\[ \hat{\beta}_{\text{GLS}} = (X' V^{-1} X)^{-1} X' V^{-1} Z, \]

\[ m_V = (Z' V^{-1} Z)^{-1} (Z' V^{-1} Z) ; \] (5.9)
V is an N×N positive-definite matrix; \( \mathbf{1} \) is an n×1 vector of 1's; SS\(_V\)(treatments) is equal to an SS\(_V\)(model) type of expression where the model is now \( Z_{ijk} = \mu + t_i + \epsilon_{ijk} \), fit by generalized least squares; I(\(-4\)) denotes the number of treatments, J(\(-3\)) is the number of columns within each treatment, and N(\(-96\)) is the total number of observations.

Consider now an analysis of variance (ANOVA) table for each of three models:

(a) the **full spatial model** where \( V = \sigma^2 \Sigma \), given by (3.12); this model incorporates heteroskedasticity among the plots as well as spatial dependence,

(b) the **heteroskedastic model** where \( V = \text{diag}(\sigma^2 \Sigma) \), and \( \Sigma \) is given by (3.12); this model allows only for unequal variances between plots and ignores spatial correlation, and

(c) the **classical model** with \( V = \sigma^2 \mathbf{I} \); this model assumes independence and homoskedasticity between plots.

Tables la, lb, and lc, give the analysis of variance table for each of these models. Notice the similarity in the decomposition of the sum of squares for the heteroskedastic and classical models, and the difference between these decompositions and that associated with the full spatial model. The model-fitting and cross-validation carried out in previous sections indicates that the full spatial model is more appropriate than the other two.

In the following sections, hypothesis tests for the parameters of the model (5.1), are developed. Differences between analyses based on
the full spatial model and the nonspatial models will again be the most marked.

5.3. Testing the hypothesis of equality among columns within treatments

One common hypothesis frequently tested in an analysis of variance is the hypothesis of equal treatment means. However, in the spatial context, this may not be a well-formulated hypothesis if there is spatial trend, i.e., if the mean depends on spatial location and is not constant from plot to plot. Thus we should check first the assumption of no spatial trend within each treatment. From (5.1), this amounts to checking for constant column means; specifically, test

\[ H_0: \beta_{11} = \beta_{12} = \beta_{13} \]
\[ \beta_{21} = \beta_{22} = \beta_{23} \]
\[ \beta_{31} = \beta_{32} = \beta_{33} \]
\[ \beta_{41} = \beta_{42} = \beta_{43} \]

against the general alternative (5.1). This hypothesis may be tested by computing the ratio

\[ F = \frac{SS_Y(\text{columns in treatments})/I(J-1)}{SS_Y(\text{residual})/(N-IJ)} \]

(5.11)
(the sums of squares may be found in the ANOVA table), and comparing it to an F-distribution on \((I(J-1), N-IJ)\) degrees of freedom. Note that if spatial correlation is present, but ignored, then an F-statistic like (5.11), but based on ordinary least squares, does not have an F-distribution. Computing the ratio using the appropriate values from each of Tables 1a, 1b, and 1c, we obtain \(F = 9.80\) for the full spatial model, \(F = 7.90\) for the heteroskedastic model and \(F = 6.27\) for the classical model.

Comparing the first of these numbers \(F = 9.80\) to an F-distribution (actually, the only comparison that is valid) with 8 and 84 degrees of freedom, we see that this value is significant at the 0.01 level, leading us to reject the null hypothesis of constant column means within treatments. Therefore, we conclude that there is significant spatial trend. Notice that the F-ratio for the classical model is much lower than that for the full spatial model. Consequently, in similar problems where the F-ratio for the full spatial model is significant and that of the classical is not, real differences could be overlooked if spatial dependence is not considered.

5.4. Testing the hypothesis of equality of average treatment effects

Now that the hypothesis of constant column mean within a treatment has been rejected, we can compare treatment-plot effects by averaging over the columns within each treatment and testing the hypothesis
against the general alternative (5.1). Since treatment effects are
confounded with location, rejection of $H_0$ may be due to a difference in
treatment effects or may be due to a difference in spatial locations.

To test the hypothesis in (5.12), we refer to the general analysis
of variance table at the beginning of this section and use

$$F = \frac{SS_Y(\text{treatments})/(I-1)}{SS_Y(\text{residual})/(NIJ)} \quad (5.13)$$

Computing the ratio for the full spatial model, the heteroskedastic model
and the classical model, we obtain 28.05, 44.95, and 43.95 respectively.

Notice that the values of the F-ratio for the heteroskedastic and
classical models are much larger than that of the full spatial model.
This is because when $V$ fails to account for the spatial correlation,
$SS_Y(\text{residual})$ is much too small. Hence the resulting F-ratios (which are
not F-distributed) are much too large. In general, by assuming (wrongly)
the classical model or the heteroskedastic model when the data are
exhibiting positive spatial dependence, more frequent declarations of
significant treatment differences are obtained than the data warrant.
Using the full spatial model and comparing $F = 28.05$ to an $F$-distribution on 3 and 84 degrees of freedom, we reject the null hypothesis (5.12) and conclude that there are treatment-plot differences.

5.5. Pairwise contrasts

We now look to pairwise contrasts to determine which pairs of treatments are significantly different with regard to the amount of soil-water infiltration. Consider, for example, testing equality of average treatment effects between the moldboard and paraplow treatments. Thus, test

$$H_0: t_1 + \frac{1}{3} \sum_{j=1}^{3} \beta_{1j} = t_2 + \frac{1}{3} \sum_{j=1}^{3} \beta_{2j},$$

(5.14)

against a general alternative, where the vector of mean-effect parameters is now

$$\beta_{1,2} = (\mu, t_1, t_2, \beta_{1,1}, \beta_{1,2}, \beta_{1,3}, \beta_{2,1}, \beta_{2,2}, \beta_{2,3})'.$$

Then using the $48 \times 1$ data vector

$$Z_{1,2} = (Z_{1,1,1}, Z_{1,1,2}, \ldots, Z_{1,3,8}, Z_{2,1,1}, \ldots, Z_{2,3,8})',$$

and by notating $\text{cov}(Z_{1,2})$ as $\sigma_{1,2}^2 \Sigma_{1,2}$, an ANOVA table with $I = 2$, $J = 3$, and $N = 48$ may be constructed. The hypothesis (5.14) may be tested by
computing the associated F-ratio (5.13), where the 48×48 positive definite matrix \( V_{1,2} \) plays the role of \( V \) in (5.5) through (5.9).

In this case, the full spatial model corresponds to taking

\[
V_{1,2} = \sigma^2_{1,2} \Sigma_{1,2} = \sigma^2_{1,2} \begin{pmatrix}
\Sigma_1 & \phi \\
\phi & \Sigma_2
\end{pmatrix},
\]

where \( \Sigma_1 \) and \( \Sigma_2 \) are as in (3.12) and \( \sigma^2_{1,2} \) is a proportionality constant. The heteroskedastic model corresponds to \( V_{1,2} = \text{diag}(\sigma^2_{1,2} \Sigma_{1,2}) \), and the classical model corresponds to \( V_{1,2} = \sigma^2_{1,2} I_{48} \).

Hypotheses similar to that of (5.14) are considered for the other five pairs: moldboard-chisel, moldboard-notill, paraplow-chisel, paraplow-notill and chisel-notill; the data vectors, the mean–effect parameters, and the covariance matrices are defined analogous to that of the moldboard-paraplow contrast above. Table 2 gives the associated values of (5.13) for each of the six contrasts, and for each of the full spatial, heteroskedastic, and classical models. Recall from Section 3 that the moldboard data showed the greatest spatial dependence. From Table 2 we see that the F-ratios for comparison of contrasts involving moldboard using the full spatial model are much lower than those for the heteroskedastic and classical models. Consequently, if we adopted (wrongly) either a heteroskedastic or classical model in the analysis of soil-water infiltration, we would be likely to conclude significant treatment differences, when in fact such differences are not supported by the data. Intuitively, the positive spatial correlation exhibited by the moldboard data reduces the effective number of observations, which in
turn may not allow rejection of a null hypothesis involving moldboard's treatment mean with other treatment means.
Measurements of soil-water infiltration were used to illustrate the effects of overlooked spatial correlation. Spatial dependence was modeled using geostatistical methods, and kriging and cross-validation were used to check model fit and adjust for outliers. A nested linear model with fixed effects was used as a basis for inference procedures. A spatial analysis of variance was proposed and used to test the hypothesis of large-scale trend, as well as the hypothesis of equality of average treatment effects. Because F-ratios do not follow an F-distribution when spatial correlation is present but overlooked, making inferences based on such ratios can lead to erroneous conclusions.
7. ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under grant number DMS-8703083.
8. REFERENCES CITED


*Journal of the International Association for Mathematical Geology*, 17, 563-586.


*Journal of the International Association for Mathematical Geology* 16, 3-18.


Table 1a. Analysis of variance associated with

\[ Z_{ijk} = \mu + t_i + \beta_{ij} + \epsilon_{ijk}, \]

for the full spatial model.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>11</td>
<td>183.23</td>
</tr>
<tr>
<td>treatments</td>
<td>3</td>
<td>95.40</td>
</tr>
<tr>
<td>columns in treatments</td>
<td>8</td>
<td>87.83</td>
</tr>
<tr>
<td>residual</td>
<td>84</td>
<td>94.12</td>
</tr>
<tr>
<td>corrected total</td>
<td>95</td>
<td>277.35</td>
</tr>
</tbody>
</table>
Table 1b. Analysis of variance associated with

\[ Z_{ijk} = \mu + t_i + \beta_{ij} + \epsilon_{ijk}, \]

for the heteroskedastic model

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>11</td>
<td>168.16</td>
</tr>
<tr>
<td>treatments</td>
<td>3</td>
<td>115.52</td>
</tr>
<tr>
<td>columns in treatments</td>
<td>8</td>
<td>52.64</td>
</tr>
<tr>
<td>residual</td>
<td>84</td>
<td>69.99</td>
</tr>
<tr>
<td>corrected total</td>
<td>95</td>
<td>238.15</td>
</tr>
</tbody>
</table>
Table 1c. Analysis of variance associated with

\[ Z_{ijk} = \mu + t_i + \beta_{ij} + \epsilon_{ijk} \]

for the classical model

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>11</td>
<td>157.83</td>
</tr>
<tr>
<td>treatments</td>
<td>3</td>
<td>114.31</td>
</tr>
<tr>
<td>columns in treatments</td>
<td>8</td>
<td>43.52</td>
</tr>
<tr>
<td>residual</td>
<td>84</td>
<td>72.83</td>
</tr>
<tr>
<td>corrected total</td>
<td>95</td>
<td>230.66</td>
</tr>
</tbody>
</table>
Table 2. F-ratios for testing the hypotheses analogous to equation (5.12) for each of the six contrast pairs: M, P, C, N denote moldboard, paraplow, chisel, and no-till respectively. Nominal degrees of freedom are 1 and 42, and $F_{1,42}^{0.05} = 4.07$

<table>
<thead>
<tr>
<th>Model</th>
<th>M-P</th>
<th>M-C</th>
<th>M-N</th>
<th>P-C</th>
<th>P-N</th>
<th>C-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>full spatial</td>
<td>0.14</td>
<td>3.36</td>
<td>1.66</td>
<td>90.73</td>
<td>53.70</td>
<td>22.13</td>
</tr>
<tr>
<td>heteroskedastic</td>
<td>2.00</td>
<td>45.66</td>
<td>23.29</td>
<td>95.67</td>
<td>52.99</td>
<td>22.13</td>
</tr>
<tr>
<td>classical</td>
<td>2.22</td>
<td>68.70</td>
<td>31.88</td>
<td>95.67</td>
<td>52.99</td>
<td>22.13</td>
</tr>
<tr>
<td>MOLDBOARD</td>
<td>PARAPLOW</td>
<td>CHISEL</td>
<td>NO-TILL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.55 27.90 12.50</td>
<td>7.54 36.64 26.47</td>
<td>10.24 8.93 14.77</td>
<td>4.30 9.75 9.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.10 35.45 6.84</td>
<td>5.40 38.82 42.02</td>
<td>6.81 8.55 11.84</td>
<td>6.10 13.41 14.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.05 53.25 13.90</td>
<td>13.43 10.67 20.33</td>
<td>3.99 1.83 7.96</td>
<td>4.48 15.38 10.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.62 39.04 18.15</td>
<td>26.49 30.28 35.20</td>
<td>7.10 4.65 5.32</td>
<td>8.67 15.29 12.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.64 34.14 28.53</td>
<td>39.82 27.52 39.65</td>
<td>2.12 5.29 8.31</td>
<td>3.54 12.56 20.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.65 23.30 25.97</td>
<td>20.19 25.15 44.42</td>
<td>6.02 3.52 5.84</td>
<td>2.22 15.21 13.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.78 18.93 38.31</td>
<td>6.48 31.78 60.04</td>
<td>6.33 4.94 8.29</td>
<td>8.58 8.88 18.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.78 31.29 10.00</td>
<td>16.20 63.32 38.71</td>
<td>8.40 2.53 5.41</td>
<td>10.35 15.32 11.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Thirty-minute cumulative soil water infiltration data (in centimeters) and their spatial locations, together with tillage treatments. Distance between readings is 3 m in the east-west direction and 1.5 m in the north-south direction within tillage treatments and 3 m between adjacent tillage treatments moldboard and paraplow and chisel and no-till; 9 m separates the closest readings associated with paraplow and chisel treatments. (Source: Cressie and Horton 1987)
Figure 2. Stem-and-leaf plot of square-root transformed, median-swept residuals \( r_{ijk} \); 0|1 denotes .1 cm\(^{1/2}\)
Figure 3a. Robust empirical semivariogram in the east-west direction for the moldboard tillage treatment.
Figure 3b. Robust empirical semivariogram in the east-west direction for the paraplow tillage treatment.
Figure 3c. Robust empirical semivariogram in the east-west direction for the chisel tillage treatment.
Figure 3d. Robust empirical semivariogram in the east-west direction for the no-till treatment
Figure 4a. Robust empirical semivariogram in the east-west direction for the moldboard tillage treatment. The superimposed dashed line represents the fitted parametric semivariogram model, fit by weighted least squares (Cressie, 1985)
Figure 4b. Robust empirical semivariogram in the east-west direction for the paraplow tillage treatment. The superimposed dashed line represents the fitted parametric semivariogram model, fit by weighted least squares (Cressie, 1985)
Figure 4c. Robust empirical semivariogram in the east-west direction for both chisel and no-tillage treatments combined. The superimposed dashed line represents the fitted parametric semivariogram model, fit by weighted least squares (Cressie, 1985)
<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>0255</td>
</tr>
<tr>
<td>1</td>
<td>4446677</td>
</tr>
<tr>
<td>0</td>
<td>99763</td>
</tr>
<tr>
<td>-1</td>
<td>222</td>
</tr>
<tr>
<td>-2</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Stem-and-leaf plot of the cross-validation statistic applied to moldboard data; -0|9 denotes -.9 cm$^{1/2}$
Figure 6. Normal probability plot of cross-validation statistic. Horizontal axis denotes expected quantile; vertical axis denotes observed quantile.
Figure 7a. Stem-and-leaf plot of residuals obtained from least squares fitting of model (5.2) using Winsorized data. The stem 2 and corresponding leaves denote data in the interval [0.20, 0.40) cm$^{1/2}$
Figure 7b. Residual plot (residual vs. expected, in cm$^{1/2}$) obtained from ordinary least squares fitting of model (5.1) using the Winsorized data.
Figure 8a. Stem-and-leaf plot for residuals obtained from generalized least-squares fitting of model (5.2) using Winsorized data. The stem 2 and corresponding leaves denote data in the interval \([0.20, 0.40)\) cm\(^{1/2}\)
Figure 8b. Residual plot (residuals vs. expected, in cm\(^{1/2}\)) obtained from generalized least squares fitting of model (5.1) using Winsorized data
Figure 9. Stem-and-leaf plot of residuals obtained from generalized least-squares fitting of model (5.2) using unedited data. The stem 2 and corresponding leaves denote data in the interval \([0.20, 0.40) \text{ cm}^{1/2}\)
This dissertation considers three important statistical aspects of spatial processes: prediction, estimation, and hypothesis testing in the linear model.

The first section contributes to the theory of prediction in a generalized-linear-model setting by considering simultaneous multiple prediction of $Z_0 = (Z_{0,1}, \ldots, Z_{0,k})$. Using the ideas of James and Stein (1961), the loss function of equation (1.5) of Section I was suggested as an index of accuracy for both stochastic and nonstochastic predictors. The associated risk function was used to compare the overall performance of the stochastic predictors developed in later subsections.

Based on a fundamental relationship between estimation and prediction, a very general class of predictors is given in equation (4.2), Section I. These predictors inherit their first- and second-order moment properties from the associated estimators of $\hat{\beta}$, and several examples from this class along with their properties are given. Finally, minimax estimators of $\hat{\beta}$ are used to construct a family of predictors with uniformly smaller risk than that of the best linear unbiased (universal kriging) predictor.

The second section presents a spatial analysis of variance of soil-water infiltration data, and an extension of the robust-resistant spatial analysis of Cressie and Horton (1987). The emphasis here is on the use of geostatistical techniques in the modeling of the intra-
treatment correlation structure. Kriging and cross-validation are used to check model fit, and Winsorization is used to adjust for outliers. Then, using a nested linear model with covariances determined by the modeled spatial correlations, various hypotheses involving main effects are tested assuming the full spatial model, the heteroskedastic model and the classical model in Section 5 of Section II. First, the assumption of no spatial trend was checked using the F-ratio of equation (5.11), Section II. Table 1 shows that the F-ratio for the classical model is much lower than that for the full spatial model, and consequently, conclusions based on the classical model (a model that is often assumed in practice) could be erroneous. Next, the hypothesis of equality of average treatment effects was tested using equation (5.13), Section II. Here, we found that the F-values for the heteroskedastic and classical models were much larger than that for the full spatial model. In this case, assuming the classical model could lead to more frequent declarations of significance than the data indicate. In any event, because F-ratios are not F-distributed when spatial correlation is present but overlooked, use of weighted estimation and spatial analysis of variance is recommended to ensure proper inference and accurate conclusions.
LITERATURE CITED


ACKNOWLEDGMENTS

I am grateful for having had the opportunity to work very closely with my major professor, Dr. Noel Cressie, not only on this dissertation, but on numerous other research projects as well. I have learned a great deal from him in the five years we have worked together. His ideas and enthusiasm are never-ending, and I have relied on them many times for the necessary courage and strength to persevere.

The comments of committee members Dr. D. A. Harville, Dr. K. J. Koehler, Dr. H. A. David, and Dr. E. H. Johnston are very much appreciated. Their suggestions helped to refine and clarify the presentation of ideas within this manuscript.

I thank Dr. Dean Isaacson for his support and encouragement; his confidence in me was a great inspiration. I also thank Dr. Oscar Kempthorne for his consideration, understanding and frankness; I enjoyed our spirited discussions.

I am deeply indebted to the Shell Oil Company Foundation, and the National Science Foundation. Their monetary support over these past two years made this research possible.

Finally, I wish to thank Martin Grondona for his help and his many suggestions. I appreciate the time he spent thinking about my research problems. Muchas gracias, amigo!