A random field approach to spatial experiments

Dale Lee Zimmerman

Iowa State University

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

Part of the Statistics and Probability Commons

Recommended Citation

Zimmerman, Dale Lee, "A random field approach to spatial experiments " (1986). Retrospective Theses and Dissertations. 8501.
https://lib.dr.iastate.edu/rtd/8501
INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

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

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
Zimmerman, Dale Lee

A RANDOM FIELD APPROACH TO SPATIAL EXPERIMENTS

Iowa State University

PH.D. 1986

University Microfilms International
300 N. Zeeb Road, Ann Arbor, MI 48106
A random field approach to spatial experiments

by

Dale Lee Zimmerman

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

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1986
# TABLE OF CONTENTS

1. **INTRODUCTION**
   - 1.1. Basic Concepts of Random Field Theory
   - 1.2. Plot Covariograms
   - 1.3. The Random Field Linear Model
   - 1.4. Spatial Configurations
   - 1.5. Additional Matrix Notation
   - 1.6. Overview

2. **MAXIMUM LIKELIHOOD APPROACHES TO PARAMETER ESTIMATION FOR RANDOM FIELD LINEAR MODELS**
   - 2.1. Non-Likelihood Estimation Procedures
   - 2.2. Maximum Likelihood Approaches
   - 2.3. Restricted Maximum Likelihood
   - 2.4. A Small-Sample Comparison of ML and REML Estimators
   - 2.5. A Relationship Between REML Estimation of Covariogram Parameters and ML Estimation of Generalized Covariance Function Parameters

3. **COMPUTATIONAL ASPECTS OF ML AND REML APPROACHES TO RFLM PARAMETER ESTIMATION**
   - 3.1. The General Spatial Configuration
   - 3.2. Remote Observational Sites
   - 3.3. Configuration I
   - 3.4. Configuration II
   - 3.5. Computational Aspects of ML Approaches for IRF-k's
   - 3.6. Evaluating Integrals Under RFLM-R
3.7. Appendix: Derivation of Result (3.37) 154

4. ASYMPTOTIC PROPERTIES OF ML ESTIMATORS 158

5. ALTERNATIVE APPROACHES TO ANALYZING SPATIAL EXPERIMENTS 172

5.1. Terminology and Notation 172

5.2. Description of Alternative Approaches 177

5.2.1. Classical approach: blocking and randomization 177

5.2.2. Row-column analysis 179

5.2.3. Ordinary covariance analysis 180

5.2.4. The Papadakis analysis 182

5.2.5. Difference-based methods 193

5.2.6. Response surface methods 202

5.3. Relationships Among Approaches 203

5.4. Properties of Treatment Parameter Estimators 210

5.5. An Empirical Comparison of Methods of Analysis 215

6. REFERENCES 230

7. ACKNOWLEDGEMENTS 238
1. INTRODUCTION

We begin by describing, in general terms, the problem to be considered. Suppose it is desired to conduct an experiment to compare the response of a collection of experimental units to two or more treatments. Suppose further that the experimental units have a spatial structure, i.e., the units are distributed throughout a region in space, with the consequence that some units are in closer proximity than others. Examples of experimental units having such structure include pens in a barn or plots in a field. We shall refer to this type of experiment as a spatial experiment.

In the design and analysis of a spatial experiment it is commonly assumed, either explicitly or implicitly, that the observations are uncorrelated. In many spatial experimental settings, however, there is reason to doubt the validity of this assumption. In agricultural field experiments, for example, it is well-known that responses on plots located in close proximity to each other are often positively correlated. Fisher (1926, 1936) and Yates (1938), recognizing the disturbing effect that this spatial correlation could have on the interpretation of the results of an experiment, proposed that one should employ randomized experimental designs, such as randomized blocks or Latin squares, and then analyze the observations by a procedure which has come to be known as the analysis of variance. Fisher argued that the process of randomization neutralizes the effect of correlations among responses, so that
the observations can, in effect, be regarded as uncorrelated. Fisher's argument was based on the now well-known randomization model in which the random allocation of treatments to units induces a probability distribution, known as the randomization distribution, on the "errors" of the model. (For a discussion of the randomization model and the randomization distribution, see Kempthorne 1952, Chapter 8.) It is well-known that, in the absence of treatment differences, the ratio of the treatment mean square to the error mean square in the analysis of variance is, under the randomization distribution, approximately distributed according to the F-distribution; this result is the basis for the classical F-test for testing treatment differences. Some statisticians (e.g., Harville, 1975; Basu, 1980) object to statistical inference which is "design-based," i.e., based on the randomization distribution; nevertheless, randomized designs and the ordinary analysis of variance remain predominant in spatial experimentation.

Recently, however, a need for approaches to spatial experimentation that deal with spatial correlation more directly has been perceived. Experimental designs that may be "better" than randomized designs when there is correlation between neighboring units have received considerable attention in the past decade (e.g., Berenblut and Webb, 1974; Dyke and Shelley, 1976; Duby et al., 1977; Freeman, 1979; Kiefer and Wynn, 1981; Martin, 1982; Bellhouse, 1984; Gill and Shukla, 1985), as have alternative methods of analysis which attempt to account for spatial correlation (e.g., Pearce and Moore, 1976; Bartlett, 1978; Kempton and Howes, 1981;
Wilkinson et al., 1983; Besag and Kempton, 1986). These designs tend to be somewhat systematic, i.e., nonrandomized. Many of the alternative methods of analysis are more complicated than the classical analysis of variance but may be more sensitive to detecting true treatment differences. Interestingly, despite the radical departure of these designs and methods of analysis from the principles of experimentation (randomization in particular) developed by Fisher and others at Rothamsted in the 1920s and 1930s, they have been proposed and developed primarily by British statisticians and agricultural scientists, including members of the Rothamsted establishment.

A different topic, but one in which spatial correlation also figures heavily, is the theory of random fields (also known as random functions, regionalized variables, or models of continuous spatial variation.) A d-dimensional random field is defined here as a collection of random variables \( \{Z_s: s \in \mathbb{R}^d\} \). Random field theory has been used to great advantage in mining and hydrology, where predicting the location of high-grade ore bodies or aquifers based on a finite sample of observations taken at known locations — a procedure known as kriging — is of tremendous economic importance. The application of random field theory to statistical prediction in the earth sciences is the main thrust of geostatistics (Matheron, 1963), a field which has amassed a huge literature in the past 15 years. Curiously enough, the application of random field theory to the design and analysis of spatial experiments has received very little attention.
The objective of this dissertation is to develop an approach to the analysis of spatial experiments that is based on random field theory. Many of the ideas and techniques developed here for spatial experiments could also be (and, to a very limited extent, are being) used to solve geostatistical problems.

We shall describe the model underlying the random field approach to spatial experiments in detail in a later section of this chapter. Before doing so, we review some basic concepts of random field theory that are relevant to spatial experiments.

1.1. Basic Concepts of Random Field Theory

The development here borrows heavily from Matern (1960), Journel and Huijbregts (1978, Chapter 2), and Ripley (1981, Chapters 1 and 4).

Definition 1.1

A d-dimensional random field is a collection of random variables \{Z_s: s \in \mathbb{R}^d\}, i.e., an infinite collection of random variables indexed by points in d-dimensional Euclidean space.

A partial but useful characterization of a random field \{Z_s: s \in \mathbb{R}^d\} is through its first two joint moments, which shall always be assumed herein to exist. The moments of a random field are functions of spatial location. Let the mean function be represented by \(m(s) = \mathbb{E}(Z_s)\). Let \(C(s, t)\) represent the covariance between \(Z_s\) and \(Z_t\); this
function of 2d variables is known as the covariogram of the random field. Thus, $\text{Var}\{Z_s\} = C(s,s)$. Associated with the covariogram is the correlogram $\rho(s,t) = C(s,t)/\sqrt{C(s,s)C(t,t)}$.

In general, estimation of the mean function and covariogram or other types of statistical inference requires knowledge of several realizations of the random field; in spatial experiments, however, usually only one realization is available. Further assumptions on the mean function and covariogram are necessary if inference is to be made from a single realization. One assumption that could be made is that of strict stationarity, which is defined as follows.

**Definition 1.2**

A d-dimensional random field $\{Z_s: s \in \mathbb{R}^d\}$ is said to be strictly stationary if the distribution of $\{Z_s: s \in \mathbb{R}^d\}$ is the same as the distribution of $\{Z_{s+t}: s \in \mathbb{R}^d\}$ for all $t \in \mathbb{R}^d$.

A more relevant concept for our purposes is that of weak stationarity, which is defined as follows.

**Definition 1.3**

A d-dimensional random field $\{Z_s: s \in \mathbb{R}^d\}$ having mean function $m(\cdot)$ and covariogram $C(\cdot,\cdot)$ is said to be weakly stationary if:

1. $m(s) = m$ (a constant) for all $s \in \mathbb{R}^d$.
2. $C(s+u,t+u) = C(s,t)$ for all $u \in \mathbb{R}^d$, $s \in \mathbb{R}^d$, $t \in \mathbb{R}^d$. 
If a random field is weakly stationary, the values of its covariogram and correlogram evaluated at two points \( s \in \mathbb{R}^d \) and \( t \in \mathbb{R}^d \) depend only on the vector difference \( s-t \). Subsequently, we let \( \hat{C}(\cdot) \) and \( \hat{\rho}(\cdot) \), where \( \hat{C} \) and \( \hat{\rho} \) are functions of \( d \) variables, represent the covariogram and correlogram, respectively, of a weakly stationary \( d \)-dimensional random field. Thus, for such a random field, \( C(s,t) \) and \( \hat{C}(h) \), where \( h = s-t \), are alternative representations of the covariogram, and \( \rho(s,t) \) and \( \hat{\rho}(h) \) are alternative representations of the correlogram. We have that
\[
\hat{C}(h) = \hat{C}(0)\hat{\rho}(h) \quad \text{for all } h \in \mathbb{R}^d.
\]

A weakly stationary random field can have the further property that the covariance between any two members \( Z_s \) and \( Z_t \) of the random field depends on only the Euclidean distance between \( s \) and \( t \).

**Definition 1.4**

A \( d \)-dimensional weakly stationary random field having covariogram \( \hat{C}(\cdot) \) is said to be isotropic if \( \hat{C}(h) = \hat{C}(g) \) for all \( h \in \mathbb{R}^d \) and \( g \in \mathbb{R}^d \) such that \( h \cdot h = g \cdot g \).

If a random field is isotropic, its covariogram \( \hat{C}(\cdot) \) and correlogram \( \hat{\rho}(\cdot) \) evaluated at a point \( h \in \mathbb{R}^d \) depend on \( h \) only through \( (h \cdot h)^{1/2} \). Subsequently, we let \( \check{C}(\cdot) \) and \( \check{\rho}(\cdot) \), where \( \check{C} \) and \( \check{\rho} \) are functions of a single variable, represent the covariogram and correlogram, respectively, of an isotropic \( d \)-dimensional random field. Consequently, the covariogram and
correlogram of such a random field can be represented as \( C(s,t) = \hat{C}(h) = \tilde{C}(r) \) and \( \rho(s,t) = \hat{\rho}(h) = \tilde{\rho}(r) \), respectively, where \( r = (h_1 \ h_2)^{1/2} \). Obviously, \( \tilde{C}(r) = \tilde{C}(0)\tilde{\rho}(r) \).

If a weakly stationary random field is not isotropic it is said to be anisotropic. Various forms of anisotropy can occur; perhaps the simplest of these to characterize is geometric anisotropy.

**Definition 1.5**

A weakly stationary random field \( \{Z_s: s \in \mathbb{R}^d\} \) is said to be geometrically anisotropic if there exists a nonsingular \( d \times d \) matrix \( B \) such that the random field \( \{Z^*_s: s \in \mathbb{R}^d\} \) is isotropic, where \( Z^*_s = B^{-1}s \).

Thus, a geometrically anisotropic random field can be transformed to an isotropic random field by making an appropriate linear transformation of the coordinates. The relationship between isotropic and geometrically anisotropic random fields becomes more transparent when we consider their isocorrelation surfaces.

**Definition 1.6**

An isocorrelation surface of a point \( s \) belonging to the index set of a \( d \)-dimensional random field whose correlogram is \( \rho(\cdot, \cdot) \) is the set of all points \( t \in \mathbb{R}^d \) for which \( \rho(s,t) \) is constant.
It is immediate from the definition of an isotropic random field that such a field has isocorrelation surfaces that are $d$-dimensional spheres. What are the isocorrelation surfaces of a geometrically anisotropic random field? Let $\{Z^*_s : s \in \mathbb{R}^d\}$ be a geometrically anisotropic random field, and suppose that $\{Z^*_s : s \in \mathbb{R}^d\}$, where $Z^*_s = Z_s^{-1}$ for some nonsingular matrix $B$, is an isotropic random field. Let $\rho(\cdot, \cdot)$ and $\rho^*(\cdot, \cdot)$ represent the correlograms of the random fields $\{Z_s : s \in \mathbb{R}^d\}$ and $\{Z^*_s : s \in \mathbb{R}^d\}$, respectively. Then,

$$
\rho(s, t) = \rho^*(Bs, Bt)
$$

$$
= \rho^*(Bs - Bt)
$$

$$
= \rho^*(\sqrt{(Bs - Bt)'(Bs - Bt)})
$$

$$
= \rho^*(\sqrt{B'Bh})
$$

which demonstrates that a geometrically anisotropic random field has ellipsoidal isocorrelation surfaces.

There are several attributes of the covariogram of a weakly stationary random field that help to distinguish it from another such covariogram: its behavior at infinity, its range, and its behavior at the origin. Many covariograms, including those which would seem to be reasonable in the setting of spatial experimentation, have the property
that the covariance between two random variables in the random field tends to zero as the distance between their locations increases, i.e.,
\[ \hat{C}(h) \to 0 \text{ as } \sqrt{\frac{1}{h^2}} \to \infty. \]
The range is the usual measure of how rapidly \( \hat{C}(h) \) approaches zero. The range of a covariogram in any given direction is the distance in that direction beyond which the values of the covariogram are "negligible" (if such a distance exists). In general, the value of the range of a random field is direction-dependent; however, for isotropic random fields the range (when it exists) is the same in every direction. The range is a somewhat unsatisfactory measure of how quickly \( \hat{C}(\cdot) \) approaches zero because it is not well-defined if \( \hat{C}(\cdot) \) does not vanish at a finite distance. In Chapter 3, we shall find it necessary to define the range more precisely.

If a covariogram is discontinuous at the origin, it is said to have a nugget effect (Journel and Huijbregts, 1978, p. 39). The presence of a nugget effect can be ascribed to the presence of both measurement error and micro-scale variability. Two types of continuous behavior at the origin can be distinguished: parabolic and linear. For the sake of concreteness we shall describe these behaviors in terms of covariograms of isotropic random fields. If \( \frac{\hat{C}(r)}{r^2} \to K \) (a constant) as \( r \to 0 \), the behavior is said to be parabolic; if \( \frac{\hat{C}(r)}{r} \to K \) as \( r \to 0 \), the behavior is said to be linear. These two types of behavior reflect how rapidly the covariance attenuates at very small distances. The local attenuation rate of the covariogram is of importance because it has a large effect on statistical inference in geostatistical settings (Ripley, 1981, p. 54).
Not every real-valued function of 2d variables can serve as the covariogram of a d-dimensional random field. It can be shown (Breiman, 1968, Chapter 11) that a necessary and sufficient condition for a function $f(\cdot, \cdot)$ of 2d variables to be the covariogram of some d-dimensional random field is that $f(\cdot, \cdot)$ be symmetric and nonnegative definite, i.e., $f(s, t) = f(t, s)$ for all $(s, t) \in \mathbb{R}^d \times \mathbb{R}^d$ and

$$n \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j f(s_i, s_j) > 0$$

for all $n$, $a_1, a_2, \ldots, a_n$, $s_1, s_2, \ldots, s_n$, where $s_i \in \mathbb{R}^d$ for all $i$. For a weakly stationary random field with covariogram $\hat{C}(\cdot)$, the symmetry and nonnegative definiteness of $\hat{C}(\cdot)$ imply the following properties:

(i) $\hat{C}(0) > 0$;

(ii) $\hat{C}(h) = \hat{C}(-h)$;

(iii) $\hat{C}(h) \leq \hat{C}(0)$.

Two theorems pertaining to the construction of covariograms are now given without proof. These theorems can be found in Matern (1960).

One way to construct covariograms is to exploit an isomorphism, of a kind, between continuous covariograms of weakly stationary random fields and characteristic functions of random vectors.

**Theorem 1.1**

$\hat{C}(\cdot)$ is a continuous covariogram of a d-dimensional weakly stationary random field iff there exists a d-dimensional random vector that has a scalar multiple of $\hat{C}(\cdot)$ as its characteristic function. Furthermore, if the distribution of the random vector depends only on its Euclidean norm, then $\hat{C}(\cdot)$ is the covariogram of an isotropic random field.
Another way to construct covariograms is to take advantage of the closure of the class of such functions under various mathematical operations.

**Theorem 1.2**

Let \( \mathcal{C} \) represent the class of all covariograms of a \( d \)-dimensional random field. Then:

1. If \( C_1 \in \mathcal{C} \) and \( C_2 \in \mathcal{C} \), then \( C_1 C_2 \in \mathcal{C} \).
2. If \( C_u \in \mathcal{C} \) for all \( u \in U \subset \mathbb{R} \), and \( \mu \) is a measure on \( U \), then 
   \[
   \int_U C_u \, d\mu(u)
   \]
   is a member of \( \mathcal{C} \).
3. If \( C_1 \in \mathcal{C} \), \( C_2 \in \mathcal{C} \), ..., and \( C = \lim_{i \to \infty} C_i \) exists, then \( C \in \mathcal{C} \).

Theorem 1.1 can be used to construct two particularly important covariograms. Consider the \( d \)-dimensional normal random vector whose mean vector is \( \mathbf{0} \) and whose covariance matrix is \( \sigma^2 I \), where \( \sigma^2 > 0 \). This random vector has probability density function

\[
g(x) = (2\pi \sigma^2)^{-d/2} \exp\left(-\frac{1}{2\sigma^2} x' x\right)
\]

and characteristic function

\[
\phi(t) = \exp\left(-\frac{1}{2} \sigma^2 t' t\right).
\]

Putting \( \theta_2 = \frac{1}{2} \sigma^2 \) and \( r = (t' t)^{1/2} \) demonstrates that a \( d \)-dimensional random field exists that has
\[ C(r) = \theta_1 \exp(-\theta_2 r^2) \] 

(1.1)

( where \( \theta_1 > 0 \) and \( \theta_2 > 0 \) ) as its covariogram. Similar arguments, applied to a \( d \)-dimensional Cauchy random vector, establish that

\[ C(r) = \theta_1 \exp(-\theta_2 r) \] 

(1.2)

( where \( \theta_1 > 0 \) and \( \theta_2 > 0 \) ) is the covariogram of a \( d \)-dimensional random field (Ripley, 1981, p. 11).

Functions (1.1) and (1.2) are referred to as the isotropic Gaussian and isotropic exponential covariograms, respectively. These two functions and the so-called isotropic spherical covariogram

\[
C(r) = \begin{cases} 
\theta_1 \left( 1 - \frac{3r}{2\theta_2} + \frac{r^3}{2\theta_2^3} \right), & \text{if } 0 \leq r \leq \theta_2, \\
0, & \text{otherwise},
\end{cases}
\] 

(1.3)

( where \( \theta_1 > 0 \) and \( \theta_2 > 0 \) ) are the three most widely used covariograms of isotropic random fields in geostatistics. Each of these covariograms is depicted in Figure 1.1. The spherical and exponential covariograms behave linearly at the origin, whereas the Gaussian covariogram has a parabolic behavior at the origin. The spherical covariogram has a finite range. The exponential and Gaussian covariograms do not have a finite range but do tend to zero monotonically as \( r \) increases. Not all covariograms asymptote to zero, however; a counterexample is
Figure 1.1. The isotropic exponential, Gaussian, and spherical covariograms with $\theta_1 = \theta_2 = 1$
\[
C(r) = \theta_1 \cos(\theta_2 r) \quad (1.4)
\]

(Where \( \theta_1 > 0 \) and \( \theta_2 \neq 0 \), which oscillates between \( \theta_1 \) and \( -\theta_1 \) as \( r \) increases.

The exponential covariogram is a member of the larger family of covariograms

\[
C(r) = \frac{\theta_1}{2^{\nu-1} \Gamma(\nu)} (\theta_2 r)^\nu K_\nu(\theta_2 r) . \quad (1.5)
\]

of isotropic random fields, where \( \theta_1 > 0 \) and \( \theta_2 > 0 \), and where \( K_\nu \) is the modified Bessel function of the second kind, order \( \nu \) (\( \nu > 0 \)). This family was introduced by Whittle (1954). When \( \nu = 1/2 \), function (1.5) reduces to the exponential covariogram (1.2), and when \( \nu = 1 \), function (1.5) reduces to

\[
C(r) = \theta_1 \theta_2 r K_1(\theta_2 r) . \quad (1.6)
\]

The covariogram given by (1.6) is monotone decreasing like the exponential covariogram, but it is flat at the origin and attenuates more slowly than the exponential covariogram. Whittle (1954) argued, using stochastic difference equations, that covariogram (1.6) is more appropriate than covariogram (1.2) when \( d = 2 \). Covariograms (1.2) and (1.6) are the only two random field covariograms which have appeared in the literature pertaining to spatial experiments (see Matern, 1960, 1972; Duby et al., 1977).
Numerous other random field covariograms have been put forward. Vecchia (1985) has recently introduced a large family of covariograms that includes Whittle's family as a special case. For a good coverage of available covariograms the reader is referred to Heine (1955), Matern (1960), Pearce (1976), and Mantaglou and Wilson (1982). The methods by which these covariograms have been constructed vary greatly, ranging from the use of Theorems 1.1 and 1.2 to derivations involving stochastic partial differential equations.

In geostatistics the variogram, rather than the covariogram, has been the preferred mode for describing the spatial variation of a random field. The variogram $2\gamma(s, t)$ of a $d$-dimensional random field $\{Z_s: s \in \mathbb{R}^d\}$ is defined as the expectation of $(Z_s - Z_t)^2$. In general, the variogram depends on both $s$ and $t$, but under the so-called intrinsic hypothesis, the variogram is a function of the vector difference $s - t$ only.

**Definition 1.7**

A $d$-dimensional random field $\{Z_s: s \in \mathbb{R}^d\}$ having mean function $m(\cdot)$ is said to satisfy the intrinsic hypothesis if:

1. $m(s) = m$ (a constant) for all $s \in \mathbb{R}^d$;

2. There exists a function $\gamma(\cdot)$ of $d$ variables such that

$$E[(Z_s - Z_t)^2] = 2\gamma(h),$$

where $h = s - t$, for all $(s, t) \in \mathbb{R}^d \times \mathbb{R}^d$.

The intrinsic hypothesis is somewhat more general than the hypothesis of weak stationarity, but when the latter holds, it can be easily demonstrated that the relationship between the variogram and the covariogram
is given by

\[ 2\hat{\gamma}(h) = 2[\hat{C}(0) - \hat{C}(h)] \]

(Journal and Huijbregts, 1978, p. 36).

It is possible to extend the notion of the intrinsic hypothesis to a larger class of random fields, and to define a function associated with this larger class that is analogous to the variogram. We now review these extensions, borrowing ideas from Delfiner (1976).

Let \( \mathcal{F}_Z \equiv \{Z_s: s \in \mathbb{R}^d\} \) represent a \( d \)-dimensional random field. Let \( S = \{s_1, \ldots, s_n\} \) represent a set of \( n \) points in \( \mathbb{R}^d \), and let \( (x_{i1}, \ldots, x_{id}) \) represent the coordinates of \( s_i \) \((i = 1, \ldots, n)\).

**Definition 1.8**

A finite linear combination \( \sum_{i=1}^{n} \lambda_i Z_{s_i} \) of random variables (or realizations of random variables) that belong to \( \mathcal{F}_Z \) is called a generalized increment of order \( k \) \((k = 0, 1, 2, \ldots)\) if the coefficients \( \{\lambda_i: i = 1, \ldots, n\} \) satisfy the conditions

\[ \sum_{i=1}^{n} P_{1}^{P_1} P_{2}^{P_2} \ldots P_{d}^{P_d} \lambda_i x_{i1} x_{i2} \ldots x_{id} = 0 \]

for all nonnegative integers \( P_1, \ldots, P_d \) such that \( P_1 + \ldots + P_d \leq k \).
Conditions (1.7) are illustrated below for the case of a two-dimensional random field when \( k = 0, 1, \) and 2.

(i) \( k = 0; \) \( \sum_{i=1}^{n} \lambda_i = 0. \)

(ii) \( k = 1; \) \( \sum_{i=1}^{n} \lambda_i = 0, \sum_{i=1}^{n} \lambda_i x_{i1} = 0, \sum_{i=1}^{n} \lambda_i x_{i2} = 0. \)

(iii) \( k = 2; \) \( \sum_{i=1}^{n} \lambda_i = 0, \sum_{i=1}^{n} \lambda_i x_{i1} = 0, \sum_{i=1}^{n} \lambda_i x_{i2} = 0, \)
\( \sum_{i=1}^{n} \lambda_i x_{i1}^2 = 0, \sum_{i=1}^{n} \lambda_i x_{i1} x_{i2} = 0, \sum_{i=1}^{n} \lambda_i x_{i2}^2 = 0, \sum_{i=1}^{n} \lambda_i x_{i1} x_{i2} = 0. \)

A generalized increment of order \( k \) can be defined for any set \( S \) of points in \( \mathbb{R}^d. \) For a given set \( S, \) define \( Y_t = \sum_{i=1}^{n} \lambda_i Z_{s_i+t} \) for \( t \in \mathbb{R}^d. \) The collection \( \{Y_t: t \in \mathbb{R}^d\} \) of generalized increments associated with a particular set \( S \) and a particular set of coefficients \( \{\lambda_i: i = 1, ..., n\} \) satisfying (1.7) is itself a \( d \)-dimensional random field, which we denote by \( \mathcal{F}_S^{\lambda}. \)

**Definition 1.9**

A random field \( \mathcal{F}_Z \) is said to be an intrinsic random function of order \( k \) (IRF-\( k \)) if, for each \( S \) and each set of coefficients \( \{\lambda_i: i = 1, ..., n\} \) satisfying (1.7), \( \mathcal{F}_S^{\lambda} \) has zero mean and is weakly stationary.

Thus, if \( \mathcal{F}_Z \) is an IRF-\( k \), then the random variable \( \sum_{i=1}^{n} \lambda_i Z_{s_i+\Delta} \) where \( \Delta \in \mathbb{R}^d \), has mean zero and has a variance which, for all sets \( \{\lambda_i: i = 1, ..., n\} \)}
..., n) that satisfy (1.7), does not depend on $\Delta$. Furthermore, the covariance between $\sum_{i=1}^{n} \lambda_i Z_{s_i + \Delta_i}$ and $\sum_{i=1}^{n} \lambda_i Z_{s_i + \Delta_i}$, where $\Delta_1, \Delta_2 \in \mathbb{R}^d$, depends on $\Delta_1$ and $\Delta_2$ only through their vector difference $\Delta_1 - \Delta_2$.

Clearly, an IRF-0 is a random field that satisfies the intrinsic hypothesis, for $Z_{s} - Z_{s_j}$ is a generalized increment of order 0 (as is seen by taking $\lambda_i = 1$, $\lambda_j = -1$, and the remaining coefficients equal to 0). The converse is also true, as was noted by Matheron (1973).

Clearly, a weakly stationary random field is an IRF-$k$ for any $k > 0$, but the converse need not be true. Similarly, an IRF-$k$ is an IRF-$(k+1)$ but is not, in general, an IRF-$(k-1)$ (where, for the case $k = 0$, we follow Cressie (1986) by defining an IRF-(-1) as a weakly stationary random field). For example, consider a one-dimensional random field $\mathcal{F}_1$ whose mean function is equal to the constant $\alpha_1$ and whose covariogram is $C(s, t; \theta_1) = \theta_1 \min\{s, t\}$ for $s, t \in \mathbb{R}$, where $\theta_1 > 0$. Obviously, this random field is not weakly stationary. Let $\sum_{i=1}^{n} \lambda_i Z_{s_i}$ be a linear combination of members of $\mathcal{F}_1$. For any real numbers $\Delta_1$ and $\Delta_2$, define $g_1 = \sum_{i=1}^{n} \lambda_i Z_{s_i + \Delta_1}$ and $g_2 = \sum_{i=1}^{n} \lambda_i Z_{s_i + \Delta_2}$.

Then, $E(g_1) = \alpha_1 \sum_{i=1}^{n} \lambda_i$ and

$$E(g_1 g_2) = \theta_1 \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \min\{s_i + \Delta_1, s_j + \Delta_2\}$$

$$= \theta_1 \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \min\{s_i + \Delta_1, s_j + \Delta_1 + (\Delta_2 - \Delta_1)\}$$

$$= \theta_1 \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \min\{s_i, s_j + \Delta_2 - \Delta_1\} + \theta_1 \Delta_1 \left( \sum_{i=1}^{n} \lambda_i \right)^2.$$
When \( \sum \lambda_i = 0 \), i.e., when \( g_1 \) and \( g_2 \) are generalized increments of order 0, the mean of \( g_1 \) equals 0 and the covariance between \( g_1 \) and \( g_2 \) depends on \( \Delta_1 \) and \( \Delta_2 \) only through the value of \( \Delta_2 - \Delta_1 \). Therefore, \( \mathcal{F}_1 \), while not weakly stationary, is an IRF-0.

What can be said about the mean function of an IRF-k? Delfiner (1976) observed that the mean function of an IRF-k is not unique; rather, an IRF-k is actually defined only up to an equivalence class of random fields whose mean functions are equal up to a polynomial of degree less than or equal to k.

The second-order dependence properties of an IRF-k are often characterized by a function known as the generalized covariance function, defined as follows.

**Definition 1.10**

Let \( m \) and \( n \) be two positive integers, and let \( \mathcal{F}_Z = \{ Z(s) : s \in \mathbb{R}^d \} \) be a \( d \)-dimensional IRF-k. A continuous and even function \( \mathcal{G}(\cdot) \) of \( d \) variables is said to be a generalized covariance function of order \( k \) of \( \mathcal{F}_Z \) if

\[
\text{Cov}\left\{ \sum_{i=1}^{m} \lambda_i \tilde{Z}_i, \sum_{j=1}^{n} \nu_j \tilde{Z}_j \right\} = \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \nu_j \mathcal{G}(\tilde{s}_i - \tilde{t}_j)
\]  

for all generalized increments \( \sum_{i=1}^{m} \lambda_i \tilde{Z}_i \) and \( \sum_{j=1}^{n} \nu_j \tilde{Z}_j \) of order \( k \).

It is proven by Matheron (1973) that any continuous IRF-k possesses a generalized covariance function of order \( k \), and that a function \( \mathcal{G}(\cdot) \) of
d variables is a generalized covariance function of order k for some IRF-k if and only if it is k-conditionally nonnegative definite, i.e., if and only if \( \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \hat{G}(s_i - s_j) \geq 0 \) for all n, all \( \{s_1, ..., s_n\} \), and all \( \{\lambda_1, ..., \lambda_n\} \) satisfying (1.7). Matheron also pointed out that a generalized covariance function of order k of an IRF-k is unique up to a polynomial of degree 2k, and that \( \hat{G}(h)/(h^h h) \rightarrow 0 \) as \( (h^h h) \rightarrow \infty \).

In order to be explicit about the order of a generalized covariance function, we shall subsequently attach a subscript to the function. Thus, for example, \( \hat{G}_2(\cdot) \) shall represent a generalized covariance function of order 2. Furthermore, at times we shall want to emphasize the dependence of a generalized covariance function of order k on an unknown parameter vector \( \vartheta \), in which case we shall write \( \hat{G}_k(\cdot; \vartheta) \).

If the generalized covariance function \( \hat{G}_k(\cdot; \vartheta) \) of an IRF-k satisfies the property

\[
\hat{G}_k(h; \vartheta) = \hat{G}_k[(h^h h)^{1/2}; \vartheta] \quad \text{for all } h \in \mathbb{R}^d,
\]

for some function \( \hat{G}_k(\cdot; \vartheta) \) of one variable, the IRF-k is said to be isotropic. Otherwise, the IRF-k is said to be anisotropic.

1.2. Plot Covariograms

We have seen that random field covariograms allow one to model the covariance structure of random variables indexed by points in d-dimensional space. In most spatial experiments, however, treatments are applied to (and responses measured over) regions rather than points.
Therefore, it is necessary to consider models for the covariance among observations made on regions, which we call plot covariograms. In this section we describe covariograms of this type and their relationship to the covariograms of random fields.

We begin by giving a precise definition of a plot covariogram. Let \( \mathcal{R} = \{R_i: i = 1, \ldots, n\} \) represent a set of \( n \) regions of "plots" in \( \mathbb{R}^d \). Suppose that a random variable is associated with each plot; let \( \{Z_i: i = 1, \ldots, n\} \) represent these random variables. Then, the plot covariogram is defined as the function \( C(R_i, R_j) = \text{Cov}(Z_i, Z_j) \), and the plot correlogram is defined as the function \( \rho(R_i, R_j) = C(R_i, R_j) / \sqrt{C(R_i, R_i) \cdot C(R_j, R_j)} \). The domain of both functions is \( \mathbb{R} \times \mathbb{R} \).

Most plot covariograms proposed to date are derived from spatial autoregression models defined on a lattice of points in the plane. We now define a special class of spatial autoregression models known as "conditional" spatial autoregression models, and then state a theorem pertaining to the covariance matrix of the observations for this class. The class of conditional spatial autoregression models is relevant because several recently developed methods for analyzing spatial experiments, as well as certain special cases of the method proposed in this dissertation, are quite similar to generalized least squares methods as they apply to a particular subclass of this class. These relationships are examined in Chapter 5.

A conditional spatial autoregression (CAR) is a collection of random variables \( \{Z_i: i = 1, \ldots, n\} \) such that
\[ E(Z_i | Z_j, j \neq i) = \mu_i + \sum_{j=1}^{n} c_{ij} (Z_j - \mu_j) \]  

(1.9)

and

\[ \text{var}(Z_i | Z_j, j \neq i) = \sigma^2, \]

where the \( \{\mu_i\} \), the \( \{c_{ij}\} \), and \( \sigma^2 \) are unknown parameters satisfying \( \sigma^2 > 0 \), \( c_{ii} = 0 \) (\( i = 1, \ldots, n \)), and \( \mu_i = E(Z_i) \) (\( i = 1, \ldots, n \)). Let \( V_g \) represent the covariance matrix of the joint distribution of \( \{Z_i: i = 1, \ldots, n\} \). The following theorem expresses \( V_g \) in terms of \( \sigma^2 \) and the \( \{c_{ij}\} \).

**Theorem 1.3**

Let \( C \) represent the matrix whose \((i,j)\)th element is \( c_{ij} \), where \( c_{ij} \) is defined as in (1.9). If \( I-C \) is symmetric and positive definite, then \( V_g = \sigma^2 (I-C)^{-1} \).

**Proof:**


Conditional spatial autoregression models, as well as other spatial autoregression models, have been fitted to data from uniformity trials, i.e., experiments in which all plots are treated in the same manner (see, e.g., Künsch, 1982). The fits have not always been good. Attempts to fit spatial autoregression models to the data from comparative experiments involving two or more treatments have not yet appeared in the literature.
A possible reason for the failure of spatial autoregression models to fit some uniformity data is that the assumption (implicit in such models) that the data are realizations of random variables defined on a lattice of points is unrealistic. It may be more realistic to assume that each observation is an average of random variables that are defined at all points within a region (Ripley, 1981, p. 94). In geostatistics, such averages are called regularizations.

Definition 1.11

A regularization $Z_A$ of a d-dimensional random field $\{Z_S : s \in \mathbb{R}^d\}$ is the mean value of $Z_S$ over a region $A$ of volume $|A|$, i.e.,

$$Z_A = \frac{1}{|A|} \int_A Z_S ds,$$

where this integral is defined as the mean square limit of approximating sums.

While a regularization is, strictly speaking, a random variable, we shall use the term to describe a realization of a random variable as well.

The equality given by the following theorem is a useful link between random field covariograms and plot covariograms.

Theorem 1.4

Let $Z_A$ and $Z_B$ be two regularizations of a d-dimensional random field $\{Z_S : s \in \mathbb{R}^d\}$ whose mean function is $m(\cdot)$ and whose covariogram is $C(\cdot, \cdot)$. Then,

$$\text{Cov}(Z_A, Z_B) = \frac{1}{|A| |B|} \int_A \int_B C(s, t) ds dt.$$
Proof:

Define \( \mu_A = \mathbb{E}\{Z_A^\cdot\} \) and \( \mu_B = \mathbb{E}\{Z_B^\cdot\} \). By Fubini's theorem,

\[
\mu_A = \frac{1}{|A|} \int_A m(s) ds \quad \text{and} \quad \mu_B = \frac{1}{|B|} \int_B m(s) ds.
\]

Again, using Fubini's theorem,

\[
\text{Cov}\{Z_A^\cdot, Z_B^\cdot\} = \mathbb{E}\{(Z_A^\cdot - \mu_A^\cdot)(Z_B^\cdot - \mu_B^\cdot)\}
\]

\[
= \mathbb{E}\left[\frac{1}{|A|} \int_A (Z_s - m(s)) ds \right] \left[\frac{1}{|B|} \int_B (Z_t - m(t)) dt \right]
\]

\[
= \mathbb{E}\left[\frac{1}{|A| |B|} \int_A \int_B [Z_s - m(s)] [Z_t - m(t)] ds dt \right]
\]

\[
= \frac{1}{|A| |B|} \int_A \int_B C(s, t) ds dt.
\]

Q.E.D.

1.3. The Random Field Linear Model

We now describe the model underlying the approach to the analysis of spatial experiments that is taken in this dissertation. The notation and terminology introduced in this section will continue to apply throughout the entire dissertation.

Suppose that a spatial experiment is to be conducted on a bounded (but not necessarily finite) set of points \( \mathbb{R} \) in \( \mathbb{R}^d \). The set \( \mathbb{R} \) shall be called the experimental material. Accordingly, one of \( t \) (\( t \geq 2 \)) treatments is to be applied to each of \( n \) exhaustive, mutually exclusive, and connected subsets of \( \mathbb{R} \). Let \( R_1, R_2, \ldots, R_n \) denote these subsets.
Subsequent to the application of treatments, a certain "response variate" is to be observed on each of the subsets $R_1, \ldots, R_n$. (For example, in agricultural field experimentation, the response variate is often the yield of a crop.) Let $y_i$ denote the observed response variate on $R_i$ ($i = 1, \ldots, n$). We shall refer to $y_i$ as the $i$th observation and to $R_i$ as the $i$th observational site.

Now, suppose that we regard the outcome of the spatial experiment as a single realization of a random field $\mathcal{F}_y \equiv \{Y_s: s \in \mathbb{R}^d\}$. Accordingly, $y_1, \ldots, y_n$ are to be viewed as observations of $\mathcal{F}_y$ on $R_1, \ldots, R_n$. Suppose further that members of $\mathcal{F}_y$ have the representation

$$Y_s = m(s; \beta) + Z_s,$$

where $m(s; \beta)$ denotes the expectation of $Y_s$ (which is assumed to exist), $\beta = \{\beta_i\}$ is a px1 vector of unknown parameters, and $\mathcal{F}_Z \equiv \{Z_s: s \in \mathbb{R}^d\}$ is a random field such that $E[Z_s] = 0$ and $\text{Var}(Z_s)$ exists for all $s \in \mathbb{R}^d$. We assume that $m(s; \beta)$ is a linear function of the elements of $\beta$.

The existence of $\text{Var}(Z_s)$ for all $s \in \mathbb{R}^d$ implies, by the Schwartz inequality, that $\text{Cov}(Z_s, Z_t)$ exists for all $(s, t) \in \mathbb{R}^d \times \mathbb{R}^d$. Let $C(\cdot, \cdot; \beta)$ and $\rho(\cdot, \cdot; \beta)$ represent the covariogram and the correlogram, respectively, of $\mathcal{F}_Z$ (or equivalently, of $\mathcal{F}_y$), where $\beta = \{\beta_i\}$ is an mx1 vector of unknown parameters. Put $\phi = (\beta', \theta')$. The parameter space for $\phi$ is taken to be $\{\phi: \beta \in \mathbb{R}^p, \theta \in \Theta\}$, where $\Theta$ is a specified subset of $\mathbb{R}^m$. 
Since we have regarded the outcome of the spatial experiment as a realization of the random field \( \mathcal{F}_Y \), equation (1.10) implies that the model for the observational vector \( \mathbf{y} = (y_1, \ldots, y_n)' \) can be represented matricially as

\[
\mathbf{y} = \mathbf{X}\hat{\mathbf{\beta}} + \mathbf{e},
\]  

(1.11)

where \( \mathbf{X} \) is an \( n \times p \) matrix of rank \( p^* \) whose elements are known functions of the spatial location of observational sites, \( \mathbf{e} \) is an \( n \times 1 \) random vector such that \( E[\mathbf{e}] = \mathbf{0} \) and \( \text{Var}[\mathbf{e}] = \mathbf{V} \), and \( \mathbf{V} \) is an \( n \times n \) matrix whose elements are known functions of \( \hat{\mathbf{\beta}} \). We specify the parameter space for \( \hat{\mathbf{\beta}} \) as \( \{ \hat{\mathbf{\beta}} : \hat{\mathbf{\beta}} \in \mathbb{R}^p, \mathbf{\beta} \in \mathbb{R} \} \), where \( \mathbb{R} \) is either the set \( \Theta_0 \) of \( \hat{\mathbf{\beta}} \)-vectors for which \( \mathbf{V} \) is positive definite or, alternatively, some given subset of \( \Theta_0 \). Subsequently, we refer to model (1.11) as the random field linear model (RFLM).

Two special cases of the RFLM, corresponding to two types of sites at which \( \mathcal{F}_Y \) is observed, are of major importance. If the observational sites are a set of points \( \{ s_1, \ldots, s_n \} \in \mathbb{R}^d \), then \( y_i = \mathbf{X}_{s_i} \hat{\mathbf{\beta}} \), the \( i \)th element of \( \mathbf{X}\hat{\mathbf{\beta}} \) is \( m(s_i; \hat{\mathbf{\beta}}) \), the \( i \)th element of \( \mathbf{e} \) is \( e_{s_i} \), and the elements \( \{ v_{ij} \} \) of \( \mathbf{V} \) are given by \( v_{ij} = C(s_i, s_j; \Theta) \) \((i = 1, \ldots, n; j = 1, \ldots, n)\). Subsequently, we refer to this RFLM as RFLM-P.

Alternatively, if the observational sites are a set of \( d \)-dimensional regions \( \{ S_1, \ldots, S_n \} \), we regard the observations as regularizations of \( \mathcal{F}_Y \). Consequently, \( y_i = \frac{1}{|S_i|} \int_{S_i} \mathbf{Y}(s) \, ds \), the \( i \)th element of \( \mathbf{X}\hat{\mathbf{\beta}} \) is
\[
\frac{1}{|s_i|} \int_{s_i} m(s, \theta) ds \text{ (which is linear in the elements of } \theta \text{ since the integrand has this property), the } i\text{th element of } e \text{ is } \frac{1}{|s_i|} \int_{s_i} Z_s ds, \text{ and using Theorem 1.4,}
\]

\[
v_{ij} = \frac{1}{|s_i| |s_j|} \int_{s_i} \int_{s_j} C(s, t; \theta) ds dt \tag{1.12}
\]

\((i = 1, \ldots, n; j = 1, \ldots, n)\). Subsequently, we refer to this RFLM as RFLM-R.

Both RFLM-P and RFLM-R (and, indeed, any RFLM) are special cases of a classical statistical linear model known as the general mixed linear model (see, e.g., Harville, 1977).

When it is important to distinguish the covariance matrix of an RFLM-P from that of an RFLM-R, we shall denote the former by \(V_P\) and its elements by \(\{v_{ij}^P\}\), and the latter by \(V_R\) and its elements by \(\{v_{ij}^R\}\).

If \(Y\) is an IRF-k, matrices \(K_k^P\) and \(K_k^R\), corresponding to RFLM-P and RFLM-R, can be defined. In conjunction with RFLM-P, let \(K_k^P\) denote the \(n \times n\) matrix whose elements \(\{k_{ij}^{(k)P}\}\) are given by

\[
k_{ij}^{(k)P} = \hat{G}_k(s_i, s_j; \theta) \quad (i = 1, \ldots, n; j = 1, \ldots, n),
\]

where \(\hat{G}_k(*)\) is the generalized covariance function of \(Y\). In conjunction with RFLM-R, let \(K_k^R\) denote the \(n \times n\) matrix whose elements \(\{k_{ij}^{(k)R}\}\) are given by

\[
k_{ij}^{(k)R} = \frac{1}{|s_i| |s_j|} \int_{s_i} \int_{s_j} \hat{G}_k(s, t; \theta) ds dt.
\]

It is possible for $\Theta$, as defined above, to be empty, as the following example illustrates for an RFIM-P. Suppose that $d = 2$ and $Z$ is isotropic with covariogram (1.4). Suppose further that $Z$ is observed at three points $s_1$, $s_2$, and $s_3$ in $\mathbb{R}^2$ that form an equilateral triangle of side $2\pi/\theta_2^*$, where $\theta_2^*$ is a nonzero scalar, and that $\Theta = \{\theta_1, \theta_2: \theta_1 \in \mathbb{R}, \theta_2 = \theta_2^*, \text{ and } Z \text{ is positive definite}\}$. (Note that this parameter space is a subset of $\Theta_0$.) Then

$$V = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

which is singular for any $\theta_1$. Thus, $\Theta$ is empty. Hereafter, to avoid any difficulties caused by $\Theta$ being empty, we assume that $\Theta$ is a nonempty subset of $\mathbb{R}^m$.

Two particular subsets of $\Theta_0$ will be of importance in the sequel. Let $\Theta_1$ represent the subset of $\Theta_0$ for which the second-order partial derivatives of the elements of $V$ with respect to the elements of $\Theta$ exist, i.e.

$$\Theta_1 = \{\Theta: \Theta \in \Theta_0 \text{ and } \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \text{ exists (}i = 1, \ldots, m; j = 1, \ldots, m)\}. \quad (1.13)$$

Let $\Theta_2$ represent the subset of $\Theta_0$ for which the second-order partial derivatives of the elements of $V$ with respect to the elements of $\Theta$ are continuous functions of the elements of $\Theta$, i.e.,
\[ \Theta_2 = \{ \Theta: \Theta \in \Theta_0 \text{ and } \partial^2 \Phi / \partial \Theta_i \partial \Theta_j \text{ is a continuous function of} \] 
\[ \text{the elements of } \Theta \ (i = 1, \ldots, m; j = 1, \ldots, m) \}. \] (1.14)

From standard results on differentiability and continuity of functions of more than one variable, it is easy to see that \( \Theta_2 \subseteq \Theta_1 \).

A special class of RFLMs is the class for which \( \mathcal{F}_Y \) has mean function

\[ m(s; \Theta) = \sum_{i=1}^{q} \alpha_i f_i(s) + \sum_{j=1}^{t} \tau_j g_j(s), \]

where the \( \{\alpha_i\} \) and the \( \{\tau_j\} \) are unknown parameters, the \( \{f_i(\cdot)\} \) are functions of spatial location that do not depend on which subsets of \( \mathcal{S} \) receive any given treatment, and

\[ g_j(s) = \begin{cases} 1, & \text{if } s \text{ receives the } j \text{th treatment,} \\ 0, & \text{otherwise.} \end{cases} \]

For members of this class, model equation (1.11) can be re-expressed as

\[ y = X_\alpha \zeta + T\tau + e, \] (1.15)

where \( X_\alpha \) and \( T \) are known \( nxq \) and \( nxt \) matrices, respectively, \( \zeta = (\alpha_1, \ldots, \alpha_q)' \), and \( \tau = (\tau_1, \ldots, \tau_t)' \). Subsequently, we refer to a member of this class of RFLMs as a treatment-additive RFLM.

Another class of RFLMs is the class for which \( \mathcal{F}_Y \) has mean function
\[ m(s; \beta) = \sum_{i=1}^{q} \alpha_i f_i(s), \]

where the \( \{\alpha_i\} \) and \( \{f_i(\cdot)\} \) are defined as they were for the class of
treatment-additive RFLMs. Kriging and other geostatistical inference
procedures are often approached using a member of this class (see,
e.g., Matheron, 1971).

1.4. Spatial Configurations

The spatial configuration of the observational sites \( R_1, \ldots, R_n \)
at which \( \mathcal{X}_Y \) is observed is an important aspect of an RFLM, primarily
through the effect it can have on the structure of the covariance
matrix \( V \). Two spatial configurations, in particular, shall be studied
extensively in this dissertation. These spatial configurations are
described after introducing two definitions.

Definition 1.11

Let \( R_1, \ldots, R_n \) represent \( n \) subsets of \( \mathbb{R}^d \). Then \( R_1, \ldots, R_n \) are
said to be isometric if they are all of the same size and shape.

Definition 1.12

Let \( R_1, \ldots, R_n \) represent \( n \) isometric subsets of \( \mathbb{R}^d \). Then \( R_1, \ldots, R_n \)
are said to be superimposable if \( d \times 1 \) vectors \( \{a_i : i = 1, \ldots, n\} \)
exist such that \( \{x : \tilde{x} \in R_i\} = \{\tilde{x} + a_i : \tilde{x} \in R_i\} \) for all \( i = 1, \ldots, n \).
Note that under any RFLM-P, the subsets $R^1, \ldots, R^n$ are isometric and superimposable (since the subsets, under such an RFLM, are merely points). However, under RFLM-R, requiring isometry and/or superimposability imposes quite restrictive conditions on the subsets. Subsets having Configuration I, defined as follows, must satisfy even more restrictive conditions.

**Definition 1.13**

Let $R^1, \ldots, R^n$ represent $n$ superimposable subsets of $\mathbb{R}^d$. Let $H$ be a real number, and let $a = (H, 0, 0, \ldots, 0)$ be a $d \times 1$ vector. Then, $R^1, \ldots, R^n$ are said to have Configuration I if \( \{x: x \in R_i^i\} = \{x + (i-1)a: x \in R^i\} \) for all $i = 1, \ldots, n$. The number $H$ is called the span of Configuration I.

Clearly, an RFLM in which the observational sites have Configuration I is necessarily an RFLM-P or an RFLM-R. Below, we illustrate examples of Configuration I, under both RFLM-P and RFLM-R, when $d = 1$ and $d = 2$. In each example $H = 1$ and $n = 5$.

A second type of spatial configuration, referred to as Configuration II and defined as follows, includes Configuration I as a special case.

**Definition 1.14**

Let $R^1, \ldots, R^n$ represent $n$ superimposable subsets of $\mathbb{R}^d$, where $d \geq 2$. Let $H_1$ and $H_2$ be two real numbers, and let $a = (H_1, 0, \ldots, 0)^T$
Figure 1.2. Examples of Configuration I. (a) $d = 1$, RFLM-P; (b) $d = 2$, RFLM-P; (c) $d = 1$, RFLM-R; (d) $d = 2$, RFLM-R.
and \( b = (0, H_2, 0, \ldots, 0) \)' be two \( dx \times 1 \) vectors. Then \( R_1, \ldots, R_n \) are said to have Configuration II if:

1. there exist integers \( R \) and \( C \) such that \( n = RC \).
2. the subsets can be re-labelled, using double subscripts, as

\[ R_{11}, R_{12}, \ldots, R_{1C}, R_{21}, R_{22}, \ldots, R_{RC} \]

such that \( \{ \bar{x} : \bar{x} \in R_{ij} \} = \{ \bar{x} + (i-1)a + (j-1)b : \bar{x} \in R_{11} \} \) for all \( i = 1, \ldots, C \) and all \( j = 1, \ldots, R \).

The numbers \( H_1 \) and \( H_2 \) are called the within-row span and the within-column span, respectively, of Configuration II.

An RFLM in which the observational sites have Configuration II is necessarily an RFLM-P or an RFLM-R. Furthermore, such a configuration consists of \( R \) rows and \( C \) columns of superimposable subsets. Note that Configuration I is a special case of Configuration II for which \( R = 1 \).

Below, we illustrate examples of Configuration II under both RFLM-P and RFLM-R. In each example \( R = 3, C = 5, H_1 = 1, \) and \( H_2 = 2 \).

An important example of observational sites that have Configuration I or Configuration II arises in the setting of agricultural field experimentation. Commonly, in such experiments, the experimental material \( R \) is a rectangular parcel of land that is to be partitioned into isometric contiguous rectangular "plots." If the parcel is partitioned to form a single row of plots, the spatial configuration is Configuration I [see Figure 1.4(a)]. Alternatively, if the parcel is partitioned to form rows and columns, the plots have Configuration II [see Figure 1.4(b)].
Figure 1.3. Examples of Configuration II. (a) under RFLM-P; (b) under RFLM-R
Throughout this dissertation, we shall describe special features that RFLM's have when the spatial configuration of observational sites is Configuration I or Configuration II. In such discussions $H$ shall always represent the span of Configuration I, and $R$, $C$, $H_1$, and $H_2$ shall always represent the number of rows, the number of columns, the within-row span, and the within-column span, respectively, of Configuration II. Moreover, when the observational sites have Configuration II, the sites shall be re-labelled, using double subscripts, as $R_{11}, R_{12}, \ldots, R_{RC}$. The corresponding observations shall be similarly relabelled as $y_{11}, y_{12}, \ldots, y_n$, and shall be arranged in the observational vector $y$ in lexicographic order, i.e., $y = (y_{11}, y_{12}, \ldots, y_{RC})'$. 

Figure 1.4. Depictions of (a) Configuration I, and (b) Configuration II, when a rectangular region in $\mathbb{R}^2$ is partitioned into isometric contiguous rectangular subregions.
1.5. Additional Matrix Notation

We shall often require matrix notation in addition to that introduced in Section 1.3. We now introduce such notation, which shall be in effect throughout the dissertation.

Let \( \mathbf{1}_n \) represent the \( nx1 \) vector whose elements are all equal to one, and let \( \mathbf{I}_n \) represent the \( nxn \) identity matrix. When the dimensions of \( \mathbf{1}_n \) and \( \mathbf{I}_n \) are clear from the context, we may omit the subscripts.

For any matrix \( \mathbf{A} \), let \( \mathcal{C}(\mathbf{A}) \) and \( \mathcal{N}(\mathbf{A}) \) represent the column space of \( \mathbf{A} \) and the null space of \( \mathbf{A} \), respectively. If the dimensions of \( \mathbf{A} \) are \( axb \), we may sometimes denote the \( (i,j) \)th of \( \mathbf{A} \) by \( (\mathbf{A})_{ij} \) \( (i = 1, \ldots, a; \ j = 1, \ldots, b) \); this notation will be most useful when \( \mathbf{A} \) is a matrix to which subscripts have already been attached. Further, if \( \mathbf{B} \) is a matrix whose dimensions are \( cxe \), let \( \mathbf{A} \otimes \mathbf{B} \) represent the direct (or Kronecker) product of \( \mathbf{A} \) and \( \mathbf{B} \), i.e., the \( acxe \) matrix given by

\[
\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix}
\mathbf{A}^1_1 & \mathbf{A}^1_2 & \cdots & \mathbf{A}^1_e \\
\mathbf{A}^2_1 & \mathbf{A}^2_2 & \cdots & \mathbf{A}^2_e \\
& \ddots & \ddots & \\
\mathbf{A}_c^1 & \cdots & \mathbf{A}_c^e
\end{bmatrix},
\]

where \( b_{ij} \) is the \( (i,j) \)th element of \( \mathbf{B} \).
1.6. Overview

Conceptualizing the outcome of a spatial experiment as a realization of a random field can provide the framework for a unified approach to several issues that arise in spatial experimentation, including estimation of treatment effects and other model parameters, optimal plot shape and size, and optimal allocation of treatments to experimental units. Each of these issues is important, but only the first shall be dealt with in this dissertation. In Chapter 2, we consider the estimation of the parameters of an RFIM by maximum likelihood approaches. These approaches can be quite computation-intensive, so in Chapter 3 computational aspects unique to the spatial context are considered that make their implementation more practical. Sufficient conditions for the asymptotic normality of the parameter estimators are given in Chapter 4 for certain special cases. Finally, in Chapter 5 some previously proposed methods for dealing with correlation in spatial experiments are compared, both analytically and empirically, to the random field approach proposed herein.
2. MAXIMUM LIKELIHOOD APPROACHES TO PARAMETER ESTIMATION FOR RANDOM FIELD LINEAR MODELS

We saw in Chapter 1 that regarding the outcome of a spatial experiment as a realization of the random field $\mathcal{F}_Y$ given by (1.10) leads to a model for the observations, namely model (1.11), that we have called the random field linear model. In this chapter we consider the estimation of the parameters of an RFLM by various implementations of the method of maximum likelihood. Although the focus of this dissertation is spatial experiments, the maximum likelihood approaches described in this chapter are applicable to any setting where an RFLM arises. In geostatistical settings where an RFLM is appropriate, attention has centered on estimating the vector of covariance parameters $\vartheta$ so that a subsequent kriging analysis can be performed. In such contexts, $\vartheta$ is regarded as a nuisance parameter so far as the estimation of $\vartheta$ is concerned (though not, of course, in the subsequent kriging analysis because such an analysis involves the prediction of random variables whose distribution depends on $\vartheta$). Consequently, methods for estimating $\vartheta$ which circumvent the estimation of $\vartheta$ are often used. This is in contrast to the spatial experimental setting, where primary interest lies in estimating certain estimable linear combinations of the elements of $\vartheta$, namely, treatment contrasts. However, it would seem that estimating $\vartheta$ well is a prerequisite to obtaining good point estimates and, particularly, good interval estimates, of treatment contrasts.
2.1. Non-likelihood Estimation Procedures

Before considering the estimation of the parameters of an RFLM via maximum likelihood procedures, let us briefly mention some alternative estimation procedures currently used as a preliminary to kriging. Some, but not all, of these methods are applicable to spatial experiments.

The most widely used estimation methods in a pre-kriging analysis consist of a wide variety of techniques which we refer to as 'two-step' procedures. Though they apply more generally, for ease of exposition we discuss these procedures only as they apply to a d-dimensional random field \( \mathcal{F}_W \equiv \{ W_s : s \in \mathbb{R}^d \} \) that satisfies the intrinsic hypothesis and is observed at points \( s_1, \ldots, s_n \) in \( \mathbb{R}^d \). Suppose further that \( \mathcal{F}_W \) is isotropic, and let \( \gamma(\cdot) \) (a function of one variable) represent its variogram. One begins a two-step procedure by estimating the variogram at each of a finite number of distances. The classical estimator of the variogram is the method of moments estimator

\[
\hat{\gamma}(r_j) = \frac{1}{N_j} \sum_{i,j} (W_{s_i} - W_{s_k})^2 \quad (j = 0, 1, \ldots, J),
\]

where \( I_j \) is the set \( \{(s_i, s_k) : M_j \leq \sqrt{(s_i - s_k)'(s_i - s_k)} < M_{j+1}\} \), \( \{M_j\} \) is an increasing sequence of nonnegative real numbers, \( M_0 = 0 \), \( N_j \) is the cardinality of \( I_j \), \( r_j = (M_j + M_{j+1})/2 \), and \( J \) is a positive integer. This estimator was proposed by Matheron (1963) and is unbiased for \( \hat{\gamma}(r_j) \) (under the intrinsic hypothesis) when \( \sqrt{(s_i - s_k)'(s_i - s_k)} = r_j \) for all
\((s_i, s_k) \in I_j\) (as can happen when the spatial configuration of the \(\{s_i\}\) is highly regular). If the random field is weakly stationary and isotropic, the method of moments estimator of the covariogram

\[ \hat{C}(r_j) = \frac{1}{N_j} \sum_{i,j} (s_i - \bar{s})(s_j - \bar{s}) \quad (j = 0,1,\ldots,J), \]

where \(\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s_i\), could alternatively be used. However, \(\hat{C}(r_j)\) is not an unbiased estimator of \(C(r_j)\); its bias is known to be \(O(1/N_j)\) (Fuller, 1976, Section 6.2). An ad hoc rule for choosing \(J\) and the sequence \(\{M_j\}\), given \(n\) and \(\{s_i: i = 1,\ldots,n\}\), is given by Journel and Huijbregts (1978, p. 194), who also suggest that anisotropy can be dealt with by obtaining separate estimates of the variogram or covariogram in each of several directions.

The second step of any two-step estimation procedure is to fit a parametric model [e.g., the exponential covariogram (1.2)] to the estimated variogram or covariogram. This has usually been accomplished either "by eye" or by least squares; however, because the variance of \(\hat{\gamma}(r_j)\) and \(\hat{C}(r_j)\) is, in general, different for different \(j\), and covariances between \(\hat{\gamma}(r_j)\) and \(\hat{\gamma}(r_{\ell})\) (where \(j \neq \ell\)) are not zero, it is more appropriate to use generalized or weighted least squares, as suggested by Cressie (1985).

While two-step estimation procedures are quite ad hoc and little is known about their properties, they continue to receive attention in the geostatistical literature, (see, e.g., Armstrong, 1984). Recently,
however, efforts have been made to replace two-step procedures with less ad hoc methods. In addition to the maximum likelihood approaches discussed in the following section, two alternative procedures (Kitanidis, 1985; Marshall and Mardia, 1985) which have been proposed for RFLM's are minimum variance quadratic unbiased estimation (MIVQUE) (Rao, 1971b) and minimum norm quadratic unbiased estimation (MINQUE) (Rao, 1971a). Contrary to an assertion made by several authors, MIVQUE, in its most general form, does not require a normality assumption. MINQUE requires a normality assumption only to obtain formulae for the variances of the parameter estimates. Although MIVQUE and MINQUE have several good properties, they are of limited use in connection with weakly stationary random field models because their implementation requires that the covariance matrix $V$ of the RFLM be a linear function of the elements of $\theta$ which, in turn, requires that the covariogram of $T_{X}$ be linear in the elements of $\theta$. The most widely used random field covariograms [e.g., covariograms (1.1)-(1.3)] are nonlinear in the elements of $\theta$ and cannot be rendered linear by a reparametrization. However, there are useful variograms and generalized covariance functions that are linear in their parameters; MIVQUE and MINQUE could be used in connection with these functions.

Although maximum likelihood approaches to RFLM parameter estimation are described and advocated in this chapter, the author does not suggest that all features of the ad hoc two-step estimation procedure should be discarded. Rather, the first step of the two-step procedure is viewed
here as a potentially valuable diagnostic tool for selecting an appropriate covariogram (or variogram), and the second step (in conjunction with the first) may be useful for obtaining rough estimates of parameters that can be used in a maximum likelihood approach as initial values for one of the iterative maximization algorithms described in the following section.

2.2. Maximum Likelihood Approaches

The method of maximum likelihood has been used in recent years to estimate the parameters of a particular subclass of mixed linear models known as the ordinary fixed, mixed, and random ANOVA models (see Harville, 1977). These models have the same form as model (1.11) but the covariance matrix is a linear function of the elements of \( \hat{\Theta} \), i.e.,

\[
V(\hat{\Theta}) = \hat{\Theta}_1 V_1 + \hat{\Theta}_2 V_2 + \ldots + \hat{\Theta}_m V_m \quad \text{for nxm matrices} \{V_i\} ; \\
\text{the elements of} \quad \hat{\Theta} \quad \text{are frequently referred to as variance components.}
\]

More recently, some researchers have suggested that the method of maximum likelihood (ML) can be successfully applied to the estimation of the parameters of an RFLM; see, e.g., Kitanidis and Vomvoris (1983) or Mardia and Marshall (1984). The apparent reason why ML estimation was shunned previously in the RFLM context (in favor of the ad hoc two-step procedures described in the previous section) is the bias in the ML estimator of \( \hat{\Theta} \), which is reported to be serious, for small to moderate sized samples due to the necessity of estimating \( \hat{\Theta} \) as well as \( \Theta \) (Matheron, 1971). Despite this drawback, maximum likelihood procedures have several advantages over ad hoc
procedures. We discuss maximum likelihood approaches for estimating the parameters of RFLMs in this section; a variant of maximum likelihood which reduces the aforementioned bias will be described in Section 2.3.

Throughout this chapter we take \( \tilde{\beta} \) to be any solution to the normal equations

\[
x'V^{-1}x\tilde{\beta} = x'V^{-1}y.
\]

(2.1)

The existence of \( V^{-1} \) is guaranteed by the restriction that \( \theta \) belong to \( \Theta_0 \) or some subset of \( \Theta_0 \). The precise nature of \( \Theta_0 \) for many RFLM's is discussed later in this section.

Note that in (2.1) the dependence of \( V \) (and hence \( \tilde{\beta} \)) on \( \Theta \) has been suppressed. When we desire to stress the functional dependence of a matrix or vector on one or more parameter vectors, we shall append the appropriate argument(s). Thus, \( V(\theta) \) and \( \tilde{\beta}(\theta) \) are different representations of \( V \) and \( \tilde{\beta} \), and if \( \theta_0 \) is a particular value of \( \theta \), then \( \tilde{\beta}(\theta_0) \) is the value of \( \tilde{\beta} \) when \( \theta = \theta_0 \).

If an RFLM is adopted for \( y \) and if \( y \) is assumed to have a multivariate normal distribution, then the ordinary log-likelihood function is (apart from an additive constant which does not depend on \( \tilde{\beta} \) or \( \theta \))

\[
L(\beta, \theta; y) = - \frac{1}{2} \log|V| - \frac{1}{2}(y - x\beta)'V^{-1}(y - x\beta),
\]

(2.2)
defined for \( \beta \in \mathbb{R}^p \) and \( \theta \in \Theta \). Maximum likelihood estimators of \( \beta \) and \( \theta \) are defined to be values which maximize \( L(\beta, \theta; y) \), i.e., values \( \hat{\beta} \) and \( \hat{\theta} \) (not necessarily unique) that satisfy
\[
L(\hat{\beta}, \hat{\theta}; \gamma) = \sup_{\beta \in \mathbb{R}^p, \theta \in \Theta} L(\beta, \theta; \gamma) .
\]

We shall require that if two different values of \( \gamma \) produce the same log-likelihood function, then the same values of \( \hat{\beta} \) and \( \hat{\theta} \) will be chosen to be the ML estimates for the second value of \( \gamma \) as for the first value of \( \gamma \).

We put \( \hat{\phi} = (\hat{\beta}' , \hat{\theta})' \).

It is possible to simplify the determination of \( \hat{\beta} \) and \( \hat{\theta} \) somewhat by exploiting the well-known result that for any fixed \( \hat{\theta} \), \( L \) is maximized with respect to \( \beta \) at \( \hat{\beta} = \hat{\beta}(\hat{\theta}) \). Putting \( \hat{\beta}(\hat{\theta}) \) into expression (2.2) yields, after some simplification, the function

\[
L^*(\theta; \gamma) = - \frac{1}{2} \log|V| - \frac{1}{2\gamma'} P_V y ,
\]

defined for \( \hat{\theta} \in \Theta \), where

\[
P_V = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}
\]

(\text{for any matrix } A, A^{-1} \text{ represents any generalized inverse of } A, \text{ i.e., any solution to } AA^{-1} A = A). Determining \( \hat{\beta} \) and \( \hat{\theta} \) by maximizing \( L \) is entirely equivalent to maximizing \( L^*_1 \) over \( \Theta \) to obtain \( \hat{\theta} \) and then putting \( \hat{\beta} = \hat{\beta}(\hat{\theta}) \). The advantage of maximizing \( L^*_1 \) rather than \( L \) is that the dimensionality of the maximization problem is reduced from \( (p+m) \) to \( m \).

The property of \( \Theta_1 \)-factorability, defined as follows, permits a further reduction in the dimensionality of the maximization problem.

\[
L^*_1(\theta; y) = \sup_{\beta \in \mathbb{R}^p, \theta \in \Theta} L(\beta, \theta; y) .
\]
Let $\theta_2$ represent the \((m-1)\times 1\) vector that consists of the last \(m-1\) elements of $\theta$, i.e., $\theta_2 = (\theta_2, \theta_3, \ldots, \theta_m)'$.

**Definition 2.1**

A d-dimensional random field whose correlogram is $\rho(\cdot, \cdot; \theta)$ is said to be $\theta_1$-factorable if $\rho(\cdot, \cdot; \theta)$ depends on $\theta$ only through the value of $\theta_2$.

Note that a random field having any of covariograms (1.1)-(1.6) is $\theta_1$-factorable.

Suppose that $\mathbb{F}_Y$ is $\theta_1$-factorable. Then, the covariance matrix $V$ of the observations can be expressed as $V(\theta) = \theta_1 W(\theta_2)$. That is, $V$ is a scalar multiple of a matrix whose elements are functions of only the last $m-1$ components of $\theta$. In this case, the dimensionality of the problem of obtaining maximum likelihood estimates can be further reduced from $m$ to $m-1$, for it is well known that for any fixed value of $\theta_2$, $L_1^*$ is maximized by taking

$$
\theta_1 = \frac{1}{n} \left[ y - \bar{y}(\theta_2) \right]' W(\theta_2) [W(\theta_2)]^{-1} \left[ y - \bar{y}(\theta_2) \right],
$$

where $\bar{y}(\theta_2)$ is any solution to the equations

$$
X' [W(\theta_2)]^{-1} x = X' [W(\theta_2)]^{-1} y.
$$

Note that any solution to equations (2.6) is also a solution to equations (2.1) in this case, irrespective to the value of $\theta_1$. Substituting
expression (2.5) into expression (2.3) yields a function which differs by no more than an additive constant from the function

\[ L^*_2(\theta_2; y) = -\frac{1}{2} \log|W| - \frac{n}{2} \log(y' P_w y), \tag{2.7} \]

where

\[ P_w = W^{-1} - W^{-1} X (X' W^{-1} X)^{-1} X' W^{-1}. \tag{2.8} \]

Following Bard (1974, p. 65), we call \( L^*_2 \) the concentrated log-likelihood function.

Since \( L_1^* \) and \( L_2^* \) are nonlinear functions of the elements of \( \theta \), and \( \theta \) is constrained to a known parameter space \( \Theta \), the determination of ML estimates \( \hat{\theta} \) and \( \hat{\theta} \) is a well-formulated constrained nonlinear optimization problem to which any one of a variety of iterative numerical algorithms can be applied. The most commonly used and well studied of these algorithms are gradient algorithms. Gradient algorithms could be used to maximize \( L \), \( L_1^* \), or \( L_2^* \); however, for ease of exposition we restrict attention to gradient algorithms as they apply to the maximization of \( L_1^* \).

In a gradient algorithm, one computes the \((k+1)\)st iterate \( \hat{\theta}^{(k+1)} \) by updating the \( k \)th iterate \( \hat{\theta}^{(k)} \) according to the equation

\[ \hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \rho^{(k)} M^{(k)} q^{(k)}, \tag{2.9} \]

where \( \rho^{(k)} \) is a scalar, \( M^{(k)} \) is an \( m \times m \) matrix, and \( q^{(k)} \) is the vector
gradient of $L^*_1$ evaluated at $\theta = \hat{\theta}^{(k)}$, i.e.,
$\hat{\theta}^{(k)} = \frac{\partial L^*_1}{\partial \theta} \bigg|_{\theta = \hat{\theta}^{(k)}}$. The
corresponding matrix product of $M^{(k)}$ and $\hat{\theta}^{(k)}$ can be thought of as defining the search
direction (relative to the kth iterate $\hat{\theta}^{(k)}$), while $\rho^{(k)}$ serves to
define the size of the step to be taken in that direction.

Two gradient algorithms that are applicable to the maximization of
a log-likelihood function are the Newton-Raphson procedure and the method
of scoring. In the Newton-Raphson procedure, $M^{(k)} = \{J^{(k)}\}^{-1}$ and $\rho^{(k)} = 1$,
where $J^{(k)}$ is the mxm matrix whose (i,j)th element is $-\frac{\partial^2 L^*_1}{\partial \theta_i \partial \theta_j} \bigg|_{\theta = \hat{\theta}^{(k)}}$. In the method of scoring, $M^{(k)} = \{B^{(k)}_{\theta \theta}\}^{-1}$ and $\rho^{(k)} = 1$, where $B^{(k)}_{\theta \theta}$ is the
Fisher information matrix associated with $L^*_1$ evaluated at $\hat{\theta}^{(k)}$, i.e., $B^{(k)}_{\theta \theta}$ is the mxm matrix whose (i,j)th element is $E\left\{ -\frac{\partial^2 L^*_1}{\partial \theta_i \partial \theta_j} \bigg|_{\theta = \hat{\theta}^{(k)}} \right\}$. Thus, the method of scoring is identical to the Newton-Raphson procedure
except that the second-order partial derivatives are replaced by their
expectations.

Certain difficulties in the convergence properties of the Newton-
Raphson and scoring algorithms can be overcome by introducing certain
modifications. In the modified versions, $\rho^{(k)}$ is not automatically set
equal to one. Rather, $\rho^{(k)}$ is determined so that $L^*_1(\hat{\theta}^{(k+1)};y)$ is at
least somewhat larger than $L^*_1(\hat{\theta}^{(k)};y)$. Negative values of $\rho^{(k)}$ are
permissible.

In order to implement an iterative algorithm, an initial estimate
$\hat{\theta}^{(0)}$ must be supplied. The choice of starting value may be based on
prior knowledge or on the results of one of the two-step estimation
procedures described in Section 2.1.

So far, in our discussion of numerical algorithms, we have not accounted for the fact that in maximizing $L$, $L_1^*$, or $L_2^*$, $\theta$ must be restricted to the specified parameter space $\theta$ which is equal to $\theta_0$ or equal to some subset of $\theta_0$. Most weakly stationary RFLM's can be parametrized so that $\theta_0$ is of the form

$$\Theta_0 = \{\theta: \theta_1 > 0, \ldots, \theta_m > 0\}.$$  \hfill (2.10)

For example, in the case of the covariograms (1.1)-(1.3), $\Theta_0 = \{\theta: \theta_1 > 0, \theta_2 > 0\}$.

Both the Newton-Raphson and scoring algorithms can yield iterates which lie outside the parameter space $\theta$. In fact, both algorithms can converge to a value of $\theta$ not in $\Theta$. Techniques for modifying the algorithms to accommodate constraints of the form (2.10) include partial stepping strategies (see, e.g., Jennrich and Sampson, 1976), interior penalty techniques such as that of Carroll (1961), the gradient projection method, and the transformation technique. The latter three of these four techniques are reviewed by Harville (1977) in the context of mixed and random ANOVA models; their application to an RFLM is essentially the same.

In order for a gradient algorithm to be applicable, the first-order partial derivatives of $L_1^*$ with respect to the elements of $\theta$ must exist at $\theta^{(0)}$ and at all iterates produced by the algorithm. Moreover, the second-order partial derivatives of $L_1^*$ with respect to the elements of $\theta$ must exist at $\theta^{(0)}$ and at all subsequent iterates if the Newton-Raphson
or scoring procedure is to be employed. This is equivalent to requiring, for all \((s,t) \in \mathbb{R} \times \mathbb{R}\), that the second-order partial derivatives of \(C(s,t;\theta)\) with respect to the elements of \(\theta\) exist at \(\theta^{(0)}\) and at all subsequent iterates. This property is ensured if \(\theta\) is restricted to \(\Theta_1\), where \(\Theta_1\) is given by (1.13). For some RFLM's, e.g., those for which \(Y\) has covariogram (1.1) or (1.2), \(\Theta_1 = \Theta_0\); thus, no modification is actually needed. Note further that no modification is needed for mixed and random ANOVA models since, in these models, \(V\) is linear in the parameters.

It is clear that in estimating the parameters of an RFLM by maximum likelihood, there is considerable flexibility with respect to the function to be maximized (i.e., \(L\), \(L^*_1\), or \(L^*_2\)), the algorithm used to maximize that function, and the method for dealing with the constraints. We now review some results by previous authors on ML estimation as applied to RFLM's.

Mardia and Marshall (1984) described how the method of scoring could be applied to the maximization of the full log-likelihood function \(L\) of an RFLM, though they did not take into account the constraints on the parameters. They preferred the method of scoring to the Newton-Raphson algorithm, apparently because the expectations of the second-order partial derivatives of \(L\) are easier to compute than the second order partial derivatives themselves. The derivative formulae needed to obtain ML estimates of \(\hat{\beta}\) and \(\hat{\theta}\) by either algorithm are well-known. We state these formulae in the following lemma.
Lemma 2.1

Suppose that $\theta \in \Theta_1$. For the full log-likelihood function $L$:

(i) The vector of first-order partial derivatives with respect to the elements of $\phi$ is $q_\phi = (q_\beta, q_\theta)'$, where

$$ q_\beta = -x'v'^{-1}x + x'v'^{-1}y, $$

and where $q_\theta$ is an mx1 vector whose ith element is

$$ -\frac{1}{2} \text{tr} \left( v'^{-1} \frac{\partial v}{\partial \theta_i} \right) + \frac{1}{2} (y-x\beta)'v'^{-1} \frac{\partial v}{\partial \theta_i} v'^{-1} (y-x\beta) \quad (i=1, \ldots, m). $$

(ii) The matrix of second-order partial derivatives with respect to the elements of $\phi$ is

$$ J = \begin{bmatrix} J_{\beta \beta} & J_{\beta \theta} \\ J_{\theta \beta} & J_{\theta \theta} \end{bmatrix}, $$

where

$$ J_{\beta \beta} = -x'v'^{-1}x, $$

$J_{\beta \theta}$ has ith column

$$ x'v'^{-1} \frac{\partial v}{\partial \theta_i} v'^{-1}x - x'v'^{-1} \frac{\partial v}{\partial \theta_i} v'^{-1}y \quad (i=1, \ldots, m), $$

and $J_{\theta \theta}$ has $(i,j)$th term
\[-\frac{1}{2} \text{tr}(V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} - V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j})
\]

\[+ (y - \hat{X} \beta) V^{-1} (\frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j} + \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j} - \frac{\partial^2 V}{\partial \theta_i \partial \theta_j}) V^{-1} (y - \hat{X} \beta))\]

\[(i=1, \ldots, m; \quad j=1, \ldots, m).\]

(iii) The information matrix (which, by definition, is given by

\[-E(J)\] is

\[B = \begin{bmatrix}
B_{\beta \beta} & 0 \\
0 & B_{\theta \theta}
\end{bmatrix},\]

where

\[B_{\beta \beta} = X' V^{-1} X\]

and the \((i,j)\)th element of \(B_{\theta \theta}\) is

\[\frac{1}{2} \text{tr}(V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j}) \quad (i=1, \ldots, m; \quad j=1, \ldots, m).\]

\[\text{Proof: } \text{See Searle (1970).}\]

Taking into account the block diagonal form of \(B\), the \(\hat{\theta}\) and \(\hat{\beta}\) components of the \((k+1)\)st iterate \(\hat{\phi}^{(k)}\) of the method of scoring are

\[\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + (B_{\theta \theta})^{-1} (k)\]

and

\[\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + (B_{\beta \beta})^{-1} (k)\]

(2.11)
\[ \hat{\beta}^{(k+1)} = [x'(v^{(k)})^{-1}x - x'(v^{(k)})^{-1}y, \]  

(2.12)

respectively, where \( v^{(k)} = v(\hat{\beta}^{(k)}) \).

Mardia and Marshall (1984) considered the following modification of this algorithm. For the initial value of \( \hat{\beta} \), equations (2.11) could be iterated until some convergence criterion was satisfied. Then, a new estimate of \( \hat{\beta} \) could be obtained by iterating equations (2.12) just once. Equations (2.11) could then be iterated again, using the updated estimate of \( \hat{\beta} \), until another convergence criterion was satisfied. This process could be repeated until the estimates of both \( \hat{\beta} \) and \( \hat{\theta} \) satisfied some convergence criterion. Mardia and Marshall (1984) reported that the unmodified scoring algorithm is more efficient than the modified algorithm (on the average).

The direct implementation of scoring or Newton-Raphson to the maximization of the full log-likelihood function \( L \) ignores the result that leads to expression (2.3) and, consequently, is likely to be less efficient than a procedure that first computes \( \hat{\theta} \) by maximizing \( L'_1 \) with respect to \( \hat{\theta} \) and then sets \( \hat{\beta} = \hat{\beta}(\hat{\theta}) \). Further, it is to be expected that if \( Y \) is \( \theta_1 \)-factorable, then the ML estimates could be obtained more efficiently by maximizing \( L'_2 \) than by maximizing \( L'_1 \).

Cook and Pocock (1983) estimated the parameters of an RFIM by maximizing \( L'_2 \) via a grid search rather than via an iterative procedure. They indicated that a grid search is generally superior to iterative procedures, though they admitted that this conclusion may have been due...
to the flat nature of the likelihood surface for their particular set of data.

In the case of a one-dimensional random field, Kitanidis and Vomvoris (1983) proposed that the log-likelihood function of the vector of differences \( \mathbf{w} = (y_1 - y_2, y_2 - y_3, \ldots, y_{n-1} - y_n) \) be maximized rather than \( L \). The log-likelihood function associated with \( \mathbf{w} \) is

\[
L_w(\theta, \mathbf{w}) = -\frac{1}{2} \mathbf{v}^T A \mathbf{v} - \frac{1}{2} (\mathbf{w} - \mathbf{Ax})^T (A^T)^{-1} (\mathbf{w} - \mathbf{Ax}),
\]

where \( A \) is an \((n-1)\times n\) matrix of the form

\[
A = \begin{bmatrix}
1 & -1 \\
& 1 & -1 \\
& & \ddots & \ddots \\
& & & 1 & -1
\end{bmatrix}.
\]

Under the hypothesis of weak stationarity, \( \mathbf{x} \mathbf{e} = \beta \mathbf{1}_n \), where \( \beta \) is a scalar, implying that \( \mathbf{Ax} = 0 \) and hence that \( L_w \) does not depend on \( \beta \). Maximizing \( L_w \) rather than \( L \) requires a sacrifice of one observation, but Kitanidis and Vomvoris, believing that the bias reduction in the estimate of \( \theta \) could be substantial, advocated this technique for weakly stationary one-dimensional random fields. In regard to a computational algorithm, Kitanidis and Vomvoris suggested that \( L_w \) be maximized by the method of
scoring, but like Mardia and Marshall (1984), they ignored the question of how to account for constraints on $\beta$.

The Kitanidis-Vomvoris approach may be useful for solving one-dimensional geostatistical problems but it is not appropriate for "one-dimensional" spatial experiments, i.e., experiments in which the observational sites $R_1, R_2, \ldots, R_n$ comprise a single row of plots (e.g., Configuration I), because in the latter context $X$ is generally not a scalar multiple of a vector of ones and the estimation of $\beta$ (or at least certain estimable functions of the elements of $\beta$) is of great interest. However, an extension of the Kitanidis-Vomvoris approach to accommodate an arbitrary $X$-matrix is possible; the extension is known as restricted maximum likelihood estimation and is discussed in the following section. Restricted ML estimation is also more general than the Kitanidis-Vomvoris approach in that it can be used in connection with a random field of any dimension.

2.3. Restricted Maximum Likelihood

Implementations of the maximum likelihood approach such as those proposed by Mardia and Marshall (1984), by Cook and Pocock (1983), and (in a special case) by Kitanidis and Vomvoris (1983) are superior in several respects to the ad hoc methods described in Section 2.1. Some advantages of the maximum likelihood approach are:

1. the estimates of $\beta$ inherently belong to $\theta$;

2. the covariance matrix of the estimators can be estimated, since the inverse of the Fisher information matrix associated with the
log-likelihood function is under certain conditions, the "large-sample"
covariance matrix of the ML estimates. The meaning of "large-sample"
in the RFLM context is discussed in Chapter 4;

(3) the implementations are appropriate for any RFLM, regardless of
whether $F_Y$ is weakly stationary or whether $F_Y$ satisfies the
intrinsic hypothesis.

Notwithstanding its advantages, the maximum likelihood approach has
several drawbacks which we have previously noted and which may be
summarized as follows:

(1) extensive computations are required;
(2) the form (e.g., multivariate normal) of the distribution of $Y$ must
   be specified;
(3) the estimate of $\theta$ may be badly biased in small to moderately sized
   samples when $\theta$ must be estimated.

The first two of these drawbacks may not be serious in practice. In
many important geostatistical and spatial experimental situations, the
necessary computations can (as is discussed in Chapter 3) be drastically
reduced by taking advantage of the special features of these situations.
Furthermore, if the observations do not appear to be multivariate normal,
a transformation may render them more so; in any case, the maximization
of $L$ may seem sensible (from an ad hoc point of view) even when $Y$ is not
multivariate normal, since the likelihood equations consist of the
differences between the observed and expected values of certain linear
and quadratic forms.
While the first two of the aforementioned drawbacks of ML estimation may not be serious in practice, the bias of the ML estimator of \( \hat{\theta} \) has been viewed as a serious drawback. In fact, this bias appears to be the major reason why geostatisticians have, until very recently, ignored ML and instead have emphasized the use of variograms, generalized covariance functions, and ad hoc estimation procedures.

In many statistical settings the bias of an ML estimator can be ignored when \( n \) is reasonably large, say \( n > 30 \). However, this is not necessarily true for ML estimates of the parameters of RFLMs. Mardia and Marshall (1984) reported the results of a small simulation study in which the bias properties of \( \hat{\beta} \) and \( \hat{\sigma} \) in samples of size \( n = 36, 64, \) and 100 were examined for a particular RFLM in which \( m(s;\beta) = \beta \) (a constant) for all \( s \in \mathbb{R}^d \). They reported the bias of \( \hat{\beta} \) to be insignificant but found considerable bias in one or more of the elements of \( \hat{\sigma} \), even when \( n = 100 \). Mardia and Marshall apparently did not realize that, for this RFLM, \( \hat{\beta} \) is in fact unbiased as an immediate consequence of results proved by Kackar and Harville (1981). However, it appears that in order for the bias of the ML estimates of the covariance parameters of an RFLM to be negligible when appreciable spatial correlation is present and the parameters of the mean function must be estimated, a very large value of \( n \) may be required. A heuristic explanation for this is that the "effective" number of independent observations is much less than \( n \) when observations are highly correlated.

Bias is also a problem in the ML estimation of the variance components of mixed and random ANOVA models (Harville, 1977). To reduce
this bias, a variant of ML estimation known as restricted maximum likelihood (REML) estimation was developed by Patterson and Thompson (1971, 1974). Although developed primarily for applications to mixed and random ANOVA models, the same general method can be used to estimate the parameters of any mixed linear model, including an RFLM. Surprisingly, although REML estimates of variance components have generally proven to be less biased than their ML counterparts (Swallow and Monahan, 1984), the potential of REML to reduce the bias of RFLM parameter estimates has apparently not yet been recognized. We propose that REML estimation should be considered for RFLMs as an alternative to ML estimation. To help justify this proposal, the empirical bias and mean square error of ML and REML estimators of the parameters of RFLMs are compared in Section 2.4. In the remainder of this section we describe the essential features of the REML approach.

In REML, the method of maximum likelihood is applied to certain linearly independent linear combinations of the observations called error contrasts, rather than to the observations themselves.

**Definition 2.2**

Under model (1.9), a linear combination \( a'y \) of the elements of \( y \) is called an error contrast if \( E(a'y) = 0 \) for all \( \hat{\beta} \) and all \( \hat{\theta} \in \Theta \).

Clearly, \( a'y \) is an error contrast if and only if \( X'a = 0 \).
Definition 2.3

A set of error contrasts \{a_1 y, a_2 y, \ldots, a_k y\} is said to be linearly independent if \(a_1, a_2, \ldots, a_k\) are linearly independent vectors.

The maximum number of linearly independent error contrasts in any set of error contrasts is \(n-p^*\) [recall \(p^* = \text{rank}(X)\)] (Harville, 1977).

Let \(u = A y\) represent any vector of \(n-p^*\) linearly independent error contrasts. If \(y\) has a multivariate normal distribution, then, under model (1,11), \(u\) has a multivariate normal distribution with mean vector 0 (since \(A X = 0\)) and covariance matrix \(A VA\). (Note that this distribution is free of \(\beta\).) Now, \(A VA\) is positive definite because \(A\) necessarily has full column rank. Hence the log-likelihood function associated with \(u\), apart from an additive constant, is

\[
L_u(\theta; u) = -\frac{1}{2} \log|VA| - \frac{1}{2u} (A VA)^{-1} u. \tag{2.13}
\]

Lemma 2.2

The log-likelihood function (2.13) and the log-likelihood function associated with any other set of \(n-p^*\) linearly independent error contrasts differ by at most an additive constant (which does not depend on \(\beta\) or \(\theta\)) from the function

\[
L_1(\theta; y) = -\frac{1}{2} \log|V| - \frac{1}{2} \log|X^* V^{-1} X^*| - \frac{1}{2 y} P_v y, \tag{2.14}
\]

defined for \(\theta \in \Theta\), where \(X^*\) is an \(n \times p^*\) matrix whose columns are any \(p^*\) linearly independent columns of \(X\), and \(P_v\) was given by (2.4).

An REML estimate of $\tilde{\theta}$ is defined to be a value $\tilde{\theta}$ of $\theta$ at which $L_1$ attains its maximum over $\theta$. As we did in the context of ML estimation, we shall require that if two different values of $y$ produce the same restricted log-likelihood function, then the same value of $\tilde{\theta}$ will be chosen to be the REML estimate for the second value of $y$ as for the first value of $y$. An REML estimate of $\tilde{\theta}$ is $\tilde{\theta} = \tilde{\beta}(\tilde{\theta})$. (Actually, there is no inherent REML estimator of $\tilde{\beta}$ because $L_1$ does not depend on $\tilde{\beta}$; however, one might justify the use of $\tilde{\beta}(\tilde{\theta})$ since, with $\tilde{\theta}$ set equal to $\tilde{\theta}$, $\tilde{\beta}(\tilde{\theta})$ maximizes the ordinary log-likelihood function $L$ with respect to $\tilde{\theta}$.)

When $f_\theta$ is $\theta_1$-factorable, the restricted log-likelihood function $L_1$, like $L^*_1$, can be concentrated by substituting

$$\theta_1 = \frac{1}{n-p} \left[ \text{y-X}^\top \hat{\beta}(\theta_2) \right]^\top \left[ W(\theta_2) \right]^{-1} \left[ \text{y-X}^\top \hat{\beta}(\theta_2) \right]$$

into (2.14), thereby reducing (by one) the dimensionality of the maximization problem. The concentrated restricted log-likelihood function differs by no more than an additive constant from

$$L_2(\theta_2; y) = - \frac{1}{2} \log |W| - \frac{1}{2} \log |X^\top W^{-1} X|$$

$$- \left( \frac{n-p}{2} \right) \log (y^\top P_w y),$$

(2.15)

where $P_w$ was given by (2.8).
Any of the gradient algorithms discussed in the previous section [modified to accommodate constraints of the form (2.10)] can be applied to the maximization of or $L_2$. The necessary first- and second-order partial derivatives of $L_1$ and $L_2$ and the expectations of the second-order partial derivatives of $L_1$ are given by the following lemma, which is a generalization of results given by Callanan (1985, pp. 105-107).

**Lemma 2.3**

Suppose that $\theta \in \Theta_1$. Then for $i = 1, \ldots, m$ and $j = 1, \ldots, m$,

$$
\frac{\partial L_1}{\partial \theta_i} = - \frac{1}{2} \text{tr}(P_v \frac{\partial^2 v}{\partial \theta_i} + \frac{1}{2} y' P_v \frac{\partial v}{\partial \theta_i} P_v y),
$$

$$
\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_j} = - \frac{1}{2} \text{tr}[(P_v \frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j})] + \frac{1}{2} y' P_v (\frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - 2 \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j} P_v y),
$$

and

$$
\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_j} = - \frac{1}{2} \text{tr}(P_v \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j}).
$$

Furthermore, for $i = 2, \ldots, m$ and $j = 2, \ldots, m$,

$$
\frac{\partial L_2}{\partial \theta_i} = - \frac{1}{2} \text{tr}(P_w \frac{\partial^2 w}{\partial \theta_i} + \frac{n-p}{2} (y' P_w y)^{-1} y' P_w \frac{\partial w}{\partial \theta_i} P_w y),
$$

and
\[ \frac{\partial^2 L_2}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} \text{tr}[P_w(\frac{\partial^2 W}{\partial \theta_i \partial \theta_j} - \frac{\partial W}{\partial \theta_i} P_w \frac{\partial W}{\partial \theta_j})] + \frac{n-p}{2} \left( (y' P_{w^*})^{-1} y' P_w (\frac{\partial W}{\partial \theta_i} - 2 \frac{\partial W}{\partial \theta_i} P_w \frac{\partial W}{\partial \theta_j}) P_{w^*} \right) + \left( (y' P_{w^*})^{-2} (y' P_w \frac{\partial W}{\partial \theta_i} P_{w^*}) (y' P_w \frac{\partial W}{\partial \theta_j} P_{w^*}) \right). \]

Proof:

From (2.14),

\[ \frac{\partial L_1}{\partial \theta_i} = -\frac{1}{2} \text{tr}(y' (v^{-1} \frac{\partial V}{\partial \theta_i} - \frac{\partial V}{\partial \theta_i} X (X' y^{-1} X')^{-1} (X' y^{-1} X')^{-1} \frac{\partial V}{\partial \theta_i} y^{-1} V^{-1} X')) \]

\[ - \frac{1}{2} y' \frac{\partial P_v}{\partial \theta_i} y. \]

Now, since \( X(X' v^{-1} X')^{-1} X \) = \( X(X' v^{-1} X')^{-1} X' \), we have that

\[ \frac{\partial P_v}{\partial \theta_i} = \frac{\partial V^{-1}}{\partial \theta_i} - \frac{\partial V^{-1}}{\partial \theta_i} X (X' v^{-1} X')^{-1} X' v^{-1} \]

\[ - v^{-1} X (X' v^{-1} X')^{-1} X' v^{-1} \frac{\partial v}{\partial \theta_i} \]

\[ - v^{-1} X' (X' v^{-1} X')^{-1} X' v^{-1} \frac{\partial v}{\partial \theta_i} \]

\[ = - v^{-1} \frac{\partial V}{\partial \theta_i} v^{-1} + v^{-1} \frac{\partial V}{\partial \theta_i} v^{-1} X (X' v^{-1} X')^{-1} X' v^{-1} \]
Thus,
\[
\frac{\partial L_1}{\partial \theta_i} = -\frac{1}{2} \text{tr}(\partial V^{-1} \frac{\partial V}{\partial \theta_i}) + \frac{1}{2} \text{tr}[V^{-1} X^*(X' V^{-1} X')^{-1} X'^* V^{-1} \frac{\partial V}{\partial \theta_i} P_v] + \frac{1}{2} V' P_v \frac{\partial V}{\partial \theta_i} P_v V' 
\]
\[
= -\frac{1}{2} \text{tr}(P_v \frac{\partial V}{\partial \theta_i}) + \frac{1}{2} V' P_v \frac{\partial V}{\partial \theta_i} P_v V' .
\]

Also, using (2.16),
\[
\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} \text{tr}(-P_v \frac{\partial V}{\partial \theta_j} P_v \frac{\partial V}{\partial \theta_i} + P_v \frac{\partial^2 V}{\partial \theta_i \partial \theta_j}) 
\]
\[
+ \frac{1}{2} V' (-P_v \frac{\partial V}{\partial \theta_j} P_v \frac{\partial V}{\partial \theta_i} P_v + P_v \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} P_v 
\]
\[
- P_v \frac{\partial V}{\partial \theta_i} P_v \frac{\partial V}{\partial \theta_j} P_v V') .
\]
\[
= - \frac{1}{2} \text{tr}[P_v \left( \frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j} \right)]
\]

\[
+ \frac{1}{2} y P_v \left( \frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - 2 \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j} \right) P_v y,
\]

and

\[
E(\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_j}) = - \frac{1}{2} \text{tr}[P_v \left( \frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j} \right)]
\]

\[
+ \frac{1}{2} \text{tr}[P_v \left( \frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - 2 \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j} \right) P_v]
\]

since \( P_x = 0 \). Now, \( P_v P_v = V \), as is easily verified, so that

\[
E(\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_j}) = - \frac{1}{2} \text{tr}[P_v \left( \frac{\partial^2 v}{\partial \theta_i \partial \theta_j} - \frac{\partial v}{\partial \theta_i} P_v \frac{\partial v}{\partial \theta_j} \right)].
\]

Recalling (2.15) and (2.16), and using \( P_w = (1/\theta_1) P_v \), we find that

\[
\frac{\partial^2 L_2}{\partial \theta_i} = - \frac{1}{2} \text{tr}(W^{-1} \frac{\partial W}{\partial \theta_i})
\]

\[
+ \frac{1}{2} \text{tr}\left[\left(Y^* W^{-1} W^{-1} \frac{\partial W}{\partial \theta_i} (X^* W^{-1} X)^{-1}\right)\right]
\]

\[
+ \left(\frac{n-p}{2}\right) (y P_w y)^{-1} y P_w \frac{\partial W}{\partial \theta_i} P_w y
\]

\[
= - \frac{1}{2} \text{tr}(P_w \frac{\partial W}{\partial \theta_i}) + \left(\frac{n-p}{2}\right) (y P_w y)^{-1} y P_w \frac{\partial W}{\partial \theta_i} P_w y,
\]
and

\[
\frac{\partial^2 L_2}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} \text{tr}(-P_w \frac{\partial^2 w}{\partial \theta_i \partial \theta_j} P_w \frac{\partial^2 w}{\partial \theta_i \partial \theta_j} + P_w \frac{\partial^2 w}{\partial \theta_i} \frac{\partial^2 w}{\partial \theta_j})
\]

\[
+ \left(\frac{n-p}{2}\right) \{(y' P_w y)^{-1} (y' P_w \frac{\partial^2 w}{\partial \theta_i} P_w y) - (y' P_w \frac{\partial^2 w}{\partial \theta_i} P_w y)^2\}
\]

\[
- P_w \frac{\partial^2 w}{\partial \theta_i} P_w \frac{\partial^2 w}{\partial \theta_j} y
\]

\[
= -\frac{1}{2} \text{tr}[P_w \left(\frac{\partial^2 w}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 w}{\partial \theta_i} \frac{\partial^2 w}{\partial \theta_j}\right)]
\]

\[
+ \frac{n-p}{2} \{(y' P_w y)^{-1} (y' P_w \frac{\partial^2 w}{\partial \theta_i} P_w y) - 2 \frac{\partial^2 w}{\partial \theta_i} P_w \frac{\partial^2 w}{\partial \theta_j} P_w y\}
\]

\[
+ (y' P_w y)^{-2} (y' P_w \frac{\partial^2 w}{\partial \theta_i} P_w y)(y' P_w \frac{\partial^2 w}{\partial \theta_j} P_w y)\].
\]

Q.E.D.

As discussed earlier, the Kitanidis-Vomvoris approach, in which the log-likelihood function \(L_w\) of first differences of observations of a weakly stationary random field is maximized, is a special case of the REML approach. This can now be easily seen, since the vector \(\mathbf{y}\) of first differences is a vector of linearly independent error contrasts under an RFIM for which \(\mathcal{F}_Y\) is weakly stationary.
There is some controversy as to whether, as a consequence of reducing the data to error contrasts, the REML approach results in a loss in information. Several statisticians (e.g., Patterson and Thompson, 1971; Harville, 1977), upon observing that both $L^*_1$ and $L_1$ depend on the data only through a set of linearly independent error contrasts, have argued that the REML approach does not ignore any information actually used by the full ML approach (although the two approaches do use that information somewhat differently). Others (e.g., Rao, 1977) have not been persuaded by such arguments and have maintained that the estimation of $\theta$ and $\beta$ should be based on the full likelihood function $L$. The results of Corbeil and Searle (1976) and the Monte Carlo investigations of Swallow and Monahan (1984) for random and mixed ANOVA models suggest that which of the two approaches (REML or ML) is "best" depends on the specifics of the model (such as the values of $n$ and $p^*$) and on the true value of $\theta$. Similar investigations for RFIM's need to be carried out; the results of one such study are given in the following section.

2.4. A Small-Sample Comparison of ML and REML Estimators

In this section we report the results of a Monte Carlo simulation study in which the bias and mean squared error (MSE) of ML estimators are compared to those of REML estimators in the context of an RFIM. No claim is made that this study is exhaustive; indeed, the computing resources required to make the investigation as thorough as that conducted by Swallow and Monahan (1984) for one-way random ANOVA models
would have been prohibitive. However, the present study is extensive enough to demonstrate that the large small-sample biases experienced in ML estimation (Matheron, 1971; Mardia and Marshall, 1984) are essentially eliminated in REML estimation.

For convenience we take the spatial setting to be one where \( \mathbf{Y} \) is a two-dimensional isotropic random field that is observed at the points of an \( N \times N \) square lattice, with unit spacing between points. This spatial configuration of observations is a special case of Configuration II. This setting is chosen because (as will become evident in Chapter 3) its structure is one for which the computational burden of parameter estimation can be reduced considerably. The vector of \( n = N^2 \) observations was generated according to the model

\[
\mathbf{y} = \mathbf{m} + \mathbf{e},
\]

where the \( i \)th element of \( \mathbf{m} \) is obtained by evaluating the function

\[
\mathbf{m}(\mathbf{s}; \alpha) = \alpha_1 + \alpha_2 s_1 + \alpha_3 s_2 + \alpha_4 s_1^2 + \alpha_5 s_2^2 + \alpha_6 s_1 s_2,
\]

at the points of the lattice, and where \( s_1 \) and \( s_2 \) represent the elements of \( \mathbf{s} \), and \( \mathbf{e} \) has a multivariate normal distribution with mean vector \( \mathbf{0} \) and covariance matrix \( \mathbf{V} \). The elements of \( \mathbf{V} \) are determined by the covariogram

\[
\bar{C}(r; \theta) = \begin{cases} 
\theta_1 (1 - \frac{3r}{2\theta_2} + \frac{r^3}{2\theta_2^3}), & 0 \leq r \leq \theta_2, \\
0, & \text{otherwise}.
\end{cases}
\]
Thus, the mean function of $F_\gamma$ is a full second-order polynomial function of the coordinates $s_1$ and $s_2$, and the covariogram is the isotropic spherical covariogram (1.3). We refer to this RFLM as RFLM #1.

In the first part of this study, the effect of increasing sample size on the bias and MSE of ML and REML estimates was investigated. Three choices of $N$ — $N = 4, 6, \text{ and } 8$ — were considered, corresponding to $n = 16, 36, \text{ and } 64$, respectively. The true values of the parameters were taken to be $\alpha = (10, 2, 1, 5, 0, 0)'$ and $\theta = (1, 3)'$. The true value of $\theta$, the stated sample sizes, and the spherical covariogram were chosen to make the results somewhat comparable with the simulation results of Mardia and Marshall (1984).

To obtain $y$ we first generated a realization of an $nx1$ multivariate normal random vector $z$ having mean 0 and identity covariance matrix. Each element of $z$ was obtained using the normal pseudo-random number generator of the Statistical Analysis System (SAS User's Guide, 1979, p. 444). Once $z$ was obtained, we put

$$y = m + U z,$$

where $U$ is an upper triangular matrix such that $U'U = V$ (such a matrix always exists when $V$ is positive definite; see, e.g., Theorem 8.6.2 of Graybill, 1983).

For each choice of $N$ we generated 100 independent replications of $y$. For each replication we obtained the ML and REML estimates (which
in this case are unique) of the coefficients of the full second-order response surface \( m(s; a) \) and of the covariogram parameters \( \theta_1 \) and \( \theta_2 \).

This was accomplished by maximizing the concentrated functions \( L_2^* \) and \( L_2 \), respectively, using a Newton-Raphson algorithm with unit stepsize, modified to account for the constraint \( \theta_2 > 0 \). In addition, \( L_2 \) and \( L_2^* \) were evaluated at each of several values of \( \theta_2 \) to ascertain if the Newton-Raphson procedure had indeed converged to a global maximum.

In Table 2.1, the empirical bias and empirical MSE of the ML estimators \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) and the REML estimators \( \tilde{\theta}_1 \) and \( \tilde{\theta}_2 \) over 100 replications of RFIM #1 are displayed. Results for the estimators \( \hat{\gamma} \) and \( \tilde{\gamma} \) of \( \gamma \) are not given because these estimators are unbiased under RFIM #1 (Kackar and Harville, 1981) and because differences in MSE between the elements of \( \hat{\gamma} \) and \( \tilde{\gamma} \) were imperceptibly small. Estimates of the standard deviation of the empirical bias and empirical MSE are also given in Table 2.1. The standard deviation of the empirical bias was estimated by the sample standard deviation of the 100 bias estimates. The standard deviation of the empirical MSE was estimated by the bootstrap estimate of the standard deviation (Efron, 1982) based on 100 bootstrap replications of size 50.

The most interesting feature of Table 2.1 is the marked superiority of \( \tilde{\theta}_1 \) and \( \tilde{\theta}_2 \) to \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) when \( N = 4 \) and \( N = 6 \). Though the ML estimators had smaller variances than their REML counterparts, the differences in bias were sufficiently large that the MSE's of \( \tilde{\theta}_1 \) and \( \tilde{\theta}_2 \) were smaller than those of \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \). The superiority of REML estimation relative to ML estimation appeared to decrease, however, as sample size increased:
Table 2.1. Empirical bias (a), and empirical MSE (b) of RFLM #1 covariogram parameter estimators. Estimates of the standard deviations of these estimators are given in parentheses.

<table>
<thead>
<tr>
<th>N</th>
<th>( \hat{\theta}_1 )</th>
<th>( \hat{\theta}_2 )</th>
<th>( \tilde{\theta}_1 )</th>
<th>( \tilde{\theta}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-0.760 (.010)</td>
<td>-1.990 (.010)</td>
<td>-0.421 (.032)</td>
<td>-1.041 (.077)</td>
</tr>
<tr>
<td>6</td>
<td>-0.488 (.017)</td>
<td>-1.230 (.055)</td>
<td>-0.016 (.047)</td>
<td>0.021 (.103)</td>
</tr>
<tr>
<td>8</td>
<td>-0.258 (.020)</td>
<td>-0.530 (.055)</td>
<td>0.103 (.035)</td>
<td>0.365 (.088)</td>
</tr>
</tbody>
</table>

with respect to MSE, the REML estimates were significantly better than the ML estimates when \( N = 4 \) and \( N = 6 \), but the ML estimates were slightly (but not significantly) superior when \( N = 8 \). This is not surprising because as the sample size increases (but the dimension of \( \lambda \) remains constant), the effect of the estimation of \( \lambda \) on the ML estimation of \( \theta \) diminishes. Note, however, that the REML estimates are still less biased than their ML counterparts even when \( N = 8 \).

Several interesting results in addition to those displayed in Table 2.1 emerged from this investigation. The results pertain to three general aspects of the Newton-Raphson procedure: the rate at which it
converges to a local maximum, the problem of constraints on the parameter space, and the problem of determining whether a point of convergence is a global maximum.

There was no appreciable difference in this study between ML and REML estimation with respect to the rate at which the Newton-Raphson procedure converged. In both cases the convergence was usually quite rapid; convergence of the estimate of $\theta_2$ to within .00001 of a point of local maximum often occurred in four to eight iterations, only rarely requiring more than 15 iterations.

The method of step halving was used to enforce the constraint on $\theta_2$ in the maximization of $L_2^*$ and $L_2$. In carrying out the maximization, the actual constraint on the parameter space, i.e., $\theta_2 > 0$, was replaced by the constraint $\theta_2 > 1$. This was done for the following reasons. The spherical covariogram vanishes when $r > \theta_2$, implying in the case of a square lattice of points with unit spacing that the observations are uncorrelated when $0 < \theta_2 \leq 1$. This, in turn, implies that $L_2^*$ and $L_2$ are flat, i.e., $3L_2^*/\partial \theta_2 = 3L_2/\partial \theta_2 = 0$, for $\theta_2 \in (0,1]$. Thus, if $L_2^*$ or $L_2$ attains its maximum at a value of $\theta_2$ in the interval $(0,1]$, then the ML or REML estimate of $\theta_2$ (and therefore of all the RFLM parameters) is not unique, and any other value in $(0,1]$ is also a maximizing value (in particular, $\theta_2 = 1$ is a maximizing value). In those trials for which this phenomenon occurred, we set the estimator of $\theta_2$ equal to one. As indicated in Table 2.2, this occurred more often for ML estimation than for REML estimation (which is consistent with the negative bias of the ML estimator), and decreased in frequency as N increased.
Table 2.2. Frequency (in 100 trials) with which the estimator of $\hat{\theta}_2$ equalled one

<table>
<thead>
<tr>
<th>N</th>
<th>$\hat{\theta}_2$</th>
<th>$\tilde{\theta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>99</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In order to check if the estimate obtained via the Newton-Raphson procedure was a point of global maximum, $L_2^*$ and $L_2$ were evaluated, for every trial, at each of the values $\hat{\theta}_2 = 1.0, \tilde{\theta}_2 = 1.1, \ldots, \tilde{\theta}_2 = 6.0$. This procedure revealed that more often than not, $L_2^*$ and $L_2$ have more than one local maximum, and that these local maxima often occur at values of $\tilde{\theta}_2$ that are quite close to each other. Not surprisingly, therefore, the Newton-Raphson procedure sometimes converged to a value that was not a point of global maximum. Because the Newton-Raphson procedure appears to converge very rapidly in the vast majority of trials, its use would be preferred (to the evaluation of $L_2^*$ or $L_2$ at numerous values of $\tilde{\theta}_2$) if the value to which the estimator converges is always a point of global maximum. Unfortunately, this was not the case in a significant number of trials of RFLM #1. Consequently, if Newton-Raphson is to be used, it is advisable to repeat the procedure several times, using a different initial value of $\tilde{\theta}_2$ on each trial. If all the trials converge to the same value, the investigator can be more confident that the value is a
global maximum. If the different trials produce more than one value of \( \theta_2 \), \( L^*_2 \) or \( L_2 \) should be evaluated at each point. As a practical matter, the point for which \( L^*_2 \) or \( L_2 \) is largest could then be regarded as the ML or REML estimate. In this particular study, by starting the Newton-Raphson procedure at each of the two values \( \theta_2 = 2.0 \) and \( \theta_2 = 4.0 \), we obtained the global maximum of \( L^*_2 \) or \( L_2 \) in more than 90% of the trials.

The fact that a Newton-Raphson algorithm (or any other iterative algorithm) does not always converge to a point of global maximum tends to make one less confident about the simulation results of Mardia and Marshall (1984), who do not mention whether or how they ascertained that the estimates obtained by their scoring algorithm were indeed points at which \( L \) achieved a global maximum.

Another part of the Monte Carlo investigation focused on how the number of parameters in the mean structure of an RFLM, i.e., the value of \( p \), affected the relative merits of the ML and REML estimation. Table 2.3 gives the empirical bias and MSE of ML and REML estimates over 100 replications of an RFLM that is exactly the same as RFLM #1 (for which \( p = 6 \)) except that its mean structure includes only the intercept term \( \alpha_1 \) (so that \( p = 1 \)). It should be noted that several extra replications of this RFLM, which we refer to as RFLM #2, were necessary when \( N = 4 \) to obtain 100 usable replications because the REML estimators did not always exist, i.e., \( L^*_2 \) or \( L_2 \) was arbitrarily close to its supremum over \( \Theta \) for large \( \theta_2 \) but did not actually attain its supremum. Table 2.3 indicates that with respect to MSE there is scant difference between the ML and
REML estimators for RFLM #2 when N = 4 and N = 6. The REML estimates are considerably less biased than their ML counterparts when N = 4, but $\hat{\theta}_2$ is less biased than $\tilde{\theta}_2$ when N = 6. These differences in bias are more than twice as large as the larger of the two estimators.

Table 2.3. Empirical bias (a), and empirical MSE (b) of RFLM #2 covariogram parameter estimators. Estimates of the standard deviations of these estimators are given in parentheses.

<table>
<thead>
<tr>
<th>(a)</th>
<th>N</th>
<th>$\hat{\theta}_1$</th>
<th>$\hat{\theta}_2$</th>
<th>$\tilde{\theta}_1$</th>
<th>$\tilde{\theta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>-.302 (.029)</td>
<td>-.662 (.083)</td>
<td>-.159 (.037)</td>
<td>-.252 (.099)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-.088 (.027)</td>
<td>-.045 (.085)</td>
<td>-.002 (.029)</td>
<td>0.200 (.089)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b)</th>
<th>N</th>
<th>$\hat{\theta}_1$</th>
<th>$\hat{\theta}_2$</th>
<th>$\tilde{\theta}_1$</th>
<th>$\tilde{\theta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>.176 (.025)</td>
<td>1.124 (.176)</td>
<td>.164 (.038)</td>
<td>1.050 (.247)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.082 (.015)</td>
<td>.723 (.161)</td>
<td>.085 (.018)</td>
<td>.834 (.180)</td>
</tr>
</tbody>
</table>

estimated standard deviations. A comparison of Table 2.1 with Table 2.3 suggests that the superiority of REML estimation to ML estimation in small samples diminishes as p decreases.

Another factor that may affect the bias and MSE properties of ML and REML estimates is the true value of $\theta$. It would seem that REML estimation would improve relative to ML estimation as the degree of
spatial correlation increases (since an increase in the degree of spatial
correlation corresponds to a decrease in the "effective" number of inde­
pendent observations). However, we did not investigate this hypothesis.

2.5. A Relationship Between REML Estimation of Covariogram
Parameters and ML Estimation of Generalized Covariance
Function Parameters

Our goal in the final section of this chapter is to reveal a rela­
tionship that exists, when \( \mathcal{F}_Y \) is an IRF-k and the RFLM satisfies certain
other conditions, between REML estimation of the parameters of the
covariogram of \( \mathcal{F}_Y \) and ML estimation of the parameters of the generalized
covariance function of \( \mathcal{F}_Y \). The latter estimation method, as it applies
to RFLM-P, was considered by Kitanidis (1983); in what follows, we indicate
how this method is essentially equivalent to REML estimation as it applies
to RFLM-P. This equivalence will be seen to have implications for re­
ducing the computational burden of REML estimation.

Consider an RFLM-P in which \( \mathcal{F}_Y \) is an IRF-k with generalized
covariance function \( \hat{G}_k(\cdot; \theta) \). Recall that, under RFLM-P, the observational
sites are points \( s_1, \ldots, s_n \). Also recall that \( K_k \) represents the nxn
matrix whose \((i,j)\)th element is \( \hat{G}_k(s_i-s_j; \theta) \). Let \( \lambda \) and \( \gamma \) be nx1 vectors.

A very important consequence of (1.8) is that if \( \sim \gamma \) and \( \sim \gamma \) are general­
ized increments of order \( k \), then

\[
\lambda_k \sim K_k \sim \gamma = \lambda \sim V \sim \gamma.
\]

(2.17)

Some additional notation must be introduced. Let \( \mathcal{M}_{kd} \) represent the
collection of monomials in \( d \) variables of degree \( \leq k \). Let \( \{m_{kd}^{(j)}(\cdot)\} \):
\( j = 1, \ldots, \ell^* \) denote these monomials. By definition, the coefficients of a generalized increment \( \lambda' y \) of order \( k \) satisfy

\[
\sum_{i=1}^{n} \lambda^{(j)}_{i \lambda} m_i(s_i) = 0
\]

for all \( j = 1, \ldots, \ell^* \). Let \( \lambda^{'1} y, \ldots, \lambda^{'n-\ell^*} y \) denote \( n-\ell^* \) generalized increments of order \( k \), such that \((n-\ell^*)x_n\) matrix

\[
A_k' = \begin{bmatrix}
\lambda^{'1}

\vdots

\lambda^{'n-\ell^*}
\end{bmatrix}
\]

has full row rank. Finally, let \( M \) denote the \( nx\ell^* \) matrix whose \( j \)th column is comprised of the elements \( \{m^{(j)}_{kd}(s_i) : i = 1, \ldots, n\} \).

In the context of RFLM-P, Kitanidis (1983) proposed that \( \theta \) be estimated by applying the method of maximum likelihood to \( A_k' y \), i.e., by maximizing the function

\[
L_3(\mu; y) = -\frac{1}{2} \log |A_k' k_k k_k' A_k| - \frac{1}{2} y' A_k' k_k k_k' A_k^{-1} y
\]

with respect to \( \mu \in \Theta \). Subsequently, this approach for the estimation of \( \mu \) shall be called the ML-IRF-k approach. One of the gradient procedures discussed in connection with the maximization of \( L \) and \( L_1 \) could be applied to the maximization of \( L_3 \); to implement one of these procedures the matrices \( k_k' \), \( k_k' / \Theta_i \), and, for Newton-Raphson, \( k_k' / \Theta_i j \) \((i = 1, \ldots, m; j = 1, \ldots, m)\) must be computed for each iterate of \( \mu \) just as \( V^p, V^p / \Theta_i \).
and $\partial^2 V/\partial \theta_i \partial \theta_j$ ($i = 1, \ldots, m; j = 1, \ldots, m$) must be computed for the maximization of $L$ or $L'$. 

The following theorem establishes that, under certain conditions, the REML estimation of $\theta$ is equivalent to the estimation of $\theta$ by the ML-IRF-k approach.

**Theorem 2.1**

Suppose, under RFLM-P, that $\mathcal{F}_Y$ is a d-dimensional IRF-k. If $\mathcal{C}(X) = \mathcal{C}(M)$, then the estimate of $\theta$ obtained by the ML-IRF-k approach is equal to the REML estimate of $\theta$.

**Proof:**

Since $\mathcal{F}_Y$ is an IRF-k, $\Lambda_k^P \Lambda_k^P = \Lambda_k^P \Lambda_k^P$ by (2.17). Thus, the ML-IRF-k approach is equivalent to maximizing

$$-rac{1}{2} \log |\Lambda_k^P \Lambda_k^P| - \frac{1}{2} \sum_k (\Lambda_k^P \Lambda_k^P)^{-1} \Lambda_k^T y.$$

Because $\Lambda_k^Y$ is a vector of generalized increments of order $k$, $\Lambda_k^M = 0$, implying that $\Lambda_k^X = 0$. This fact, together with $\text{rank}(\Lambda_k) = n-p^*$, implies that $\Lambda_k^Y$ is a vector of $n-p^*$ linearly independent error contrasts under RFLM-P. The result follows by Lemma 2.2. Q.E.D.

Though REML and the ML-IRF-k approach produce the same estimate of $\theta$ under the conditions of Theorem 2.1, there are advantages to the latter
approach when the random field is not weakly stationary. Not surprisingly, these advantages are due, in large part, to the weak stationarity of the kth-order generalized increments of an IRF-k.

One benefit of the weak stationarity of kth-order generalized increments is that $R_K$ and each of its first- and second-order partial derivatives may have a much simpler structure than that of $V^p$ and its derivatives. Further, the covariance matrix of $n,L_k$ kth-order generalized increments, which must be formed and inverted on each iteration of a Newton-Raphson or scoring algorithm, can have a much simpler structure than that of the ordinary covariance matrix $V^p$; this result can be true when the IRF-k is weakly stationary. Sufficient conditions for these structures, their precise nature, and their computational utility are examined in Section 3.5.

Another advantage of the weak stationarity of kth-order generalized increments is that it permits one to estimate $\theta$ unbiasedly by applying a two-step estimation procedure, similar to those discussed in Section 2.1, to the generalized covariance function instead of to the covariogram. This estimate could be used as the initial value of $\hat{\theta}$ by an iterative maximization algorithm associated with a subsequent ML or REML estimation procedure.

One apparent disadvantage of the ML-IRK-k approach is that by reducing the data to generalized increments of order k, some information seems to be lost. (This apparent loss of information is similar to the aforementioned apparent loss of information when the data are reduced to
error contrasts; indeed, the two are related as a consequence of Theorem 2.1.) The larger the value of k needed to obtain weak stationarity, the greater the apparent reduction in information. Hence, it would seem that an important practical consideration is the determination of the smallest value of k for which generalized increments of order k appear to be weakly stationary. Cressie and Laslett (1986) have devised a graphical method for solving this problem.
3. COMPUTATIONAL ASPECTS OF ML AND REML APPROACHES TO RFLM PARAMETER ESTIMATION

Maximum likelihood and REML approaches for estimating the parameters of RFLMs, as described in the previous chapter, are quite computation-intensive compared to the various ad hoc approaches in common use. Numerous function evaluations, several matrix additions and multiplications, and, in general, the inversion of an \( nxn \) matrix are required to compute each iterate of a Newton-Raphson or scoring algorithm [cf. expressions (2.11) and (2.12)].

In this chapter we show, in several special cases that are of considerable practical importance, that the amount of computation required to obtain ML or REML estimates of RFLM parameters can be significantly reduced. Clearly, the amount of computation is, for the most part, determined by the structure of the covariance matrix \( V \) (or, if the ML-IRF-k approach is taken, by the structure of \( K_k \)). The structure of \( V \) and \( K_k \), in turn, shall be seen to be greatly affected by the following two features of a particular spatial setting: (1) the spatial configuration of the observational sites \( R_1, \ldots, R_n \); (2) certain properties of the covariogram or generalized covariance function of \( f_c \). We shall see that for some spatial configurations and some covariograms or generalized covariance functions, \( V \) and \( K_k \) have a structure that can be exploited computationally.

It is worth noting that MIVQUE, MINQUE, and kriging require \( V \) or \( K_k \) to be formed and inverted, so the results presented in this chapter could be useful for these estimation procedures as well as for ML and REML
estimation. Furthermore, results on the structure of $V$ are useful for establishing sufficient conditions for the asymptotic normality of parameter estimators (see Chapter 4) and could be useful for solving problems of optimal spatial design.

We use the term "computation" throughout this chapter to refer to the addition or multiplication of two real numbers. (Although multiplications require greater amounts of computing time than additions, this usage is customary in the computational literature.) Defined in this way, the addition of two $mxn$ matrices requires $mn$ computations, and the post multiplication of an $mxn$ matrix by an $nxp$ matrix requires $2mnp$ computations.

We begin this chapter by considering in detail the computations needed to obtain ML and REML estimates of RFLM parameters in the case of a general spatial configuration and in three special cases: "remote" observational sites, Configuration I, and Configuration II. In each case we characterize the structure of the covariance matrix $V$ and seek ways of exploiting that structure computationally. Results pertaining to this structure when $\mathcal{F}_Y$ is weakly stationary are given first, then extended to the less restrictive case in which $\mathcal{F}_Y$ is an IRF-$k$. We also discuss efficient procedures for evaluating integrals of the form (1.12) that arise when the model adopted for $y$ is an RFLM-R.

3.1. The General Spatial Configuration

In discussing computational aspects of RFLM parameter estimation, it is important to distinguish between the covariance matrix $V^p$ associated
with RFLM-P and the covariance matrix $V^R$ associated with RFLM-R. Provided that the covariogram of $f_Y$ can be expressed in closed form, an element of $V^p$ can be obtained exactly with one function evaluation. In contrast, an element of $V^R$ is a 2d-dimensional integral having the form (1.12); often, this integral cannot be evaluated analytically and must, therefore, be approximated using a numerical integration technique. Such techniques are discussed in Section 3.6. These techniques require several evaluations of the covariogram, thereby increasing the amount of computation over that required to obtain $V^p$.

For a completely general spatial configuration, the most that can be said about $V$ is that it is an nxn symmetric positive definite matrix whose elements are functions of $\theta$. Such a matrix can have as many as $n(n+1)/2$ distinct elements. Consequently, each step of an iterative maximization algorithm requires up to $n(n+1)/2$ evaluations of the covariogram to form $V^p$. Similarly, up to $n(n+1)/2$ function evaluations are required to form each of the matrices $\partial V^p/\partial \theta_i$ and $\partial^2 V^p/\partial \theta_i \partial \theta_j$ ($i = 1, \ldots, m; j = 1, \ldots, m$), which enter, for example, in the Newton-Raphson algorithm. Thus, for RFLM-P, the total number of function evaluations needed to evaluate the covariance matrix and its partial derivatives at each iteration is, in general, $(m^2+m+1)n(n+1)/2$, while for RFLM-R an integer multiple of $(m^2+m+1)n(n+1)/2$ function evaluations may be necessary.

Recall that if $f_Y$ is $\theta_1$-factorable, then $V = \theta_1 W(\theta_2)$ for some nxn matrix $W$. In this case, we have seen that the ML and REML estimates of $\theta$ and $\theta_2$ can be obtained by maximizing $L^*_{2}$ and $L_2$, respectively, both of
which are free of $\theta_1$. For RFLM-\$P, it is easily verified that to evaluate $W$ and its first- and second-order partial derivatives requires a total of $(m^2-m+1)n(n+1)/2$ function evaluations.

Since $V$ is symmetric, its storage requirements are approximately half of what they would be for an arbitrary matrix. However, $O(n^3)$ computations are required to obtain $V^{-1}$ unless $V$ has additional structure. In the case of some (though not all) spatial configurations, weak stationarity, isotropy, and certain other properties of $F_1$ insure that $V$ will possess useful additional structure.

3.2. Remote Observational Sites

A property of all covariograms commonly used to solve geostatistical problems is that $C(s, t; \theta)$ decreases as the distance between $s$ and $t$ increases. This property motivates a notion of remoteness, which is made precise by the following definition.

**Definition 3.1**

Let $\varepsilon > 0$ be given, and let $S$ and $T$ represent two disjoint subsets of $\mathbb{R}^d$. For a given covariogram $C(s, t; \theta)$ of a $d$-dimensional random field, we say that $S$ and $T$ are $\varepsilon$-remote subsets if $|C(s, t; \theta)| < \varepsilon$ for all $s \in S$ and $t \in T$.

Clearly, this definition is dependent on the form of the covariogram in addition to the values of the underlying parameters.
The notion of remoteness made precise by Definition 3.1 is intimately related to the loosely-defined notion of the "range" of a covariogram. In the case of the isotropic spherical covariogram

\[
C(r; \theta) = \begin{cases} 
\theta_1 \left[ 1 - \frac{3r}{2\theta_2} + \frac{r^3}{2\theta_2^3} \right], & 0 \leq r \leq \theta_2, \\
0, & \text{otherwise},
\end{cases}
\]  

(3.1)

the range is well-defined, since the covariogram vanishes at the finite distance \( \theta_2 \). However, for covariograms that vanish only asymptotically the range is not well-defined. In contrast, the meaning of \( \varepsilon \)-remoteness is always clear. For example, for the isotropic exponential covariogram

\[
C(r; \theta) = \theta_1 \exp(-\theta_2 r),
\]  

(3.2)

two points in \( \mathbb{R}^d \) are \( \varepsilon \)-remote if the distance between them is greater than \((1/\theta_2) \log(\theta_1/\varepsilon)\).

This attempt to make the notion of remoteness precise is similar in spirit to Diamond and Armstrong's (1984) attempt to rigorously define "similarity" among variogram models. Their underlying purpose was to demonstrate the robustness of kriging estimators to misspecification of the variogram. Here, the objective is to approximate the covariance matrix \( V \) of an RFIM by a matrix whose structure is simpler
than that of \( V \), allowing the computational burden of ML and REML estimation to be reduced.

How does the concept of \( \varepsilon \)-remoteness help to relieve the computational burden of ML and REML estimation for RFLM's? If the observational sites \( R_1, \ldots, R_n \) can be grouped into \( K \) sets \( T_1, T_2, \ldots, T_K \) such that \( T_i \) and \( T_j \) are \( \varepsilon \)-remote for all \( i \neq j = 1, \ldots, K \) (note that \( K \) depends on \( \varepsilon \)), then we can arrange the observations \( y_1, \ldots, y_n \) in the observational vector \( y \) in such a way that the covariance matrix \( V \) differs "very little" from the block-diagonal matrix

\[
V_{\varepsilon} = \begin{bmatrix}
B_{11} & & \\
& B_{22} & \\
& & \ddots \\
& & & B_{kk}
\end{bmatrix}
\]

(The notion of "very little" is made precise following the introduction of Theorem 3.1.) Here, \( B_{jj} \) is a \( c_j \times c_j \) symmetric matrix, where \( c_j \) is the number of observational sites belonging to \( T_j \). The subscript \( \varepsilon \) is used to stress the dependence of the block diagonal matrix \( V_{\varepsilon} \) on \( \varepsilon \).

If \( \max\{c_j\} \) is small relative to \( n \), \( V_{\varepsilon} \) is much easier to invert than \( V \). The following theorem provides a useful bound on how much the inverse of \( V_{\varepsilon} \) differs from \( V^{-1} \) for small \( \varepsilon \). Define the matrix norm

\[
\|A\| = \max_{i,j} |a_{ij}|
\]

often called the row-sum norm.
Theorem 3.1

Let $\eta$ represent an arbitrary positive number. Suppose there exists $\varepsilon > 0$ and a positive integer $K$ such that $T_1, T_2, \ldots, T_k$ are pairwise $\varepsilon$-remote. Let $V$ be partitioned into $c_i \times c_j$ blocks $B_{ij}$ ($i=1, \ldots, K; j=1, \ldots, K$), where $c_i$ is the number of observational sites contained in $T_i$. Suppose further that the diagonal blocks $\{B_{jj}\}$ satisfy

$$\sum_{j=1}^{K} \sum_{u=1}^{c_j} \sum_{v=1}^{c_j} |B_{jj}^{-1}|_{uv} < \frac{\eta}{\varepsilon}. \quad (3.3)$$

Then,

$$\| I - \mathbb{V}_\varepsilon \| < \eta.$$

Proof:

We have

$$\mathbb{V}_\varepsilon^{-1} = \begin{bmatrix} I & B_{12}B_{22}^{-1} & \ldots & B_{1k}B_{kk}^{-1} \\ B_{21}B_{11}^{-1} & I & \vdots \\ \vdots & \ddots & \ddots \\ B_{k1}B_{11}^{-1} & \ldots & \ldots & I \end{bmatrix}.$$

Since the $(i,j)$th element of $B_{ij}^{-1}$ is $\sum_{u=1}^{c_j} (B_{ij})_{ju} (B_{jj}^{-1})_{uv}$, and since $T_i$ and $T_j$ are $\varepsilon$-remote, we have
Thus,

\[ \| I - V \varepsilon^{-1} \| = \max \{ \max \left( \sum_{i=1}^{c_i} \sum_{v=1}^{j \neq i} \sum_{u=1}^{k} \left| (B_{ij}B_{jj})_{uv} \right| \right) \} \]

\[ < \varepsilon \max \left[ \sum_{i=1}^{c_i} \sum_{j \neq i} \sum_{v=1}^{k} \sum_{u=1}^{k} \left| (B_{ij}^{-1})_{uv} \right| \right] \]

\[ < \varepsilon \sum_{j=1}^{k} \sum_{v=1}^{k} \sum_{u=1}^{k} \left| (B_{ij}^{-1})_{uv} \right| \]

The desired result follows by hypothesis. Q.E.D.

An important consequence of the bound in Theorem 3.1 is that, when \( \eta < 1 \), the elements of \( V^{-1} \) may be approximated to as high an accuracy as desired by using a recursive formula which requires no matrix inversion except that of \( V_\varepsilon \) (see, e.g., Fadeeva, 1959, pp. 99-102). The recursive formula is

\[ (V^{-1})^{(\ell)} = (V^{-1})^{(\ell-1)} + (V^{-1})^{(\ell-1)}[I-V(V^{-1})^{(\ell-1)}] \]

\[ (\ell=1, 2, \ldots), \]
where \( (V^{-1})^{(0)} = V^{-1}_\varepsilon \). A bound on the error is

\[
\| (V^{-1})^{(2)} - V^{-1}_\varepsilon \| < \| V^{-1}_\varepsilon \| \frac{\pi^2 \varepsilon}{1-\eta}.
\]

Thus, if \( \eta < 1 \), the accuracy of the elements of \( V^{-1} \) increases geometrically.

Since both sides of (3.3) are functionally dependent on \( \varepsilon \), it may be necessary to find \( \varepsilon \) by trial and error. If numerous trial values of \( \varepsilon \) are required before a suitable value is discovered, the whole purpose of this technique may be defeated because verification of (3.3) for a trial value of \( \varepsilon \) does require the inversion of the blocks \( B_{11}, \ldots, B_{kk} \) on the block diagonal of \( V \).

The method described here for \( \varepsilon \)-remote observational sites makes ML and REML estimation possible with fewer function evaluations and smaller order matrix inversions than in the general case. In fact, if \( V^{-1}_\varepsilon \) is deemed close enough to \( V^{-1} \) that it need not be improved upon using (3.4), then the method requires only \( O([\max\{c_j\}]^3) \) computations for matrix inversion.

Note that \( \mathcal{X}_y \) need not be weakly stationary nor isotropic for this technique to be applicable; however, these two properties, if they held even approximately, would be of some benefit in grouping the observational sites into sets \( T_1, \ldots, T_k \) that are pairwise \( \varepsilon \)-remote.
3.3. Configuration I

We now investigate the structure of $V$ and methods for exploiting that structure when $\mathcal{F}_Y$ is observed at subsets $R_1, \ldots, R_n$ of $\mathbb{R}$ whose spatial configuration is Configuration I. Configuration I and the notation and terminology associated with it were described in Section 1.4. The same notation and terminology will apply in this section.

Recall that, in the context of Configuration I, the RFLM is necessarily either a special case of RFLM-P or a special case of RFLM-R. Also recall that the covariance matrices associated with RFLM-P and RFLM-R are $V^P$ and $V^R$, respectively.

The covariance matrix $V$ can be a highly patterned matrix when the observational sites have Configuration I. One such patterned matrix is a Toeplitz matrix, defined as follows.

**Definition 3.2**

An $n \times n$ matrix $A = \{a_{ij}\}$ is said to be a Toeplitz matrix if

$$a_{i+s,j+s} = a_{ij}$$

for all possible $i$, $j$, and $s$.

Note that an $n \times n$ matrix $A$ is a symmetric Toeplitz matrix if and only if its elements $\{a_{ij}\}$ depend on $i$ and $j$ only through $|i-j|$.

Mardia and Marshall (1984) observed, for the case $d=1$, that $V^P$ is a symmetric Toeplitz matrix when $\mathcal{F}_Y$ is weakly stationary and the spatial configuration of observational sites is Configuration I. The following theorem establishes a similar result for $V^R$ and extends both results to arbitrary values of $d$. 
Theorem 3.2

Suppose that a single realization of a d-dimensional random field $X$ is observed at sites $R_1, \ldots, R_n$ having Configuration I. Suppose further that $X$ is a weakly stationary d-dimensional random field. Then, $V^R$ and $V^P$ are symmetric Toeplitz matrices.

Proof:

Let $e_1$ denote the dxl vector $(1, 0, \ldots, 0)$, and let $s$ be any integer. Because the observational sites in RFLM-R are regions $S_1, \ldots, S_n$, and because of the manner in which the observations are arranged in the observational vector $y$ for Configuration I (cf. Section 1.4), the elements $\{v^R_{ij}\}$ of $V^R$ are given by

$$v^R_{ij} = \int_{S_i} \int_{S_j} C(s,t) ds dt.$$

The superimposability of $S_1, \ldots, S_n$ and the weak stationarity of $X$ imply that the elements of $V^R$ satisfy

$$v_{i+s, j+s} = \int_{S_{i+s}} \int_{S_{j+s}} C(s,t) ds dt$$

$$= \int_{S_i} \int_{S_j} C(s + Hse_l, t + Hse_l) ds dt$$

$$= \int_{S_i} \int_{S_j} C(s,t) ds dt$$

$$= v^R_{ij}.$$
Similarly, the elements \( \{v_{ij}^P\} \) of \( V^P \) satisfy

\[
v_{i+s,j+s}^P = C(s_{i+s}, s_{j+s})
\]

\[
= C(s_i + sH_e, s_j + sH_e)
\]

\[
= C(s_i, s_j)
\]

\[
= v_{ij}^P.
\]

Now, \( v_{ij}^R = v_{ji}^R \) and \( v_{ij}^P = v_{ji}^P \), which imply that \( V^R \) and \( V^P \) are symmetric Toeplitz matrices.

Q.E.D.

The following corollary is an immediate consequence of Theorem 3.2.

**Corollary**

Under the conditions of Theorem 3.2, \( \partial V/\partial \theta_i \) and \( \partial^2 V/\partial \theta_i \partial \theta_j \) (\( i = 1, \ldots, m; j = 1, \ldots, m \)) are symmetric Toeplitz matrices.

Because an \( nxn \) symmetric Toeplitz matrix is comprised of at most \( n \) distinct elements, an obvious consequence of Theorem 3.2 and its corollary is that \( V \) and its first- and second-order partial derivatives with respect to the elements of \( \theta \) can be obtained with \( O(n) \) computations (as compared to \( O(n^2) \) in the general case). In fact, it is easily verified that the exact number of elements and derivatives of elements to be evaluated on each iteration when the method of scoring is applied to \( L^*_1 \) or \( L_1 \), when
Newton-Raphson is applied to $L_1^*$ or $L_1$, or when Newton-Raphson is applied to $L_2^*$ or $L_2$ does not exceed $(m+1)n$, $(m^2+m+1)n$, and $(m^2-m+1)n$, respectively.

It would seem that the special structure of a Toeplitz matrix might be used to advantage in obtaining its inverse (when the latter exists). This is indeed the case. Ray (1970) derived a Toeplitz matrix inversion algorithm that reduces computations by a factor of approximately $n^3/[2(n/2)^3] = 4$ from what they would be for an arbitrary nonsingular matrix. For large $n$, however, an algorithm given first by Trench (1964) and later elucidated by Zohar (1969) is more efficient. After introducing a definition and three lemmas, we will briefly describe the Trench-Zohar algorithm.

The cross-diagonal of an $n \times n$ matrix is the diagonal array of elements extending from the $(n,1)$ element of the matrix to the $(1,n)$ element.

**Definition 3.3**

An $n \times n$ matrix $A = \{a_{ij}\}$ is said to be persymmetric if $a_{n+1-j,n+1-i} = a_{ij}$ for all $i$ and $j$, i.e., elements symmetric with respect to the cross-diagonal are equal.

The following lemma is easily proved using the definition of a Toeplitz matrix and that of a persymmetric matrix.

**Lemma 3.1**

A Toeplitz matrix is persymmetric.

The converse of Lemma 3.1 is not true, however, as the following example indicates. Consider the matrices
Both $M_1$ and $M_2$ are persymmetric, but only $M_1$ is Toeplitz.

Define $E$ to be the $n \times n$ "exchange" matrix having ones along the cross-diagonal and zeros elsewhere, i.e.,

$$
E = \begin{bmatrix}
1 & 1 \\
& 1 \\
& & 1 \\
& & & \ddots \\
& & & & 1 \\
\end{bmatrix}.
$$

The matrix $E$ is called the $n \times n$ exchange matrix because premultiplying a matrix $A$ having $n$ rows by $E$ reverses the order of the rows of $A$, and postmultiplying a matrix $A$ having $n$ columns by $E$ reverses the order of the columns of $A$. Thus, if $A = \{a_{ij}\}$ is an $n \times n$ matrix, then the $(i,j)$th element of $EAE$ is $a_{n+1-i,n+1-j}$. Note also that $EE = I$, so that $E^{-1} = E$.

The following lemma provides an alternative way to characterize a persymmetric matrix using the exchange matrix.
Lemma 3.2

An nxn matrix $A = \{a_{ij}\}$ is persymmetric iff $EAE = A'$.

Proof:

Let $b_{ij}$ represent the $(i,j)$th element of $EAE$. Since $b_{ij} = a_{n+1-i,n+1-j}$, we have that $b_{ij} = a_{ji}$ iff $a_{ij} = a_{n+1-j,n+1-i}$, i.e.,

$EAE = A'$ iff $a_{ij} = a_{n+1-j,n+1-i}$ for all $i$ and $j$. Q.E.D.

The next lemma was proved by Zohar (1969). We give a proof that is slightly simpler than that given by Zohar.

Lemma 3.3

The inverse of an nxn nonsingular persymmetric matrix $A$ is persymmetric.

Proof:

It follows from Lemma 3.2 that $A^{-1} = (EA'E)^{-1} = E^{-1}(A')^{-1}E^{-1} = E(A^{-1})'E$, i.e., $A^{-1}$ is persymmetric. Q.E.D.

Lemma 3.3 may lead one to speculate that the inverse of a Toeplitz matrix is itself a Toeplitz matrix. This is not true except under special conditions (obtained by Greville, 1983) which cannot be satisfied by a symmetric matrix unless all of its off-diagonal elements are equal. Therefore, if the conditions of Theorem 3.2 are satisfied,
and if the covariogram is nonconstant over that part of its domain (excluding \( \{0\} \)) encompassed by \( R \) (as would be the case, e.g., if the covariogram was monotone decreasing), then \( V^{-1} \) is not Toeplitz. However, since \( V \) is Toeplitz, it is persymmetric according to Lemma 3.1; hence, the symmetry of \( V \) and Lemma 3.3 imply that \( V^{-1} \) is symmetric and persymmetric.

The symmetry and persymmetry of \( V^{-1} \) imply that it has fewer distinct elements than an arbitrary \( nxn \) matrix. The exact number of distinct elements of such a matrix is given by the following lemma.

**Lemma 3.4**

Let \( A = \{a_{ij}\} \) be an \( nxn \) matrix that is both symmetric and persymmetric. The maximum number of distinct elements of \( A \) is \( (n+1)^2/4 \) if \( n \) is odd, and \( n(n+2)/4 \) if \( n \) is even.

**Proof:**

By hypothesis, \( a_{ij} = a_{ji} \) and \( a_{ij} = a_{n+1-j,n+1-i} \). Thus, the distinct elements of \( A \) are contained in its left "wedge" of elements

\[
\begin{align*}
a_{11}, \\
a_{21}, a_{22}, \\
\vdots \\
a_{(n+1)/2,1}, \ldots, a_{(n+1)/2,(n+1)/2}
\end{align*}
\]
\(^{(n+3)/2,1'} \ldots, ^{(n+3)/2,(n-1)/2}
::
\(^{a_n-1,1'}, ^{a_{n-1,2'}, \text{ and}}
\^{a_{n1}}
\text{for } n \text{ odd, or}
\^{a_{11'}}
\^{a_{21'}, ^{a_{22'}}}
::
\^{a_{(n/2),1'}} \ldots, ^{a_{(n/2),(n/2')}}
\^{a_{(n/2)+1,1'}} \ldots, ^{a_{(n/2)+1,n/2}}
::
\^{a_{n-1,1'}, ^{a_{n-1,2'}, \text{ and}}
\^{a_{n1}}
\text{for } n \text{ even. Therefore, if } n \text{ is odd, } A \text{ has at most}
1 + 3 + \ldots + n = 2[1+2+\ldots+(n+1)/2] - [(n+1)/2]
= 2[(n+1)/2][(n+1)/2+1] - [(n+1)/2]
= (n+1)^2/4
\text{distinct elements, and if } n \text{ is even, } A \text{ has at most}
2 + 4 + \ldots + n = 2(1+2+\ldots+n/2)
= 2(n/2)[(n/2)+1]/2
= n(n+2)/4 \text{ distinct elements. Q.E.D.}
The Trench-Zohar algorithm for obtaining the inverse of an $n \times n$ non-
singular Toeplitz matrix $A_n$ whose principal minors are all nonzero is
based on partitioning $A_n$ as

$$A_n = \begin{bmatrix}
1 & e' \\
\vdots & \ddots \\
g & A_{n-1}
\end{bmatrix}$$

and then applying the well-known bordering method (see, e.g., Fadeeva, 1959, pp. 105-111). The crucial property of $A_n$ that makes the Trench-
Zohar algorithm so efficient is that the submatrix $A_{n-1}$ is also Toeplitz.
This fact leads to the so-called regenerative property of the inverse of
$A_n$: all of the elements of $A_n^{-1}$ can be generated from its first row and
column with $O(n^2)$ computations. Furthermore, the first row and column
of $A_n^{-1}$ can be obtained with $O(n^2)$ computations by a simple recursion
formula. For a clear and detailed derivation of the Trench-Zohar
algorithm, the reader is referred to Zohar (1969). The steps of the
algorithm, adapted to the symmetric Toeplitz covariance matrix $V$ of an
RFLM under the conditions of Theorem 3.2, are outlined below. The re-
quirement that all principal minors of $V$ are nonzero is satisfied since
$V$ is positive definite for all $\theta \in \Theta$.

Notation:

$$V_n = \alpha \begin{bmatrix}
1 & u_{n-1}' \\
u_{n-1} & V_{n-1}
\end{bmatrix}, \quad u_i = (\rho_1, \rho_2, \ldots, \rho_i) (i=1, \ldots, n-1),$$

$$V_n^{-1} = \frac{1}{\alpha \lambda_{n-1}} \begin{bmatrix}
1 & g_{n-1}' \\
g_{n-1} & G_n
\end{bmatrix}.$$
Initial values for recursion:

\[ \lambda_1 = 1 - \rho_1^2, \quad g_1 = -\rho_1. \]

Recursion formula for obtaining first column of \( V^{-1}_n \) \((1 \leq i < n-1)\):

\[ y_i = -(\rho_{i+1} + \lambda_i g_i), \quad \text{where } \hat{g}_i = E g_i, \]

\[ \hat{g}_{i+1} = \begin{bmatrix} \gamma_i / \lambda_i \\ \hat{g}_i + (\gamma_i / \lambda_i) g_i \end{bmatrix}, \]

\[ \lambda_{i+1} = \lambda_i - \frac{\gamma_i^2}{\lambda_i} \]

Formation of \( V^{-1}_n \):

\[ (V^{-1}_n)^{-1} = 1 / \alpha \lambda_{n-1}, \]

\[ (V^{-1}_n)_{j+1,1} = (1/\alpha \lambda_{n-1}) (g_{n-1})_{j+1} \quad \text{for } 1 \leq j \leq n-1, \]

\[ (V^{-1}_n)_{j+1,k+1} = (V^{-1}_n)_{j,k} + (1/\alpha \lambda_{n-1}) (g_{n-1} g_{n-1} - \hat{g}_{n-1} \hat{g}_{n-1})_{jk} \quad \text{(3.6)} \]

\[ \text{for } 1 \leq k \leq \min\{j,n-j-1\}, 1 \leq j \leq n-2. \]

The remaining elements of \( V^{-1}_n \) can be filled in by virtue of its symmetry and persymmetry.
A careful examination of this algorithm reveals that $2i+2$ multiplications and $2i+1$ additions are required for the $i$th recursion; hence, the computations involved in obtaining the first column of $V^{-1}_n$ are $O(n^2)$. Furthermore, it is obvious from inspection that (3.6) requires $O(n^2)$ computations. Thus, under the conditions of Theorem 3.2, $V$ can be inverted with $O(n^2)$ computations.

There are certain special symmetric Toeplitz matrices whose inverses can be obtained with even fewer computations than are required by the Trench-Zohar algorithm. Under the conditions of Theorem 3.2, the covariance matrices corresponding to particular covariograms have one of these special forms. Consider, in particular, the symmetric Toeplitz matrix

$$
\begin{pmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\
1 & \rho & & & \\
\sigma^2 & 1 & \ddots & & \\
& \ddots & \ddots & \ddots & \\
\text{symm} & & \ddots & \ddots & 1 \\
& & & \rho & \\
& & & 1 & \\
\end{pmatrix}
$$

(3.7)

It is well-known that the covariance matrix of $n$ realizations of a weakly stationary first-order autoregressive time series is of this form. Under the conditions of Theorem 3.2, $V^P$ (but not $V^R$) is also of this form when the covariogram of $F_Y$ is the isotropic exponential covariogram (3.2), as can be seen by putting $\theta_1 = \sigma^2$ and $e^{-\theta_2 H} = \rho$. 
The inverse of (3.7) exists when $\sigma^2 > 0$ and $|\rho| < 1$, and has the particularly simple form

\[
\begin{bmatrix}
1 -\rho & & & \\
(1+\rho^2) -\rho & 1+\rho^2 & & \\
 & \ddots & \ddots & \\
& & & 1+\rho^2 -\rho \\
\end{bmatrix}
\]  

(3.8)

Note that (3.8) has only four distinct elements; thus, the matrix given by (3.7) can be inverted with $O(1)$ computations.

Another special type of symmetric Toeplitz matrix is the symmetric Toeplitz band matrix of order $2k+1$.

**Definition 3.4**

An $n \times n$ matrix $A = \{a_{ij}\}$ is said to be a band matrix of order $2k+1$ if $a_{ij} = 0$ whenever $|i-j| > k$.

A band matrix of order 3 is commonly called a tridiagonal matrix.

An explicit formula for the inverse of an $n \times n$ positive definite tridiagonal symmetric Toeplitz matrix
is (assuming $a_1 \neq 0$)

$$a_{ij} = \begin{cases} 
\frac{(1-b^{-2n-2j+2})(b^{j+i+1}_0 - b^{j+1}_0)}{(a_1/a_0)(1-b^2)(1-b^{2n+2})} & , i \leq j , \\
\alpha_{ii} & , i > j , 
\end{cases} \quad (3.9)$$

where $b = \frac{1}{2}(a_1/a_0)(\sqrt{1-4(a_1/a_0)^2} - 1)$ and $a_{ij}^*$ is the $(i,j)$th element of $A^{-1}$ (Graybill, 1983, p. 286). This formula does involve $O(n^2)$ computations but the number of computations is a fraction of those required by the Trench–Zohar algorithm.

For Toeplitz band matrices of order greater than 3, nonrecursive inversion formulae, if they exist, do not yet appear to be available. However, Trench (1974) showed that, in the special case of a Toeplitz band matrix, his algorithm for Toeplitz matrix inversion can be simplified. The simplified algorithm still requires $O(n^2)$ computations.
for the "regenerative" part of the algorithm but only requires $O(n)$ computations for the "recursive" part.

Covariograms that give rise to symmetric Toeplitz band matrices under the conditions of Theorem 3.2 must have a finite range, as do the isotropic spherical covariogram (3.1) and the so-called linear covariogram

$$ C(r; \theta) = \begin{cases} \theta_1 (1 - \frac{r}{\theta_2}), & r \leq \theta_2, \\ 0, & \text{otherwise}, \end{cases} $$

where $\theta_1 > 0$ and $\theta_2 > 0$. In these cases, the order of the band matrix $V$ depends on the magnitude of the "correlation decay" parameter $\theta_2$ relative to the span $H$ of Configuration I. For example, if $H$ exceeds $\theta_2$, then $V^P$ is diagonal and $V^R$ is tridiagonal.

Many covariograms whose range does not exist may yield a covariance matrix $V = \{v_{ij}\}$ that is "nearly" a band matrix in the sense that $|v_{ij}| < \varepsilon$ whenever $|i-j| > k$ for some "small" positive $\varepsilon$ and some positive integer $k$. An approximate ML or REML approach similar to that proposed in Section 3.2 could be developed in this setting, though it would have little or no merit unless the covariogram attenuated so rapidly that a diagonal or tridiagonal symmetric Toeplitz matrix was an excellent approximation to $V$. 
3.4. Configuration II

In this section, we consider the structure of $V$ and methods for exploiting that structure when $F_Y$ is observed at subsets $R_1, \ldots, R_n$ of $\mathbb{R}$ whose spatial configuration is Configuration II. The notation and terminology associated with Configuration II were described in Section 1.4 and apply throughout this section. An RFLM in the context of Configuration II is necessarily either a special case of RFLM-P or a special case of RFLM-R. Recall that the observational sites are labelled as $s_{11}^*, s_{12}^*, \ldots, s_{RC}^*$ under RFLM-P, and as $s_{11}^a, s_{12}^a, \ldots, s_{RC}^a$ under RFLM-R, according to the scheme described in Section 1.4.

Certain highly patterned block matrices are important in the context of Configuration II. Two of these matrices are a block Toeplitz matrix and a block persymmetric matrix. The definitions of these matrices, given below, are natural extensions of Definitions 3.2 and 3.3 to block matrices.

Definition 3.5

An RCxRC block matrix $A$ comprised of $C \times C$ blocks $A_{ij}$ ($i = 1, \ldots, R; j = 1, \ldots, R$) is said to be block Toeplitz if $A_{i+s,j+s} = A_{ij}$ for all possible $i, j,$ and $s$.

Definition 3.6

An RCxRC block matrix $A$ comprised of $C \times C$ blocks $A_{ij}$ ($i = 1, \ldots, R; j = 1, \ldots, R$) is said to be block persymmetric if $A_{R+1-j,R+1-i} = A_{ij}$ for all possible $i$ and $j$. 
The following lemma is easily proved using the definitions of a block Toeplitz matrix and a block persymmetric matrix.

**Lemma 3.5**

A block Toeplitz matrix is block persymmetric.

The matrix $M_2$ — refer to equation (3.5) — is an example of a block Toeplitz matrix for which $R=3$ and $C=2$.

Subsequently, we shall refer to a symmetric matrix that is block Toeplitz or block persymmetric as a symmetric block Toeplitz matrix or a symmetric block persymmetric matrix.

**Theorem 3.3**

Suppose that a single realization of a $d$-dimensional random field $\mathcal{F}_Y$ is observed at sites $R_1, \ldots, R_n$ having Configuration II.

(i) If $\mathcal{F}_Y$ is weakly stationary, then the covariance matrix $V$ of both RFLM-P and RFLM-R has the symmetric block Toeplitz matrix representation

$$V = \begin{pmatrix}
V_0 & V_1 & V_2 & \cdots & V_{R-1} \\
V_1 & V_0 & & & \\
V_2 & & & & \\
& & & \ddots & \\
V_{R-1} & \cdots & V_1 & & V_0
\end{pmatrix}, \quad (3.10)$$

where $V_i$ ($i = 0, 1, \ldots, R-1$) is a $C \times C$ Toeplitz matrix.
(ii) If \( \mathcal{X}_Y \) is isotropic in addition to being weakly stationary, then each of the blocks \( V_0, V_1, \ldots, V_{R-1} \) is a \( C \times C \) symmetric Toeplitz matrix, in which case \( V \) has the representation

\[
V = \begin{bmatrix}
V_0 & V_1 & V_2 & \cdots & V_{R-1} \\
V_1 & V_0 & \cdots & \cdots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
V_{R-1} & \cdots & V_1 & V_0 
\end{bmatrix}.
\]

\[\text{(3.11)}\]

Proof of (i):

Without loss of generality, let us superimpose a \( d \)-dimensional rectangular coordinate system onto the layout in such a manner that its first two axes are aligned with the rows and columns of the layout. Let \( C(\cdot, \cdot) \) denote the covariogram of \( \mathcal{X}_Y \). Because of the manner in which the observations are arranged in \( y \) for Configuration II, the covariance matrix of the observations has the form

\[
V = \begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1R} \\
A_{21} & A_{22} & \cdots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
A_{R1} & \cdots & A_{R2} & A_{RR}
\end{bmatrix},
\]

where
\begin{align*}
(A_{ik})_{j\ell} &= \int_{S_{ij}} \int_{S_{k\ell}} C(s,t) ds dt \\
\text{under RFLM-R, and} \\
(A_{ik})_{j\ell} &= C(s_{ij}, s_{k\ell}) \\
\text{under RFLM-P. Let } u \text{ and } v \text{ be given integers, and let } u = (u_{H_1}, v_{H_2}, 0, \ldots, 0) \text{ be a } d \times 1 \text{ vector. Under RFLM-R, the isometry of } S_{11}, S_{12}, \ldots, S_{RC} \text{ and the weak stationarity of } \mathcal{F}_Y \text{ imply that} \\
&= \int_{S_{ij}} \int_{S_{k\ell}} C(s_{ij}, s_{k\ell}) ds dt \\
&= \int_{S_{ij}} \int_{S_{k\ell}} C(s_{ij}, s_{k\ell}) ds dt. \quad (3.12)
\end{align*}

Similarly, under RFLM-P, the weak stationarity of \( \mathcal{F}_Y \) implies that

\begin{align*}
C(s_{ij}, s_{k\ell}) &= C(s_{ij}, s_{j\ell}) \\
&= C(s_{ij}, s_{j\ell}). \quad (3.13)
\end{align*}

From (3.12) and (3.13), we have, under both RFLM-R and RFLM-P, that

\begin{align*}
(A_{i+u, k+u})_{j\ell} &= (A_{ik})_{j\ell} \text{ for all possible } i, j, k, \ell, \text{ and } u, \text{ implying}
\end{align*}
that $V$ is block Toeplitz in either case. Further, (3.12) and (3.13) imply that $(A_{ik})_{j+v, \ell+v} = (A_{ik})_{j, \ell}$ for all possible $i, j, k, \ell, v$, and hence that $A_{ik}$ ($i = 1, \ldots, R; k = 1, \ldots, R$) is Toeplitz. Thus, under both RFLM-R and RFLM-P, $V$ is a symmetric block Toeplitz matrix whose blocks are Toeplitz, so that $V$ has the representation (3.10).

Q.E.D.

Proof of (ii):

If $Z_Y$ is isotropic as well as weakly stationary, then, using the isometry of $S_{11}, S_{12}, \ldots, S_{RC}$, we have, under RFLM-R,

$$
(A_{ik})_{ij} = \int_{S_{i\ell}} \int_{S_{k\ell}} C(s, t) ds dt
$$

$$
= \int_{S_{i\ell}} \int_{S_{kj}} C((s-t)'(s-t))^{1/2} ds dt
$$

$$
= \int_{S_{ij}} \int_{S_{k\ell}} C((s-t)'(s-t))^{1/2} ds dt
$$

Thus,

$$(A_{ik})_{ij} = \int_{S_{ij}} \int_{S_{k\ell}} C(s, t) ds dt
$$

$$
= (A_{ik})_{j\ell}
$$

for all possible $i, j, k, \ell$. Similarly, under RFLM-P,
\[(A_{ik})_{lj} = C(s_{ij}, s_{kj})\]
\[= C([(s_{ij} - s_{kj})'(s_{ij} - s_{kj})]^{1/2})\]
\[= C([(s_{ij} - s_{kl})'(s_{ij} - s_{kl})])\]
\[= C(s_{ij}, s_{kl})\]
\[= (A_{ik})_{lj}.\]

Hence, if \(\mathcal{F}_Y\) is isotropic, then the blocks of \(V\) are symmetric, and \(V\) has the representation (3.11) under both RFLM-R and RFLM-P. Q.E.D.

Subsequently, we define \(V_0, V_1, \ldots, V_{R-1}\) to be the \(C\times C\) blocks of the first row of blocks of the covariance matrix \(V\) when the spatial configuration of sites is Configuration II and the RFLM satisfies the conditions of Theorem 3.3(i).

The following corollary is a simple consequence of Theorem 3.3.

**Corollary**

Under the conditions of Theorem 3.3(i), \(\partial V/\partial \theta_i\) and \(\partial^2 V/\partial \theta_i \partial \theta_j\) \((i = 1, \ldots, m; j = 1, \ldots, m)\) are symmetric block Toeplitz matrices with \(C\times C\) Toeplitz blocks. Under the conditions of Theorem 3.3(ii), the blocks of these matrices are symmetric.

For two-dimensional RFLMs, it can be shown that the results of part (ii) of Theorem 3.3 holds under a somewhat weaker condition than
isotropy. This condition, which is called reflective anisotropy, is defined as follows.

**Definition 3.7**

A weakly stationary two-dimensional random field whose covariogram is \( \hat{C}(\cdot) \) is said to be reflectively anisotropic if \( \hat{C}(x,y) = \hat{C}(x,-y) \) for all \((x,y) \in \mathbb{R}^2\).

That reflective anisotropy is sufficient for the blocks of \( V \) to be symmetric is now illustrated for the case of RFIM-R. If \( F_Y \) is reflectively anisotropic, then

\[
\int \int_{S_{ijkl}} \hat{C}(s_1-t_1, s_2-t_2) ds_1 ds_2 dt_1 dt_2 \quad \hat{C}(s_1-t_1, t_2-s_2) ds_1 ds_2 dt_1 dt_2
\]

\[
\int \int_{S_{ijkl}} \hat{C}(s_1-t_1, s_2-t_2) ds_1 ds_2 dt_1 dt_2 .
\]

Clearly, whether or not a two-dimensional random field satisfies Definition 3.7 can depend on how the coordinate axes are oriented; consequently it may be of little value in practice to know (or hypothesize) that \( F_Y \) is reflectively anisotropic unless some prior information on
the characteristics of $\mathcal{F}$ is available. In the absence of such information, isotropy must, in practice, be assumed in order for $V$ to have the representation (3.11). It will be seen in the sequel that the symmetry of the blocks helps to ease the burden of inverting $V$, so isotropy is a useful property from a computational standpoint.

Theorem 3.3 and its corollary have direct implications regarding the number of function evaluations required at each iteration of a gradient algorithm for obtaining ML or REML parameter estimates. Clearly, an RCxRC symmetric block Toeplitz matrix comprised of CxC Toeplitz blocks is completely determined by the elements in its first row and the elements in the first column of each block in its first row of blocks. Thus, the number of distinct elements in such a matrix does not exceed $RC+(R-1)(C-1) = 2RC-C-R+1$ as compared to $RC(RC+1)/2$ for an arbitrary RCxRC symmetric matrix. If the CxC Toeplitz blocks are symmetric as well, the RCxRC matrix is completely determined by its first row of elements; consequently, the number of distinct elements does not exceed RC. For large R and C, the symmetry of the blocks reduces the number of function evaluations to about half of what they would be when the blocks are not symmetric. Thus, under the conditions of Theorem 3.3(i) or (ii), the number of function evaluations needed at each iteration of the maximization procedure is of $O(RC)$ as compared to $O(R^2C^2)$ in the general case.

Lemma 3.6

An RCxRC block persymmetric matrix $A$ with CxC persymmetric blocks is persymmetric.
Proof:

Let \( A_{ij} : i = 1, \ldots, R; j = 1, \ldots, R \) denote the \( C \times C \) blocks of \( A \), and let \( \{ a_{ij} : i = 1, \ldots, RC; j = 1, \ldots, RC \} \) denote the elements of \( A \).

Note that \( (A_{ij})^k \ell = a^{(i-1)C+k,(j-1)C+\ell} \) for \( i = 1, \ldots, R, j = 1, \ldots, R, k = 0, 1, \ldots, C-1, \) and \( \ell = 0, 1, \ldots, C-1 \). By the block persymmetry of \( A \) and the persymmetry of each of its blocks,

\[
(A_{ij})^k \ell = (A_{R+1-j,R+1-i})^k \ell \\
= (A_{R+1-j,R+1-i})^{C+1-\ell,C+1-k} 
\]

Thus,

\[
a^{(i-1)C+k,(j-1)C+\ell} = a^{(R-j)C+C+1-\ell,(R-i)C+C+1-k} \\
= a^{RC+1-(j-1)C-\ell,RC+1-(i-1)C-k}
\]

for all possible, \( i, j, k, \) and \( \ell \). Q.E.D.

Corollary

An \( RC \times RC \) block Toeplitz matrix \( A \) with \( C \times C \) Toeplitz blocks is persymmetric.

Proof:

The result follows immediately from Lemmas 3.1 and 3.5. Q.E.D.

The corollary to Lemma 3.6 implies, under the conditions of Theorem 3.3(i), that the covariance matrix \( V \) is persymmetric. However,
V may not be Toeplitz. When V is not Toeplitz, the (RC-1)x(RC-1) sub-
matrix formed by deleting the first row and first column of V may not be persymmetric; this contrasts with the result when the configuration of observational sites is Configuration I. Consequently, the Trench–Zohar algorithm is not applicable. However, under the conditions of Theorem 3.3(i), the C(R-1)xC(R-1) submatrix formed by deleting the first row of blocks and first column of blocks of V is block persymmetric. Akaike (1973) used this property to good advantage in extending the Trench–Zohar algorithm to the inversion of a block Toeplitz matrix. Modified for a symmetric block Toeplitz matrix, Akaike's algorithm is useful for inverting the covariance matrix of an RFLM satisfying the conditions of Theorem 3.3(i). The modified algorithm, as applied to V, is now described.

Certain operations on block matrices must be defined. If A is an aCxbC matrix, then \( \hat{A} \) is defined as the aCxbC matrix formed by exchanging the ith row of CxC blocks with the \((a+1-i)\)th row of CxC blocks, \( A^* \) is defined as the aCxbC matrix formed by transposing the elements within each CxC block of A, and \( \tilde{A} \) is defined as the bCxaC matrix formed by replacing the \((i,j)\)th CxC block by the \((j,i)\)th CxC block. Also, let \([A]_{ij}\) denote the \((i,j)\)th CxC block of A.

Let \( G_i \) \((i = 1, \ldots, R)\) denote the lower right \(i\)C\(i\)C submatrix of V. The symmetric block Toeplitz structure that V has under the conditions of Theorem 3.3(i) implies that \( G_{i+1} \) has the representation
\[ G_{i+1} = \begin{pmatrix} V_0 & \tilde{Z}_i \\ \tilde{Z}_i^* & G_i \end{pmatrix}, \]

where \( \tilde{Z}_i = (V_1, V_2, \ldots, V_i) \). It follows that

\[ \tilde{G}_{i+1} = \begin{pmatrix} V_0 & \tilde{Z}_i^* \\ \tilde{Z}_i & \tilde{G}_i \end{pmatrix} \]

and \( V = G_R \). Since \( G_{i+1} \) and \( \tilde{G}_{i+1} \) are symmetric for all \( i \), so are their inverses, and we can adopt the representations

\[ \tilde{G}_{i+1}^{-1} = \begin{pmatrix} Q_i & \tilde{P}_i \\ \tilde{P}_i^* & M_i \end{pmatrix}, \quad (G_{i+1})^{-1} = \begin{pmatrix} S_i & \tilde{T}_i \\ \tilde{T}_i^* & N_i \end{pmatrix}, \]

where \( Q_i \) and \( S_i \) are CxC matrices. Like the Trench-Zohar algorithm, the modified Akaike algorithm can be divided into a recursive part and a regenerative part because all of the blocks of \( V^{-1} \) can be generated from the first column of blocks.

Let \( U_i = S_i^{-1} T_i \), \( W_i = Q_i^{-1} P_i \), and let \( I \) represent the CxC identity matrix. The recursive part of the algorithm obtains \( U_{i+1} \), \( W_{i+1} \), \( Q_i^{-1} \), and \( S_i^{-1} \) from \( U_i \), \( W_i \), \( Q_i^{-1} \), and \( S_i^{-1} \) \((0 \leq i \leq R-2)\) according to the following equations:

\[ U_{i+1} = (U_i, 0) - (U_i \tilde{Z}_i + V_{i+1}) Q_i (W_i, I), \quad (3.14a) \]
The initial values for the recursive equations are $q^{-1}_0 = v_0$, $s^{-1}_0 = v_0$, $u_1 = -v_0v_0^{-1}$, and $w_1 = -v_0v_0^{-1}$.

The final iteration of the recursive equations (3.14a)-(3.14d) essentially produces the first column of CxR blocks of $V$, since $q_{R-1} = [V^{-1}]_{1,1}$ and $[P^*_{R-1}]_{j,1} = [V^{-1}]_{j+1,1}$ ($1 \leq j \leq R-1$). The blocks $[V^{-1}]_{j+1,k+1}$ ($1 \leq k \leq j; 1 \leq j \leq R-1$) are then obtained according to the regenerative equation

$$[V^{-1}]_{j+1,k+1} = [V^{-1}]_{j,k} + [P^*_{R-1}]_{j,R-1} [R^{-1} - R^{-1} R^{-1} R^{-1}]_{j,k}. \quad (3.15)$$

The remaining blocks of $V^{-1}$ can be filled in immediately because $V^{-1}$ is symmetric.

The $i$th iteration of the recursive equations (3.14a)-(3.14d) requires the inversion of a CxC matrix, four additions of CxC matrices and two additions of Cx(i+1)C matrices, and the pre- and post-multiplication of CxC matrices by CxC and Cx(i+1)C matrices, respectively. Consequently, computations in the recursive part of the algorithm are of

$$O(C^3) + O(C^2) + O(R^2 C^2) + O(C^3) + O(R^2 C^3) = O(R^2 C^3).$$

The regenerative equations (3.15) require the addition of $R(R-1)$ CxR
matrices and the pre- and post-multiplication of two CxC matrices by (R-1)CxC and Cx(R-1)C matrices, respectively, so the computations required in this part of the algorithm are \(O(R^2C^2) + O(R^2C^3) = O(R^2C^3)\).

Therefore, the covariance matrix \(V\) of an RFLM satisfying the conditions of Theorem 3.3(i) can be inverted with \(O(R^2C^3)\) computations, compared to \(O(R^3C^3)\) in general. It becomes apparent that if the number of rows in the layout is not equal to the number of columns, then a judicious identification of rows and columns is important; specifically, rows and columns should be identified so that \(R \geq C\).

The modified Akaike algorithm given by (3.14a)-(3.14d) and (3.15) applies to any symmetric block Toeplitz matrix; there are no limitations on the structure within blocks. By Theorem 3.3, however, the covariance matrix \(V\) has blocks which are Toeplitz and, if \(\mathbb{F}_Y\) is isotropic, symmetric. One may suspect that Akaike's algorithm can be modified to take advantage of this additional structure. This suspicion is justified: modifications can be made which reduce the computations in (3.14a)-(3.14d) and (3.15) by a factor of approximately 4. To show this, we must introduce a series of definitions and matrix lemmas. Some of the lemmas are quite trivial but are included for the sake of completeness.

**Definition 3.8**

An \(nxn\) matrix \(A = \{a_{ij}\}\) is said to be centrosymmetric if \(a_{n+1-i, n+1-j} = a_{ij}\) (\(i = 1, \ldots, n; j = 1, \ldots, n\)), i.e., elements symmetric with respect to the radial center of \(A\) are equal.
**Definition 3.9**

An $R \times R$ block matrix $A$ comprised of $C \times C$ blocks $A_{ij}$ ($i = 1, \ldots, R; j = 1, \ldots, R$) is said to be block centrosymmetric if $A_{R+1-i,R+1-j} = A_{ij}$ for all $i$ and $j$, i.e., blocks symmetric with respect to the radial center of $A$ are equal.

**Definition 3.10**

An $R \times R$ block matrix comprised of $C \times C$ blocks is said to be a matrix of order $(R,C)$.

Let $\mathcal{F}(R,C)$ denote the set of all matrices of order $(R,C)$, let $\mathcal{G}(R,C)$ denote the set of all block centrosymmetric matrices of order $(R,C)$, and let $\mathcal{H}(R,C)$ denote the set of all symmetric block persymmetric matrices of order $(R,C)$ with symmetric blocks.

**Lemma 3.6**

$\mathcal{H}(R,C) \subseteq \mathcal{G}(R,C)$.

**Proof:**

Let $A \in \mathcal{H}(R,C)$, and let $\{A_{ij}: i = 1, \ldots, R; j = 1, \ldots, R\}$ denote the $C \times C$ blocks of $A$. By the block persymmetry of $A$, together with the symmetry of $A$ and each of its component blocks, we have, for all $i$ and $j$, that

$$A_{ij} = A_{R+1-j,R+1-i} = (A_{R+1-i,R+1-j})' = A_{R+1-i,R+1-j}.$$  

Q.E.D.
Corollary

Let A be a symmetric block Toeplitz matrix of order (R,C) with symmetric Toeplitz blocks. Then A is block centrosymmetric.

Proof:

The result follows immediately from Lemmas 3.5 and 3.6. Q.E.D.

The following two lemmas are generalizations to block centrosymmetric matrices of results given by Cord and Sylvester (1962) for centrosymmetric matrices.

Lemma 3.7

Let $A \in \mathcal{G}(R,C)$ and $B \in \mathcal{G}(R,C)$, and define $D = AB$. Then $D \in \mathcal{G}(R,C)$.

Proof:

Let $\{A_{i,j}\}$, $\{B_{i,j}\}$, and $\{D_{i,j}\}$ ($i = 1, \ldots, R; j = 1, \ldots, R$) denote the $C \times C$ blocks of $A$, $B$, and $D$, respectively. Then, for any $i$ and $j$,

\[
D_{i,j} = \sum_{k=1}^{R} A_{i,k} B_{k,j}
\]

\[
= \sum_{k=1}^{R} A_{i,R+1-k,R+1-k} B_{R+1-k,R+1-j}
\]

\[
= \sum_{k=1}^{R} A_{R+1-i,R+1-j,k} B_{k,R+1-j}
\]

\[
= D_{R+1-i,R+1-j}^* \quad \text{Q.E.D.}
\]
Lemma 3.8

Let $A \in \mathcal{G}(R,C)$ and $B \in \mathcal{G}(R,C)$. Define $D = A + B$. Then $D \in \mathcal{G}(R,C)$.

Proof:

Let $\{A_{ij}\}$, $\{B_{ij}\}$, and $\{D_{ij}\}$ ($i = 1, \ldots, R; j = 1, \ldots, R$) denote the $C \times C$ blocks of $A$, $B$, and $D$, respectively. Then, for any $i$ and $j$,

$$D_{ij} = A_{ij} + B_{ij}$$

$$= A_{R+1-i, R+1-j} + B_{R+1-i, R+1-j}$$

$$= [A + B]_{R+1-i, R+1-j}$$

$$= D_{R+1-i, R+1-j}.$$  
Q.E.D.

Let $F$ denote the "block exchange" matrix of order $(R,C)$ having $C \times C$ identity matrices along the cross block diagonal and zeros elsewhere, i.e.,

$$F = \begin{bmatrix} 0 & \cdots & I \\ & \ddots & \\ I & & 0 \end{bmatrix}.$$  

For $A \in \mathcal{F}(R,C)$, premultiplying by $F$ interchanges the $i$th row of blocks in $A$ with the $(n+1-i)$th row of blocks, and postmultiplying by $F$ interchanges the $j$th column of blocks in $A$ with the $(n+1-j)$th column of blocks. Note that $FF = I$, so that $F^{-1} = F$.  

The following result is easily proved by an argument similar to that used to prove Lemma 3.2.

**Lemma 3.9**

A ∈ Ω(R,C) if and only if FAF = A.

Subsequently, let Ω₁(R,C) denote the set of all nonsingular block centrosymmetric matrices of order (R,C).

**Lemma 3.10**

If A ∈ Ω₁(R,C), then A⁻¹ ∈ Ω₁(R,C).

**Proof:**

From Lemma 3.9, we have that

\[ A^{-1} = (FAF)^{-1} = F^{-1}A^{-1}F^{-1} = FA^{-1}F, \]

implying that \( A^{-1} \in \Omega_1(R,C) \). Q.E.D.

**Corollary**

Let A be a nonsingular symmetric block Toeplitz matrix of order (R,C) with symmetric Toeplitz blocks. Then \( A^{-1} \) is block centrosymmetric.

**Proof:**

The result follows immediately by Lemma 3.10 and the corollary to Lemma 3.6. Q.E.D.
It is important to note that results analogous to Lemmas 3.7 and 3.10 do not hold, in general, for block symmetric or block persymmetric matrices. Thus, if $A$ is block symmetric (persymmetric), it is not necessarily the case that $A^{-1}$ is block symmetric (persymmetric).

Similarly, if $A$ and $B$ are block symmetric (persymmetric), it is not necessarily the case that $AB$ is block symmetric (persymmetric).

The following lemma is useful for reducing the number of computations required by the recursive part of the modified Akaike algorithm for inverting a symmetric block Toeplitz matrix when that matrix belongs to $\mathcal{C}_1(R,C)$.

**Lemma 3.11**

$\mathcal{C}_1(R,C)$ is a group with respect to multiplication.

**Proof:**

The result follows directly from Lemmas 3.7 and 3.10. Q.E.D.

Lemmas 3.7 through 3.11 do not place any restrictions on the within-block structure of the block centrosymmetric matrix. The following lemma is applicable to a subclass of block centrosymmetric matrices that have a particular within-block structure. Let $\mathcal{Z}(R,C)$ denote the set of all nonsingular block centrosymmetric matrices of order $(R,C)$ that are (elementwise) symmetric and persymmetric [$\mathcal{Z}(R,C) \subseteq \mathcal{C}_1(R,C)$].
Lemma 3.12

Let $A \in \mathcal{J}(R,C)$, and let $B_{ij}$ ($i = 1, \ldots, R; j = 1, \ldots, R$) denote the $C \times C$ blocks of $A^{-1}$. Then, $B_{ij}$ is centrosymmetric for all $i$ and $j$.

Proof:

The elementwise symmetry of $A^{-1}$, together with Lemma 3.10, implies that, for any $i$ and $j$,

$$
R+1-j, R+1-i = R+1-j, R+1-i
$$

Now, $B_{ij}$ and $B_{R+1-j, R+1-i}$ are blocks of $A^{-1}$ that are symmetrically located with respect to the cross block diagonal of $A^{-1}$. Hence, the persymmetry of $A^{-1}$ (which is guaranteed by Lemma 3.3) implies that

$$
(B_{ij})_{k\ell} = (B_{R+1-j, R+1-i})_{C+1-\ell, C+1-k}
$$

$$
= (B_{ij})_{C+1-\ell, C+1-k}
$$

$$
= (B_{ij})_{C+1-k, C+1-\ell}
$$

i.e., $B_{ij}$ is centrosymmetric. Q.E.D.

Corollary

Each $C \times C$ block of the inverse of a nonsingular symmetric block Toeplitz matrix of order $(R,C)$ with symmetric Toeplitz blocks is centrosymmetric.
Proof:

The result follows immediately because the class of nonsingular symmetric Toeplitz matrices of order \((R,C)\) with symmetric Toeplitz blocks is a subclass of \(3(R,C)\) by the corollaries to Lemmas 3.6 and 3.7.

Now, we return to Akaike's algorithm to consider how the preceding lemmas help to reduce the number of computations involved in inverting the covariance matrix \(V\) under the conditions of part (ii) of Theorem 3.3. The initial values for the recursive equations (3.14a)-(3.14d) are

\[
Q_0^{-1} = S_0^{-1} = V_0, \quad U_1 = -V_1V_0^{-1}, \quad \text{and} \quad W_1 = -V_1V_0^{-1}.
\]

By Theorem 3.3, \(V_0\) and \(V_1\) are symmetric Toeplitz matrices. Thus, the corollary to Lemma 3.6 implies that \(V_0\) and \(V_1\) are centrosymmetric, i.e., \(V_0\) and \(V_1\) belong to \(\mathcal{C}_1(1,C)\). (\(V_0\) and \(V_1\) are nonsingular since \(\theta \in \Theta\).) Therefore, the recursive process begins with matrices in \(\mathcal{C}_1(1,C)\) and, by Lemmas 3.8 and 3.11, at each iteration the matrices on the left-hand side of (3.14a)-(3.14d) remain in \(\mathcal{C}_1(1,C)\). This result would be of no particular benefit were it not possible to invert centrosymmetric matrices with less computation than arbitrary matrices. Good (1970) derived a formula for inverting nonsingular centrosymmetric matrices that requires roughly one-fourth as many computations as it takes to invert an arbitrary nonsingular matrix.

For a matrix

\[
A = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\]
belonging to $\mathcal{G}_1(1,C)$, where $C$ is even, and where $A_{11}$, $A_{12}$, $A_{21}$, $A_{22}$ are all $(C/2)\times(C/2)$ matrices, Good showed that $B \equiv A^{-1}$ can be written in the form

$$B = \begin{pmatrix} B_1 & EB_2 \\ B_2E & EB_1E \end{pmatrix},$$

where

$$B_1 = \frac{1}{2}[(A_{11} + A_{12}E)^{-1} + (A_{11} - A_{12}E)^{-1}],$$

$$B_2 = \frac{1}{2}[(A_{11} + A_{12}E)^{-1} - (A_{11} - A_{12}E)^{-1}],$$

and $E$ is the $C\times C$ exchange matrix. Good obtained a similar formula for the case where $C$ is odd. Therefore, the problem of inverting $Q_{i+1}^{-1}$ and $S_{i+1}^{-1}$ ($i = 0, 1, \ldots, R-2$) in (3.14a) and (3.14b) can be reduced to the problem of adding six $(C/2)\times(C/2)$ matrices and inverting four $(C/2)\times(C/2)$ matrices (for $C$ even). For large $C$, this reduces the amount of computation required in the recursive part of the algorithm by approximately 75%.

The regenerative part of Akaike's algorithm benefits from the special structure of $V$ as well. By the corollary to Lemma 3.12 every block of $V^{-1}$ is centrosymmetric, and, by Lemma 3.3 and the corollary to Lemma 3.6, $V^{-1}$ is persymmetric. These properties reduce the number of distinct elements of $V^{-1}$ to approximately one-fourth of what they would be if $V$ was merely a symmetric block Toeplitz matrix with nonsymmetric blocks. Therefore, the computations involved in (3.15) can be reduced by approximately 75%.
In summary, the covariance matrix of an RFLM that arises when a single realization of \( \mathcal{Y} \) is observed at sites having Configuration II can, if \( \mathcal{Y} \) is weakly stationary, be inverted with \( O(R^2C^3) \) computations; a further 75% (approximately) reduction is possible if \( \mathcal{Y} \) is isotropic.

As was true in the context of Configuration I, the vanishing of the covariogram of \( \mathcal{Y} \) within the domain represented by the distances between observational sites having Configuration II may result in a covariance matrix that is a band matrix. Depending on the values of \( H_1 \) and \( H_2 \), \( \mathcal{V} \) may be a band matrix and each \( C \times C \) block of \( \mathcal{V} \) may be a band matrix. Shortcuts for inverting such matrices analogous to those of Trench (1974) for inverting band Toeplitz matrices would be useful here.

The remainder of this section is devoted to another simplification for inverting \( \mathcal{V} \) that is possible in the context of Configuration II when \( \mathcal{V} \) has another kind of structure in addition to that guaranteed by part (i) of Theorem 3.3. This additional structure occurs when \( \mathcal{Y} \) satisfies a property known as separability, defined as follows.

**Definition 3.11**

Let \( h \) represent a vector in \( \mathbb{R}^d \), and let \( h_1, \ldots, h_d \) represent the elements of \( h \). A \( d \)-dimensional weakly stationary random field is said to be separable if its correlogram \( \rho(\ast) \) satisfies

\[
\rho(h_1, h_2, \ldots, h_d) = \prod_{i=1}^{d} \rho_i(h_i) \quad \text{(for all } h_1, \ldots, h_d) \tag{3.16}
\]

for \( d \) real-valued even functions \( \rho_1(\ast), \ldots, \rho_d(\ast) \).
Theorem 3.4

Suppose that a single realization of a weakly stationary two-dimensional random field \( \mathcal{F}_Y \) is observed at sites \( R_1, \ldots, R_n \) having Configuration II. Suppose further that \( \mathcal{F}_Y \) is separable, implying that there exist even functions \( \rho_1(*) \) and \( \rho_2(*) \) such that the covariogram \( \hat{C}(*) \) of \( \mathcal{F}_Y \) satisfies the condition \( \hat{C}(x,y) = \hat{C}(0,0)\rho_1(x)\rho_2(y) \) for all \( (x,y) \in \mathbb{R}^2 \). Then the covariance matrix \( V \) of both RFIM-R and RFIM-P can be represented as

\[
V = P \otimes Q,
\]

where \( P \) and \( Q \) are \( C \times C \) and \( R \times R \) symmetric Toeplitz matrices, respectively. Under RFIM-R, the \((i,k)\)th element of \( P \) and the \((j,\ell)\)th element of \( Q \) are given by

\[
P_{ik} = \frac{\hat{C}(0,0)}{|S_{11}|} \int_{S_{11}} \rho_1(x_2-x_1+H_1|k-i|)dx_1 dx_2 \quad \text{and} \quad Q_{j\ell} = \frac{1}{|S_{11}|} \int_{S_{11}} \rho_2(y_2-y_1+H_2|j-\ell|)dy_1 dy_2,
\]

respectively. Under RFIM-P, the corresponding elements of \( P \) and \( Q \) are given by

\[
P_{ik} = \hat{C}(0,0)\rho_1(H_1i-H_1k) \quad \text{and} \quad Q_{j\ell} = \rho_2(H_2j-H_2\ell).
\]
Proof:

The separability of $\mathcal{F}_Y$ implies that $\mathcal{F}_Y$ is at least reflectively anisotropic, for

\[
\hat{c}(x, -y) = \hat{c}(0, 0) \hat{p}(x, -y) = \hat{c}(0, 0) \rho_1(x) \rho_2(-y) = \hat{c}(0, 0) \rho_1(x) \rho_2(y) = \hat{c}(x, y) .
\]

Then, as noted in the discussion following the proof of Theorem 3.3, $V$ is a symmetric block Toeplitz matrix of order $(R, C)$ with symmetric Toeplitz blocks. Using the weak stationarity of $\mathcal{F}_Y$, the isometry of the regions $S_{11}, S_{12}, \ldots, S_{RC}$, and finally the separability of $\mathcal{F}_Y$, we have, for RFLM-R, that

\[
\text{Cov}(y_{ij}, y_{kl}) = \frac{1}{|S_{ij}||S_{kl}|} \int_{S_{ij}} \int_{S_{kl}} \hat{c}(x_2, y_2; y_1) dx_1 dx_2 dy_1 dy_2
\]

\[
= \frac{1}{|S_{11}|^2} \int_{S_{11}} \int_{S_{11}} \hat{c}(x_2, y_2; y_1 + H_1 | k-i |, y_2; y_1 + H_2 | j-l |) dx_1 dx_2 dy_1 dy_2
\]

\[
= \frac{\hat{c}(0, 0)}{|S_{11}|^2} \int_{S_{11}} \int_{S_{11}} \rho_1(x_2; x_1 + H_1 | k-i |) \rho_2(y_2; y_1 + H_2 | j-l |) dx_1 dx_2 dy_1 dy_2
\]

\[
= \rho_{ik} q_{j\ell} .
\]
Similarly, for RFLM-P, we find that

\[
\text{Cov}(y_{ij}, y_{kl}) = \mathcal{C}(s_{ij} - s_{kl}) \\
= \mathcal{C}(H_1|i-k|, H_2|j-l|) \\
= \mathcal{C}(0, 0)\rho_1(H_1 i - H_1 k)\rho_2(H_2 j - H_2 l) \\
= p_{ik} q_{jl}.
\]

Thus, for either RFLM, \( V = P \otimes Q \). Furthermore, by Theorem 3.2, \( P \) and \( Q \) are symmetric Toeplitz matrices. Q.E.D.

The result of Theorem 3.4 is useful in two ways. First, only \( R + C \) function evaluations are necessary to obtain \( V \) and each of its first- and second-order partial derivatives with respect to the elements of \( \theta \). Secondly, the inverse of the direct product of two nonsingular matrices is the direct product of the two inverses; thus \( V^{-1} = P^{-1} \otimes Q^{-1} \). Because \( P \) and \( Q \) are symmetric Toeplitz matrices, they can be inverted with \( O(R^2) \) and \( O(C^2) \) computations, respectively, using the Trench-Zohar algorithm described in Section 3.3. Forming the direct product of \( P^{-1} \) and \( Q^{-1} \) involves \( O(R^2C^2) \) computations, although only about one-sixteenth of the elements of this direct product actually need to be computed because \( V^{-1} \) and each one of its \( C \times C \) component blocks is symmetric and persymmetric. Therefore, when the conditions of Theorem 3.4 are met, \( V \) can be inverted with a significant saving of computational effort.
Do any two-dimensional separable random fields exist? Since we are primarily interested in isotropic (or at least geometrically anisotropic) random fields, an even more interesting question is: do random fields exist which are both isotropic and separable? The following theorem, which is an extension of a well-known result in real analysis, provides an answer to this question of existence under a generalized definition of isotropy. Let $1 \leq p < \infty$, and let $c$ be a positive integer. Let $x$ represent a point in $\mathbb{R}^c$ and let $x_i$ $(i = 1, \ldots, c)$ represent the $i$th element of $x$.

We define $L^p$-norm in the usual way as $\|x\|_p = \left( \sum_{i=1}^{c} |x_i|^p \right)^{1/p}$.

**Theorem 3.5**

The only continuous function defined on $\mathbb{R}^c_+ \equiv \{x; x \in \mathbb{R}^c$ and $x_i > 0$ for all $i\}$ that is not identically equal to zero and that satisfies the property

$$f(\|x\|_p) = \prod_{i=1}^{c} f(x_i)$$

for all $x \in \mathbb{R}^c_+$ is $f(x) = \exp(\beta x^p)$, where $\beta$ is a constant.

**Proof:**

Suppose that $f(\cdot)$ satisfies (3.17). Consider $x \in \mathbb{R}^c_+$ such that $x_1 = x_2 = \ldots = x_c = x/c^{1/p}$, where $x$ is a nonnegative real number. Then (3.17) implies that

$$f(x) = [f(x/c^{1/p})]^c.$$
Also, (3.17) implies that \( f(x)[f(0)]^{c-1} = f(x) \) for all \( x \), which requires that \( f(0) = 1 \) unless \( f(x) = 0 \) for all \( x \), which it cannot be hypothesis. Hence \( f(0) = 1 \). By the continuity of \( f(\cdot) \), there exists an \( \eta > 0 \) such that \( f(\eta) > 0 \), whence it follows from (3.18) that \( f(x) > 0 \) for all \( x \).

Now, for any positive integer \( b \), substitute \( x_i = 1/b \) (\( i = 1, \ldots, c \)) into (3.17), giving \( f(c^{1/p}/b) = [f(1/b)]^c \), or alternatively,

\[
f(c/b) = [f(1/b)]^{c^p}.
\]

(3.19)

Thus, \( f(1) = f(b/b) = [f(1/b)]^{b^p} \), which implies that

\[
f(1/b) = [f(1)]^{b^{-p}}.
\]

(3.20)

From (3.19) and (3.20),

\[
f(c/b) = \left\{ [f(1)]^{b^{-p}} \right\}^{c^p}
\]

\[= [f(1)]^{(c/b)^p}.
\]

Since the set of rational numbers is dense in \( \mathbb{R} \), the continuity of \( f(\cdot) \) implies that

\[f(x) = [f(1)]^{x^p} \text{ for all } x \in \mathbb{R}_+,
\]

or equivalently,
f(x) = \exp(x^\log[f(1)])
= \exp(\beta x^p) \quad \text{for all } x \in \mathbb{R}_+,

where \beta \in (-\infty, \infty) \text{ since } f(1) > 0. \quad \text{Q.E.D.}

**Definition 3.12**

A weakly stationary d-dimensional random field whose covariogram
is \( \hat{C}(\cdot) \) is said to be \( L^p \)-isotropic if \( \hat{C}(h) = \hat{C}(k) \) for all \( (h, k) \in \mathbb{R}^d \times \mathbb{R}^d \)
satisfying \( \|h\|_p = \|k\|_p \).

The preceding definition of \( L^p \)-isotropy is a generalization of the
ordinary definition of isotropy. Classical isotropy is equivalent to \( L^2 \)-
isotropy under this definition.

**Corollary 1**

The only \( L^2 \)-isotropic continuous correlogram of a two-dimensional
random field that satisfies the separability condition (3.16) is the
Gaussian correlogram

\[
\tilde{\rho}(r) = \exp(-\theta_2 r^2),
\]

where \( \theta_2 > 0 \).

**Proof:**

\( L^2 \)-isotropy implies that \( \rho(x, y) = \tilde{\rho}(\sqrt{x^2 + y^2}) \), so the separability
condition becomes \( \tilde{\rho}(\sqrt{x^2 + y^2}) = \rho_1(x)\rho_2(y) \). The result follows upon
setting \( p = 2 \) in Theorem 3.5. \quad \text{Q.E.D.}
Corollary 2

The only $L^1$-isotropic continuous correlogram of a two-dimensional random field that satisfies the separability condition is the $L^1$-isotropic exponential correlogram

$$\hat{\rho}(x,y) = \exp[-\theta_2(|x| + |y|)],$$

(3.22)

where $\theta_2 > 0$.

Proof:

$L^1$-isotropy implies that $\hat{\rho}(x,y) = \hat{\rho}(|x| + |y|)$, so the separability condition becomes $\hat{\rho}(|x| + |y|) = \rho_1(x)\rho_2(y)$. The result follows upon setting $p = 1$ in Theorem 3.5.

Q.E.D.

Corollaries 1 and 2 to Theorem 3.5 establish that the only two correlograms of two-dimensional random fields that are $L^1$- or $L^2$-isotropic and that satisfy the separability condition are correlograms in common use. Recall from Chapter 1 that the behavior of the $L^2$-isotropic Gaussian correlogram (3.21) near the origin is parabolic, while that of the $L^2$-isotropic exponential correlogram is linear. The behavior of the $L^1$-isotropic exponential correlogram (3.22) at the origin is also linear with respect to the $L^1$ distance definition. Thus, both parabolic and linear behavior at the origin could be modelled by correlograms which, in the context of Configuration II, result in a covariance matrix that can be inverted with $O(R^2C^2)$ computations.
Geometric anisotropy could be incorporated into correlograms (3.21) and (3.22). Consider the two correlograms

\[ \rho(x,y) = \exp\left[-(\theta_2 x^2 + \theta_3 y^2)\right] \]

and

\[ \rho(x,y) = \exp\left[-(\theta_2 |x| + \theta_3 |y|)\right], \]

where \( \theta_2 > 0 \) and \( \theta_3 > 0 \) in both cases. These correlograms have isocorrelation contours that are \( L^2 \)- and \( L^1 \)-norm ellipses, respectively. The separability condition (3.16) holds for these correlograms, so the covariance matrix of an RFLM having one of these correlograms can be inverted with \( O(R^2 C^2) \) computations when the observational sites have Configuration II.

Modjeska and Rawlings (1983) showed that the classical empirical model developed by Fairfield Smith (1938) for relating the variance of field plot data to plot size and shape in the context of Configuration II implies an underlying plot correlogram which, in its 2-dimensional version, is

\[ \rho(s_{ij},s_{kj}) = \rho_1(i-k)\rho_2(j-k), \quad (3.23) \]

where

\[ \rho_1(x) = \frac{1}{2}\left[ (|x|+1)^{2-\theta} - 2|x|^{2-\theta} + (|x|-1)^{2-\theta} \right] \text{ for } |x| \geq 1, \]

\[ \rho_2(y) = \frac{1}{2}\left[ (|y|+1)^{2-\theta} - 2|y|^{2-\theta} + (|y|-1)^{2-\theta} \right] \text{ for } |y| \geq 1, \]
and where $\theta_2 > 0$ and $\theta_3 > 0$. Note that plot correlogram (3.23) is neither isotropic nor geometrically anisotropic, but it obviously satisfies the separability condition. Thus, the $O(r^2 c^2)$ method of inverting the covariance matrix indicated by Theorem 3.4 is appropriate in conjunction with this plot correlogram if the plot yields, which are more properly regarded as regularizations of the random field, are instead regarded as point observations.

### 3.5. Computational Aspects of ML Approaches for IRF-k's

We have seen that a highly regular spatial configuration of sites at which $\mathcal{F}_Y$ is observed (such as Configuration I or Configuration II), together with certain properties of the covariogram of $\mathcal{F}_Y$, result in an RFLM whose covariance matrix $V$ has a structure that can be exploited to alleviate the computational burden of ML or REML estimation of RFLM parameters. In Section 2.5, an estimation procedure called the ML-IRF-k approach, in which the method of maximum likelihood is applied to generalized increments of order $k$ under RFLM-P, was described. In this section we characterize the structure of matrices involved in the ML-IRF-k approach when $\mathcal{F}_Y$ is an IRF-k and the spatial configuration of observational sites is Configuration I or Configuration II, seeking, once again, structures which lead to computational efficiency.

Throughout this section we adopt the same notation as that used in Section 2.5. We consider Configuration I first.
Theorem 3.6

Suppose that a single realization of a d-dimensional IRF-k \( \mathcal{F}_Y \) is observed at sites \( R_1, \ldots, R_n \) having Configuration I. Then \( \mathbf{K}_k \), \( \partial \mathbf{K}_k / \partial \theta_i \) (\( i = 1, \ldots, m \)), and \( \partial^2 \mathbf{K}_k / \partial \theta_i \partial \theta_j \) (\( i = 1, \ldots, m; j = 1, \ldots, m \)) are symmetric Toeplitz matrices under both RFLM-P and RFLM-R.

Proof:

The proof is essentially the same as the proof of Theorem 3.2 except that the generalized covariance function \( \hat{\mathbf{G}}_k(\cdot; \theta) \) of \( \mathcal{F}_Y \) and \( \mathbf{K}_k \) now play the role of \( \mathbf{C}(\cdot, \cdot; \theta) \) and \( \mathbf{V} \). Q.E.D.

Though Theorem 3.6 holds for an IRF-k of any dimension, we shall now restrict our attention, except where noted, to the case \( d = 1 \).

The ML-IRF-k approach consists of applying the method of maximum likelihood to an \( (n-\mathbf{l}^*) \times 1 \) vector \( \Lambda y \) of linearly independent generalized increments of order \( k \) under RFLM-P, where, using the same notation as in Section 2.5, \( \mathbf{l}^* \) is the number of monomials in \( \mathbf{m}_k^d \). For a one-dimensional IRF-k, it is easily verified that \( \mathbf{l}^* = k+1 \), so that \( n-\mathbf{l}^* = n-k-1 \). Using (2.17), the log-likelihood function associated with \( \Lambda y \) differs by no more than an additive constant from the function

\[
L_3(\theta; y) = -\frac{1}{2} \log |\Lambda K_k \Lambda| - \frac{1}{2} y' \Lambda (\Lambda K_k \Lambda)^{-1} \Lambda y. \tag{3.24}
\]

The \( (n-k-1) \times (n-k-1) \) matrix \( \Lambda K_k \Lambda \) plays a major role in the sequel and is denoted by \( Q_k \). Note that \( L_3 \) depends on \( K_k \) only through \( Q_k \).
Note that we have not used a superscript on the matrix $K_k$ in (3.24). Strictly speaking, the ML-IRF-k approach, as defined in Section 2.5, applies only to RFLM-P. However, in the one-dimensional situation of Configuration I to which we are restricting attention, it is easily verified that $\Lambda_y$ is a vector of $p^*$ linearly independent error contrasts under RFLM-R as well as under RFLM-P, so that $\Lambda_{K_{k}^R} = \Lambda_{V^R}$. Consequently, the ML-IRF-k approach could be applied to RFLM-R in this setting, and we shall omit superscripts on $K_k$ in this section, understanding that $K_k$ represents a matrix that is equal to either $K_k^P$ or $K_k^R$.

A very important consequence of Theorem 2.1 is that the estimator of $\theta$ obtained by maximizing $L_3$ is the same no matter what set of $n-k-1$ generalized increments of order $k$ are used, provided, of course, that they are linearly independent. Therefore, we may as well use a set of increments that can reduce the computational burden of the estimation procedure, if indeed such a set can be devised. One such set can be constructed from what we shall call the elementary generalized increment of order $k$.

**Definition 3.13**

The elementary generalized increment of order $k$ is $\hat{\lambda}^{(k)}$, where

$$\hat{\lambda}^{(k)} = (a_1^{(k)}, a_2^{(k)}, \ldots, a_{k+2}^{(k)}, 0, 0, \ldots, 0)'$$

and

$$a_i^{(k)} = (-1)^{i-1}{k+1 \choose i-1} (i = 1, \ldots, k+2).$$
Note that the nonzero coefficients of $\lambda^{(k)}$ are the coefficients in the binomial expansion of $(1-x)^{k+1}$. As examples, the elementary generalized increments of order 0, 1, and 2 are $\lambda^{(0)}_k$, $\lambda^{(1)}_k$, and $\lambda^{(2)}_k$, respectively, where

$$
\lambda^{(0)}_k = (1, -1, 0, 0, \ldots, 0)',
$$
$$
\lambda^{(1)}_k = (1, -2, 1, 0, 0, \ldots, 0)', \text{ and }
$$
$$
\lambda^{(2)}_k = (1, -3, 3, -1, 0, 0, \ldots, 0)'.
$$

For any value of $k$, let $\Lambda_k$ represent the $n \times (n-k-1)$ matrix whose first column $\lambda^{(k)}_k$ is equal to $\lambda^{(k)}_k$ and whose remaining columns $\{\lambda^{(k)}_j: j = 2, \ldots, n-k-1\}$ are given by the relation $\lambda^{(k)}_j = \Gamma^{j-1}\lambda^{(k)}_1$, where $\Gamma$ is an $n \times n$ matrix with ones on the first subdiagonal and zeroes elsewhere, i.e.,

$$
\Gamma = \begin{bmatrix}
0 & & & \\
1 & 0 & & \\
& \ddots & \ddots & \\
& & 1 & 0
\end{bmatrix}.
$$

Below, $\Lambda_k$ is illustrated for $k=0, 1,$ and 2:

$$
\Lambda_0 = \begin{bmatrix}
1 & & & \\
-1 & 1 & & \\
& -1 & \ddots & \\
& & \ddots & 1
\end{bmatrix},
$$

$$
\Lambda_1 = \begin{bmatrix}
1 & & & \\
-1 & 1 & & \\
& -1 & \ddots & \\
& & \ddots & 1
\end{bmatrix},
$$

$$
\Lambda_2 = \begin{bmatrix}
1 & & & \\
-1 & 1 & & \\
& -1 & \ddots & \\
& & \ddots & 1
\end{bmatrix}.
$$
Let $\{\hat{\lambda}_{ij}^{(k)}: i = 1, \ldots, n; j = 1, \ldots, n-k-1\}$ denote the elements of $\hat{\Lambda}_k$. Note, by the definition of $\hat{\Lambda}_k$, that

\begin{align*}
\hat{\lambda}_{ij}^{(k)} &= 0 \text{ for } i < j , \quad (3.25) \\
\hat{\lambda}_{ij}^{(k)} &= \hat{\lambda}_{i-j+1,1}^{(k)} \text{ for } i \geq j , \quad (3.26) \\
\text{and} \\
\hat{\lambda}_{11}^{(k)} &= 0 \text{ for } i > k+2 . \quad (3.27)
\end{align*}
Put $\hat{Q}_k = \hat{A}_k K_k \hat{A}_k^*$, and let $\hat{q}_{ij}^{(k)}$ represent the $(i,j)$th element of $\hat{Q}_k$
$(i = 1, \ldots, n-k-1; j = 1, \ldots, n-k-1)$. Let $k_s^{(k)}$ represent the $s$th column of $K_k$ $(s = 1, \ldots, n)$, and let $k_{ij}^{(k)}$ represent the $(i,j)$th element of $K_k$ $(i = 1, \ldots, n; j = 1, \ldots, n)$.

**Theorem 3.7**

Suppose that a single realization of a one-dimensional IRF-$k \mathcal{F}_Y$
is observed at sites $R_1, \ldots, R_n$ having Configuration I. Then $\hat{Q}_k$, $\partial^2 \hat{Q}_k / \partial \theta_i \partial \theta_j$ ($i = 1, \ldots, m$; $j = 1, \ldots, m$) are symmetric Toeplitz matrices.

**Proof:**

The symmetry of $\hat{Q}_k$ follows from that of $K_k$. Therefore, attention can be restricted to those elements $\hat{q}_{ij}^{(k)}$ of $\hat{Q}_k$ such that $j \geq i$. We have

$$\hat{q}_{ij}^{(k)} = \sum_{s=1}^{n} \hat{\lambda}_i^{(k)} (k) \hat{\lambda}_s^{(k)} \hat{\lambda}_j^{(k)}$$

by (3.25). Now, for $s \geq i$, it follows from (3.25), (3.26), (3.27), and Theorem 3.6 that

$$\hat{\lambda}_i^{(k)} (k) \hat{\lambda}_s^{(k)} = \sum_{t=1}^{n} \hat{\lambda}_t^{(k)} (k)$$
Substituting (3.29) into (3.28) and using (3.26) once again, we find that
\[
q_{ij} = \sum_{s=j}^{n} \left( \lambda_{k}^{(k)} \right)^{t_{s-i+1}} \lambda_{s_{j}}^{(k)} 
\]
\[
= \sum_{s=j}^{n} \left( \lambda_{k}^{(k)} \right)^{t_{s-i+1}} \lambda_{s_{j}}^{(k)} 
\]

Renumbering the index of summation by putting \( r = s - j + 1 \) and then using (3.27), we have that
\[
q_{ij}^{(k)} = \sum_{r=1}^{n-j+1} \left( \lambda_{k}^{(k)} \right)^{t_{r+j-1}} \lambda_{r_{1}}^{(k)} 
\]
\[
= \sum_{r=1}^{k+2} \left( \lambda_{k}^{(k)} \right)^{t_{r+j-1}} \lambda_{r_{1}}^{(k)} ,
\]
which depends on \( i \) and \( j \) only through \( j-i \). Thus, \( \hat{Q}_{k} \) is symmetric Toeplitz. Q.E.D.
One particularly useful class of generalized covariance functions, introduced by Matheron (1973), is the class of functions

\[ G_k(r; \theta) = \sum_{l=1}^{k+1} (-1)^l \theta_l r^{2l-1} \]  

(3.31)

defined for any dimension d and any nonnegative integer k, where the parameters must satisfy the condition

\[ \sum_{l=1}^{k+1} \frac{\Gamma(\frac{2l+d-1}{2})}{\Gamma(l+\frac{1}{2}) \pi^{\frac{2l+d}{2}}} \theta_l u^{-d+2l+3} \geq 0 \quad \text{for all } u > 0. \]  

(3.32)

This class is known as the class of polynomial isotropic generalized covariance functions. It is not a very rich class for small k; for example, when k = 0, the only members of this class are linear functions of r. Condition (3.32) imposes certain inequality constraints on the elements of \( \theta \). Delfiner (1976) gives these constraints for the cases k=0, 1, 2 and d=1, 2, 3. It is worth noting that these constraints are not linear when k > 1 and d > 1; however, for the case d=1 to which we are restricting attention, the constraints are linear for k=0, 1, and 2.

The class (3.31) has at least three useful properties for the estimation of \( \theta \). First, \( G_k(r; \theta) \) is a linear function of the parameters, which makes its first-order partial derivatives with respect to the
elements of $\theta$ easy to compute and implies that all higher-order derivatives are zero. The linearity of $G^*(r;\theta)$ also facilitates the estimation of $\theta$ by one of the ad hoc regression techniques described in Section 2.1.

Secondly, $G^*(r;\theta)$ depends on $r$ in a relatively simple way, with the consequence that integrals of $G^*(r;\theta)$ over rectangular regions in $\mathbb{R}^d$ (which are relevant in the RFLM-$R$ context) are easy to evaluate. Thirdly, for a given $d$, \( \{G^*(r;\theta): k = 0, 1, \ldots\} \) is a nested sequence of functions; hence, the class (3.31) lends itself nicely to inference about the order $k$ of the IRF-$k$. As noted earlier, however, one disadvantage of generalized covariance functions of order two or higher is that when they are used in connection with random fields of dimension two or higher, the constraints on the parameters are nonlinear.

In the remainder of this section we attach an asterisk to $K^*_k$, $Q^*_k$, and $\hat{Q}_k$ and to their elements when the elements are obtained by evaluating the $k$th-order polynomial isotropic generalized covariance function $G^*_k(r;\theta)$ of a one-dimensional IRF-$k$.

The following example illustrates the symmetric Toeplitz structure of $K^*_k$ and $Q^*_k$ under an RFLM-$P$ in which $\mathcal{F}_y$ is an IRF-0, and demonstrates that $V$ and $Q_k$ do not in general possess that structure.

**Example 3.1**

Let $\mathcal{F}_1$ represent a one-dimensional random field whose mean function is constant and whose covariogram is $C(s,t;\theta) = \theta_1 \min\{s,t\}$ where $\theta_1 > 0$. It was established previously (see Section 1.1) that $\mathcal{F}_1$ is an IRF-0. The
variogram of $\mathcal{F}_1$ is the so-called linear variogram $2\gamma(r; \theta_1) = \theta_1 r$ (for $r \geq 0$). To see this, observe that

$$E[(Z_s - Z_{s+r})^2] = E[Z_s^2] + E[Z_{s+r}^2] - 2E[Z_s Z_{s+r}]$$

$$= C(s,s) + C(s+r,s+r) - 2C(s,s+r)$$

$$= \theta_1 s + \theta_1 (s+r) - 2\theta_1 s$$

$$= \theta_1 r .$$

Suppose that $\mathcal{F}_1$ is observed at the points $s_1 = 1, s_2 = 3, s_3 = 5, s_4 = 7,$ and $s_5 = 9$ in $\mathbb{R}$; note that the spatial configuration of these points is Configuration I. Then,

$$V = \theta_1 \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 3 & 3 & 3 \\ 1 & 3 & 5 & 5 & 5 \\ 1 & 3 & 5 & 7 & 7 \\ 1 & 3 & 5 & 7 & 9 \end{bmatrix}$$

and

$$K_0^* = \theta_1 \begin{bmatrix} 0 & -1 & -2 & -3 & -4 \\ -1 & 0 & -1 & -2 & -3 \\ -2 & -1 & 0 & -1 & -2 \\ -3 & -2 & -1 & 0 & -1 \\ -4 & -3 & -2 & -1 & 0 \end{bmatrix} .$$

Furthermore,
and it is easily verified that
\[
\begin{pmatrix}
2 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 2
\end{pmatrix}
\]
Another vector of linearly independent generalized increments of order 0 is
\[
\begin{pmatrix}
1 & 1 & -2 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1
\end{pmatrix}^T .
\]
However, the matrix \( Q_0^* \) corresponding to this set of increments is
\[
\begin{pmatrix}
10 & 4 & 0 & 0 \\
4 & 2 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 2
\end{pmatrix}
\]
illustrating that \( Q_k \) is not, in general, a Toeplitz matrix.
The results of Theorems 2.1, 3.2, 3.6, and 3.7 have important implications for the estimation of RFLM parameters by maximum likelihood approaches when the spatial configuration of observational sites is Configuration I. First, $K_k$ and $\hat{Q}_k$, like $V$, are symmetric Toeplitz matrices when $F_Y$ is weakly stationary. However, when $F_Y$ is not weakly stationary, $V$ is no longer symmetric Toeplitz, but $K_k$ and $\hat{Q}_k$ remain so, provided that $F_Y$ is an IRF-k. Thus, when the standard REML approach is equivalent to the ML-IRF-k approach (as is the case under the conditions of Theorem 2.1), the REML estimator of $\theta$ can be obtained by forming and inverting symmetric Toeplitz matrices (using the Trench-Zohar algorithm) even if $F_Y$ is not weakly stationary. A second implication pertains to how $L_3$ is maximized. Because $L_3$ depends on $K_k$ only through $Q_k$, the maximization of $L_3$ by a Newton-Raphson or scoring algorithm could be accomplished by computing the matrices $Q_k$, $\partial Q_k / \partial \theta_i$, and $\partial^2 Q_k / \partial \theta_i \partial \theta_j$ or $E\{\partial^2 Q_k / \partial \theta_i \partial \theta_j\}$ on each iteration rather than by computing the matrices $K_k$, $\partial K_k / \partial \theta_i$, and $\partial^2 K_k / \partial \theta_i \partial \theta_j$ or $E\{\partial^2 K_k / \partial \theta_i \partial \theta_j\}$ on each iteration. If both $K_k$ and $Q_k$ were symmetric Toeplitz matrices, then, by virtue of the smaller dimensions of $Q_k$, fewer function evaluations would be required to form $Q_k$ than to form $K_k$, and fewer computations would be required to invert $Q_k$ than to invert $K_k$. Although the conditions of Theorem 3.6 are sufficient for $K_k$ to be symmetric Toeplitz, they are not sufficient for $Q_k$ to be symmetric Toeplitz. According to Theorem 3.7, however, $\hat{Q}_k$, $\partial \hat{Q}_k / \partial \theta_i$, and $\partial^2 \hat{Q}_k / \partial \theta_i \partial \theta_j$ ($i = 1, \ldots, m; j = 1, \ldots, m$) are symmetric Toeplitz matrices. Therefore, a seemingly efficient way to carry out the estimation of $\theta$ is to maximize the function
with respect to $\theta \in \Theta$.

In Example 3.1, $K_k$ and $\hat{Q}_k$ were illustrated for an IRF-0 whose
generalized covariance function is $G^*(r;\theta)$. In fact, we saw that $\hat{Q}_0^*$
was diagonal and thereby had even more structure than the symmetric
Toeplitz structure guaranteed by Theorem 3.7. This example suggests
that $\hat{Q}_k$ may have additional structure when $\hat{G}_k(h;\theta)$ is the kth-order
polynomial isotropic generalized covariance function. The precise
nature of this structure is established by the following theorem.

**Theorem 3.8**

Suppose that a single realization of a one-dimensional IRF-$k$ $\mathcal{F}_Y$
is observed at sites $R_1, \ldots, R_n$ having Configuration I. Suppose
further that the generalized covariance function of $\mathcal{F}_Y$ is $G^*(r;\theta)$.
Then:

(i) under RFLM-R, $\hat{Q}_k^*$ and $\partial Q_k^*/\partial \theta_i$ ($i = 1, \ldots, m$) are symmetric
    Toeplitz band matrices of order $2k+3$;
(ii) under RFLM-P, $\hat{Q}_k^*$ and $\partial Q_k^*/\partial \theta_i$ ($i = 1, \ldots, m$) are symmetric
    Toeplitz band matrices of order $2k+1$.

**Proof of (i):**

Because $\hat{Q}_k^*$ is a symmetric Toeplitz matrix by Theorem 3.7, we can
restrict attention to the elements $\{q_{ij}^*(k) : j = 1, \ldots, n-k-1\}$ of the

\[ L_3(\theta; y) = -\frac{1}{2} \log |\hat{Q}_k| - \frac{1}{2y} \hat{Q}_k y \]
first row of \( \hat{Q} \). Under RFLM-R, when a one-dimensional random field is observed at sites having Configuration I, those sites are merely non-overlapping equally-sized intervals \( S_1, \ldots, S_n \) whose midpoints are equally-spaced. Define

\[
\bar{G}_k^*(i-u) = \int_{S_1} \int_{S_u} G_k^*(s-t) \, ds \, dt,
\]

i.e., \( \bar{G}_k^*(i-u) \) is the \((i,u)\)th element of \( \mathbf{K}_k^* \). The function \( \bar{G}_k^*(\cdot) \) is well defined since, according to Theorem 3.6, the \((i,u)\)th of \( \mathbf{K}_k^* \) depends on \( i \) and \( u \) only through \( i-u \). Note that \( \lambda(k)_{i-u}^* \) depends on \( i \) and \( u \) only through \( i-u \). By (3.27), this can be simplified to

\[
\lambda(k)_{i-u}^* = \sum_{u=1}^{n} a_u (k) \bar{G}_k^*(i-u) \tag{3.33}
\]

Using (3.30) and (3.33), we have that

\[
\hat{q}_{ij}^* = \lambda(k)_{i-j}^* \lambda(k)_{j-i}^* \sum_{s=1}^{k+2} \sum_{u=1}^{k+2} a_s (k) \bar{G}_k^*(s+j-1-u) a_u (k)
\]

\[
= \sum_{s=1}^{k+2} \sum_{u=1}^{k+2} \{ a_s (k) \bar{G}_k^*(s+j-1-u) \} a_u (k)
\]

\[
= \sum_{s=1}^{k+2} \sum_{u=1}^{k+2} a_s (k) a_u (k) \bar{G}_k^*(j+s-u-1).
\]
Putting $t = s - u$, an alternative representation for this sum is

\[
\sum_{t=-k-1}^{k+1} a_u (k) (k) \frac{G^*_k (j+t-1)}{u+|t|} = b^{(k)}_{t} G^*_k (j+t-k-3), \tag{3.34}
\]

where

\[
b^{(k)}_{t} = \sum_{u=1}^{2k+3} a_u (k) a_u (k) \frac{G^*_k (j+t-k-3)}{u+|t-k-2|} \quad (t=1, \ldots, 2k+3). \tag{3.35}
\]

Expression (3.35) vanishes when $j > k+3$, as we now demonstrate. From (3.34), we have that

\[
p_{1j} = \sum_{t=1}^{2k+3} b^{(k)}_{t} \int_{0}^{1} \int_{j+t-k-3}^{j+t-k-2} (-1)^{k+1} \theta_2 \sum_{t=1}^{2k+3} b^{(k)}_{t} \sum_{t=1}^{2k+3} \frac{u-v}{2^k-1} dudv.
\]

It can be shown (see Section 3.7) that for $p = 0, 1, \ldots, 2k+1$, the coefficients \(b^{(k)}_{t} : t = 1, \ldots, 2k+3\) satisfy the conditions

\[
\sum_{t=1}^{2k+3} b^{(k)}_{t} |x+t|^p = 0 \quad \forall x \in [-1, \infty). \tag{3.37}
\]
Now, $j \geq k+3$ implies that $u-v+j-k-3 \geq -1$ for all $(u,v) \in [0,1] \times [0,1]$. Thus, the integrand of (3.36) vanishes for all $\ell = 1, \ldots, k+1$, implying further that $q_{1j}^{*}(k) = 0$ when $j \geq k+3$. Therefore, $Q_k^{*}$ is a symmetric Toeplitz band matrix of order $2k+3$.

That $\partial q_{1j}^{*}/\partial \theta_i$ $(i = 1, \ldots, k+1)$ is a band matrix is established by an argument similar to that used to establish that $Q_k^{*}$ is a band matrix. Clearly,

$$\partial q_{1j}^{*}(k)/\partial \theta_i = (-1)^i \int_0^1 \int_0^{2k+3} \sum_{t=1}^{2k+3} b_t^{(k)} |u-v+j+t-k-3|^{2i-1},$$

which vanishes by (3.37) for all $i = 1, \ldots, k+1$ provided that $j \geq k+3$. Q.E.D.

**Proof of (ii):**

As in the proof of (i), we may restrict attention to the elements $\{q_{1j}^{*}(k) : j = 1, \ldots, n-k-1\}$ of the first row of $Q_k^{*}$. Using exactly the same reasoning as that used to derive expression (3.34), we obtain

$$q_{1j}^{*}(k) = \sum_{t=1}^{2k+3} b_t^{(k)} \sum_{\ell=1}^{k+1} (-1)^{\ell} \theta_{\ell} |j+t-k-3|^{2\ell-1}.$$

Thus,

$$q_{1j}^{*}(k) = \sum_{\ell=1}^{k+1} (-1)^{\ell} \theta_{\ell} \sum_{t=1}^{2k+3} b_t^{(k)} |j+t-k-3|^{2\ell-1}$$

$$= \sum_{\ell=1}^{k+1} (-1)^{\ell} \theta_{\ell} \sum_{t=1}^{2k+3} b_t^{(k)} |j+t-k-3|^{2\ell-1}$$

(3.38)
Note that when \( j \geq k+2 \) in (3.38), then \( j-k-3 \geq -1 \). Consequently, (3.37) implies that

\[
\sum_{t=1}^{2k+3} b_t^{(k)} |j+t-k-3|^{2\ell-1} = 0 \quad \text{for } \ell = 1, \ldots, k+1,
\]

by which it follows that \( Q_{1j}^{*}\) is 0 when \( j \geq k+2 \). Therefore, \( Q_{k}^{*}\) is a symmetric Toeplitz band matrix of order \( 2k+1 \). Furthermore,

\[
\frac{\partial Q_{1j}^{*}}{\partial \theta_i} = (-1)^i \sum_{t=1}^{2k+3} b_t^{(k)} |j+t-k-3|^{2i-1},
\]

which vanishes by (3.37) for all \( i = 1, \ldots, k+1 \) provided that \( j \geq k+2 \).

If the conditions of Theorem 3.8 are satisfied, the maximization of

\[
-\frac{1}{2} \log |Q_{k}^{*}| - \frac{1}{2y} Q_{k}^{*} y
\]

with respect to \( \theta \in \Theta \) (which is equivalent to REML estimation under the conditions of Theorem 2.1) can be accomplished relatively efficiently by exploiting the structure of \( Q_{k}^{*}\) and \( \frac{\partial Q_{k}^{*}}{\partial \theta_i} \) (and using the fact that \( \frac{\partial^2 Q_{k}^{*}}{\partial \theta_i \partial \theta_j} = 0 \) for all \( i \) and \( j \)). This is particularly true for an IRF-0 (in which case the aforementioned matrices are Toeplitz diagonal under RFLM-P and symmetric Toeplitz tridiagonal under RFLM-R) and for an IRF-1 (in which case the matrices are...
symmetric Toeplitz tridiagonal under RFLM-P and can thus be inverted using formula (3.9)].

Finally, for completeness, we state a theorem pertaining to the structure of $K_k$ when the spatial configuration of observational sites is Configuration II. For two-dimensional random fields, $k^* = (k+1)(k+2)/2$.

**Theorem 3.9**

Suppose that a single realization of a d-dimensional IRF-k $F_Y$ is observed at sites $R_1, \ldots, R_n$ having Configuration II. Then $K_k$, $\Theta K_k/\Theta_1$, and $\Theta^2 K_k/\Theta_1 \Theta_j$ [i = 1, ..., (k+1)(k+2)/2; j = 1, ..., (k+1)(k+2)/2] are symmetric block Toeplitz matrices whose blocks are $C \times C$ Toeplitz matrices. If, in addition, $F_Y$ is isotropic, then each block is symmetric in addition to being Toeplitz.

**Proof:**

The proof is essentially the same as that of Theorem 3.3 with the generalized covariance function $G_k(\cdot, \cdot) \Theta$ of $F_Y$ now playing the role of $C(\cdot, \cdot; \Theta)$. Q.E.D.

Although $K_k$ has block Toeplitz structure and Toeplitz structure within blocks when the configuration of sites is Configuration II (which, as we have seen, can significantly reduce the computational burden of RFLM parameter estimation), it is unfortunately much more
difficult to devise a $Q_k$ with useful structure for two-dimensional IRF-k's than it is for one-dimensional IRF-k's.

3.6. Evaluating Integrals Under RFLM-R

In order to estimate the parameters of an RFLM-R, we have seen that certain integrals of the covariogram or generalized covariance function of $\mathcal{F}_Y$ must be evaluated. Under RFLM-R, if $C(\cdot, \cdot; \theta)$ is the covariogram of $\mathcal{F}_Y$, the $(i,j)$th element of $V$ is

$$\int_{S_i} \int_{S_j} \int_{\mathbb{R}} C(s, t; \theta) ds dt \quad (i = 1, \ldots, n; j = 1, \ldots, n). \quad (3.39)$$

If $\mathcal{F}_Y$ is an IRF-k whose generalized covariance function is $\mathcal{G}_k(\cdot; \theta)$, the corresponding element of $K_k$ is

$$\int_{S_i} \int_{S_j} \mathcal{G}_k(s - t; \theta) ds dt \quad (i = 1, \ldots, n; j = 1, \ldots, n). \quad (3.40)$$

Similarly, the $(i,j)$th elements of $\partial V/\partial \theta^\perp_x$, $\partial^2 V/\partial \theta^\perp_x \partial \theta^\perp_u$, $\partial K_k/\partial \theta^\perp_x$ and $\partial^2 K_k/\partial \theta^\perp_x \partial \theta^\perp_u$ are given by

$$\int_{S_i} \int_{S_j} \frac{\partial}{\partial \theta^\perp_x} C(s, t; \theta) ds dt, \quad (3.41)$$

$$\int_{S_i} \int_{S_j} \frac{\partial^2}{\partial \theta^\perp_x \partial \theta^\perp_u} C(s, t; \theta) ds dt. \quad (3.42)$$
respectively. In each of the expressions (3.39)-(3.44), the notation \( \int_{S_1} \) represents a d-dimensional integral.

In most applications it is difficult to evaluate integrals of the form (3.39)-(3.44) analytically. Typically, the regions are rectangular and \( \mathcal{Y} \) is assumed to be \( L^2 \)-isotropic or at least geometrically \( L^2 \)-anisotropic, with the consequence that isocorrelation surfaces are ellipsoidal. Rectangular regions make the limits of integration easy to obtain in a rectangular coordinate system but not in a polar or elliptical coordinate system, while the \( L^2 \)-isotropy assumption makes the integrand more suited for integrating in a polar or elliptical coordinate system than in a rectangular coordinate system.

This difficulty can be avoided if we assume that \( \mathcal{Y} \) is \( L^1 \)-isotropic (or \( L^1 \)-geometrically anisotropic) rather than \( L^2 \)-isotropic (or \( L^2 \)-geometrically anisotropic), for then the isocorrelation surfaces are rhombuses rather than ellipsoids. For example, if the covariogram of \( \mathcal{Y} \) is

\[
C(x,y; \theta) = \theta_1 \exp\{-\theta_2(|x|+|y|)\},
\]
and the regions $S_1, \ldots, S_n$ are rectangular, then (3.39), (3.41), and (3.42) can be evaluated analytically.

When regions are rectangular and $L^2$-isotropy (or $L^2$-geometric anisotropy) is assumed, it may be necessary to evaluate (3.39)-(3.44) by numerical methods. Several numerical methods have been proposed in the geostatistical literature. One method is to replace each of the regions $S_i$ and $S_j$ by a finite set of points on a regular grid, taking the sum of the values of the function at each pair of points $\{s,t: s \in S_i, t \in S_j\}$ to be the (approximate) value of the integral (see David, 1976; Clark, 1976). This method has been widely used in mining applications. Other methods, e.g., Simpson's rule or the trapezoidal rule, could be considered.

The most efficient numerical method proposed to date is seemingly that described by Davis and David (1978). It requires that regions be isometric and superimposable, which is the case, for example, for Configurations I and II. This method employs Cauchy's algorithm (Journel and Huijbregts, 1978, p. 98) to reduce (3.39)-(3.44), which are multiple integrals of order $2d$, to multiple integrals of order $d$. For example, when $d = 1$ and the observational sites have length $J$ and have Configuration I, Cauchy's algorithm reduces (3.39) to

$$\int_{-\infty}^{\infty} u \Gamma_1(H| i-j| + J [1-u]) du ,$$

where

$$\Gamma_1(Hk+a) = \frac{1}{2} [\hat{C}(Hk+a) + \hat{C}(Hk-a)].$$
When \( d = 2 \) and the observational sites are rectangular with dimensions \( J_1 \times J_2 \) and have Configuration II, Cauchy's algorithm reduces

\[
\iint_{S_{ij} S_{k\ell}} C(s,t;\theta) ds dt \quad \text{to}
\]

\[
\frac{1}{4} \int_0^1 \int_0^1 u_1 u_2 \Gamma_2 (H_1 | i-k | + J_1 | 1-u_1 | , H_2 | j-\ell | + J_2 | 1-u_2 | ) du_1 du_2, \tag{3.46}
\]

where

\[
\Gamma_2 (H_1 x+a, H_2 y+b) = \frac{1}{4} [ C(H_1 x+a, H_2 y+b) + C(H_1 x-a, H_2 y+b) + C(H_1 x+a, H_2 y-b) + C(H_1 x-a, H_2 y-b) ].
\]

A similar formula can be obtained for integrals (3.40)-(3.44), and all of these formulae can be extended for use when \( d > 2 \) (see Journel and Huijbregts, 1978, p. 101).

Expressions (3.45) and (3.46) can be evaluated to a remarkable degree of accuracy with relatively few functions evaluations by the Gauss numerical integration method (Davis and Rabinowitz, 1984, Section 2.7). The discrete sums that approximate (3.45) and (3.46) are

\[
\sum_{u=1}^{u_0} \Sigma \lambda \Gamma_1 (H | i-j | + J [1-x_u] ) \tag{3.47}
\]

and

\[
\sum_{u=1}^{u_0} \Sigma \lambda \lambda \Gamma_2 (H_1 | i-k | + J_1 [1-x_u] , H_2 | j-\ell | + J_2 [1-x_v] ) \tag{3.48}
\]

respectively. Here, \( \lambda_u \) and \( x_u \ (u = 1, \ldots, u_0) \) are the weights and abscissae corresponding to the Gauss moment integration formula;
these weights and abscissae are given for various values of $u_0$ by Abramowitz and Stegun (1964, p. 921). The discrete sum (3.47) requires the covariogram to be evaluated at $2u_0$ values, while (3.48) requires the covariogram to be evaluated at $4u_0^2$ values. For arbitrary $d$, the number of evaluations required is $(2u_0)^d$. Computational investigations by Davis and David (1978) and Journel and Huijbregts (1978), p. 102) suggest that (3.47) and (3.48) differ from the exact value of the appropriate integral by a very small amount for quite small $u_0$; for example, the relative precision of (3.48) for integrals of $L^2$-isotropic spherical and exponential covariograms over rectangular regions was always better than 1% for $u_0$ as small as 4. Therefore, in practice, each of the integrals (3.39)-(3.44) can be obtained to a sufficient degree of accuracy with only $[(2)(4)]^d = 8^d$ function evaluations.

3.7. Appendix: Derivation of Result (3.37)

In this appendix, we show that, for all $p = 0, 1, \ldots, 2k+1$,

$$
\sum_{t=1}^{2k+3} b^{(k)}_t |x+t|^p = 0 \quad \text{for all } x \in [-1, \infty),
$$

where $b^{(k)}_t$ was defined by expression (3.35). In doing so, we make use of two well-known combinatorial identities. For easy reference, we state these identities as lemmas.
Lemma A.1

For positive integers \(a, b,\) and \(c\) such that \(a \geq c\) and \(b \geq c\),
\[
\sum_{i=0}^{c} \binom{a}{i} \binom{b}{c-i} = \binom{a+b}{c}.
\]

Lemma A.2

For any positive integer \(p\) and all \(j = 0, 1, \ldots, p-1,\)
\[
\sum_{i=0}^{p} \binom{p}{i} (-1)^{i+1} (i+1)^j = 0.
\]

Now, by definition, for \(t = 1, 2k+3,\)
\[
b_t^{(k)} = k+2-|t-k-2| \sum_{u=1}^{k+2-|t-k-2|} a_u^{(k)} a_{u+|t-k-2|}
\]
\[
= \sum_{u=1}^{k+2-|t-k-2|} (-1)^{u-1} \binom{k+1}{u-1} (-1)^{u+|t-k-2|-1} \binom{k+1}{u+|t-k-2|-1}
\]
\[
= (-1)^{t-k} \sum_{u=0}^{k+1-|t-k-2|} \binom{k+1}{u} \binom{k+1}{u+|t-k-2|}.
\]

Note that, for any \(k,\) \(b_t^{(k)}\) depends on \(t\) only through \(|t-k-2|\). Thus,
\[
b_t^{(k)} = b_{2k+4-t}^{(k)} (t = 1, \ldots, k+2).\] Substituting \(a = b = k+1\) and \(c = 2k+3-t\) in Lemma A.1, we find that, for \(t \geq k+2,\)
\[
\sum_{i=0}^{2k+3-t} \binom{k+1}{i} \binom{k+1}{2k+3-t-i} = \binom{2k+2}{2k+3-t},
\]
or equivalently,
Thus,

\[ b_t^{(k)} = (-1)^{t-k} \binom{2k+2}{t} \]

\[ = (-1)^{t-k} (-1)^{t-1} \binom{2k+2}{t-1} \]

\[ = (-1)^{k-1} a_t^{(2k+1)} \text{ for } t = 1, \ldots, 2k+3. \]

Consequently, Lemma A.2 implies that for \( p = 0, 1, \ldots, 2k+1, \)

\[
\sum_{t=1}^{2k+3} b_t^{(k)} t^p = \sum_{t=1}^{2k+3} (-1)^{t-k} \binom{2k+2}{t-1} t^p
\]

\[ = (-1)^{k-1} \sum_{t=0}^{2k+2} (-1)^t \binom{2k+2}{t} (t+1)^p \]

\[ = 0. \]

Finally, this implies that, for any \( x, \)

\[
\sum_{t=1}^{2k+3} b_t^{(k)} (t+x)^p = \sum_{t=1}^{2k+3} b_t^{(k)} \sum_{i=0}^{p} \binom{p}{i} x^i t^i p-i
\]

\[ = \sum_{i=0}^{p} \binom{p}{i} x^i \sum_{t=1}^{2k+3} b_t^{(k)} t^i p-i
\]

\[ = 0. \quad (A.1)\]
When \( x > -1 \), the equality (A.1) can be re-expressed as

\[
\sum_{t=1}^{2k+3} b_t^{(k)} |t+x|^p = 0 ,
\]

for all \( p = 0, 1, \ldots, 2k+1 \), which is what we set out to show.
4. ASYMPTOTIC PROPERTIES OF ML ESTIMATORS

One virtue of ML, as a general approach to estimation, is that, subject to certain regularity conditions, ML estimators are consistent and have an asymptotic normal distribution. In this chapter, we seek conditions that ensure the consistency and asymptotic normality of the ML estimators of RFLM parameters. It can be argued that asymptotic properties in the RFLM setting are somewhat irrelevant since the ML estimation procedure is often impractical for large n. However, we shall confine our attention to the case of a weakly stationary random field that is observed at sites having Configuration I or Configuration II, in which case we have established that the ML estimation procedure is practical even for relatively large values of n. We do not consider the asymptotic behavior of REML estimators or the estimates obtained by the ML-IRF-k approach; an asymptotic theory for these estimation procedures has yet to be developed.

Consider an RFLM for which y is comprised of c = n/q q-variate subvectors that are independently and identically normally distributed. In such an RFLM, y can be regarded as a vector of c observations, where each observation has a q-variate normal distribution, and as c \rightarrow \infty (while q is fixed), the consistency and asymptotic normality of the ML estimator \hat{\phi} of \phi can be demonstrated (provided that \hat{\phi} is an interior point of \Theta) (Anderson, 1971). (The requirement that \hat{\phi} be an interior point of \Theta is ordinarily not any restriction for an RFLM because the
parameter space for $\hat{\phi}$ is usually an open subset of $\mathbb{R}^m$ [cf. expression (2.10)]. One subclass of RFIM's that can be so regarded is that for which: (1) $F_Y$ is an isotropic jointly normally-distributed random field whose covariogram vanishes at a finite distance; (2) the observational sites $R_1, ..., R_n$ can be partitioned into $c$ "clusters" of $q$ sites each such that the distance between any two sites from different clusters exceeds the range of the covariogram; (3) the spatial configuration of sites within a cluster is the same (apart from rotations) for each cluster.

Ordinarily, however, the observational vector of an RFIM cannot be partitioned into independently and identically distributed subvectors, and we shall have to regard $y$ as a single observation from $N_n(X_\sim, V)$ (where $N_n(\mu, A)$ denotes the $n$-variate normal distribution with mean vector $\mu$ and covariance matrix $A$), no matter how large $n$ is. In this case it is not at all clear whether $\hat{\phi}$ converges in probability to $\phi$ and whether the limiting distribution of $\hat{\phi}$ is multivariate normal.

A general result pertaining to the consistency and asymptotic multivariate normality of a vector of ML estimators was established by Sweeting (1980). Conditions under which Sweeting's result is applicable to the ML estimators of RFIM parameters were derived by Mardia and Marshall (1984, Theorems 1 and 2) and are stated in the following theorem.

Let $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ represent the eigenvalues of $V$, let $\lambda_k^i$ ($k = 1, ..., n$), where $|\lambda_1^i| \leq ... \leq |\lambda_n^i|$, represent the eigenvalues of
\( \partial V / \partial \theta_i \) (i = 1, \ldots, m), and let \( \lambda^i_k \) (k = 1, \ldots, n), where \( |\lambda^i_1| \leq \ldots \leq |\lambda^i_n| \), represent the eigenvalues of \( \partial^2 V / \partial \theta_i \partial \theta_j \) (i = 1, \ldots, m; j = 1, \ldots, m). Note that the eigenvalues of \( V \) are positive. Define \( q_{ij} = \text{tr}[ (V^{-1} \partial V / \partial \theta_i) (V^{-1} \partial V / \partial \theta_j) ] \). As in Chapter 2, let \( B \) denote the information matrix associated with the full log-likelihood function \( L \). Let \( \| \cdot \| \) denote the Euclidean matrix norm, i.e., \( \| A \| = [\text{tr}(A A)]^{1/2} \) for any matrix \( A \).

**Theorem 4.1**

In the RFLM, suppose that \( y \) has a multivariate normal distribution, and assume (without loss of generality) that \( X \) has full column rank. Suppose further that the following conditions are satisfied:

(i) \( \theta \in \Theta_2; \)

(ii) \( \lim_{n \to \infty} \lambda_n < \infty, \lim_{n \to \infty} |\lambda^i_n| < \infty, \) and \( \lim_{n \to \infty} |\lambda^i_j| < \infty \) (i = 1, \ldots, m; j = 1, \ldots, m);

(iii) \( \| \partial V / \partial \theta_i \|^{-2} = o(n^{-1/2} - \delta) \) for some \( \delta > 0 \) (i = 1, \ldots, m);

(iv) \( a_{ij} = \lim_{n \to \infty} [q_{ij} / q_{ii} q_{jj}^{1/2}] \) exists (i = 1, \ldots, m, j = 1, \ldots, m)

and \( A = \{ a_{ij} \} \) is a nonsingular matrix;

(v) \( \lim_{n \to \infty} (X' X)^{-1} = 0. \)

Then, \( B^{-1/2}(\hat{\theta} - \theta) \) has a limiting \((p+m)\)-variate normal distribution with mean vector 0 and covariance matrix \( I \).

In general, it is difficult to determine whether the conditions of Theorem 4.1 are satisfied by an RFLM. The next theorem gives readily verifiable conditions that imply conditions (i)-(iii) of Theorem 4.1 when a single realization of a jointly normally-distributed weakly stationary random field is observed at sites having Configuration I or Configuration II.

**Theorem 4.2**

Suppose, under RFLM-R, that $\mathcal{X}_Y$ is a jointly normally-distributed weakly stationary random field with covariogram $\hat{C}(\cdot; \vec{\theta})$, and suppose that $\mathcal{X}_Y$ is observed at sites have Configuration II. If $\hat{C}(\cdot; \vec{\theta})$, $\frac{\partial^2}{\partial \vec{\theta}_i \partial \vec{\theta}_j} \hat{C}(\cdot; \vec{\theta})$ (i = 1, ..., m; j = 1, ..., m) are continuous and absolutely integrable for all $\vec{\theta} \in \Theta$, then conditions (i)-(iii) of Theorem 4.1 are satisfied.

Proof:

The continuity of $\frac{\partial^2}{\partial \vec{\theta}_i \partial \vec{\theta}_j} \hat{C}(\cdot; \vec{\theta})$ (i = 1, ..., m; j = 1, ..., m) for all $\vec{\theta} \in \Theta$ is equivalent to condition (i) of Theorem 4.1 by the definition of $\Theta_2$.

Now, consider condition (ii) of Theorem 4.1. Let $\vec{z} = (i-k, j-l)'$ represent the vector of "lags" between the centroids of $S_{ij}$ and $S_{k\ell}$.
(i = 1, ..., R; j = 1, ..., R; k = 1, ..., C; \ell = 1, ..., C), and let
N_k represent the number of times that lag k occurs in Configuration II.

Some thought reveals that

\[
N_k = \begin{cases} 
(C - |k_1|)(R - |k_2|), & \text{if } |k_1| < C \text{ and } |k_2| < R, \\
0, & \text{otherwise.}
\end{cases}
\]  

(4.1)

Further, let \(k_k = \{h: h = s-t, \text{ where } s \in S_i, t \in S_k, \text{ and } (i-k, j-k) = k\} \). Now, the elements of V are given by integrals of the form

\[
\int_{S_i} \int_{S_k} \hat{C}(s-t; \theta) ds dt.
\]

After a change of variable, such integrals can be alternatively represented

as

\[
\int_{K_k} \hat{C}(h; \theta) g_k(h) dh
\]

for some bounded integrable function \(g_k(\cdot)\) of d

variables. Thus, the row-sum norm of V, i.e., \(\max |\sum_{a=1}^n v_{ab}|\), is

less than or equal to \(\int_{IR} \hat{C}(h; \theta) g(h) dh\) for all R and C, where \(g(h) = \sup_{k} g_k(h)\) is a bounded function. Moreover, the spectral norm of V, i.e., \(\lambda_n\), is less than or equal to the row-sum norm of V (by Theorem 5.6.7 of Graybill, 1983), so that

\[
\lim_{n \to \infty} \lambda_n < \lim_{n \to \infty} \int_{IR} \hat{C}(h; \theta) g(h) dh < \infty
\]

by the absolute integrability of \(\hat{C}(\cdot; \theta)\). Similar arguments show that

\[
\lim_{n \to \infty} \lambda_i^n < \infty \lim_{n \to \infty} \lambda_i^{ij} < \infty.
\]

Hence, condition (ii) is satisfied.
Finally, consider condition (iii). From the symmetry of $\frac{\partial V}{\partial \theta_i}$, $\| \frac{\partial V}{\partial \theta_i} \|^2$ is equal to the sum of squares of the elements of $\frac{\partial V}{\partial \theta_i}$, i.e.,

$$
\| \frac{\partial V}{\partial \theta_i} \|^2 = \Sigma \Sigma \Sigma [ \int_{s_{ab}} \int_{s_{kl}} \frac{\partial}{\partial \theta_i} \hat{c}(s-t;\theta) ds dt ]^2.
$$

Equivalently, we have that

$$
\| \frac{\partial V}{\partial \theta_i} \|^2 = \Sigma \Sigma [ \int_{k_{k}} \int_{k_{k}} \frac{\partial}{\partial \theta_i} \hat{c}(h;\theta) g_k(h) dh ]^2
$$

\[=\Sigma \Sigma (C_1-|k_1|)(R-|k_2|) \int_{k_{k}} \int_{k_{k}} \frac{\partial}{\partial \theta_i} \hat{c}(h;\theta) g_k(h) dh ]^2,

where for the last equality we have used (4.1). Now, as $n \to \infty$ at least one of the numbers $R$ or $C$ approaches infinity as well. Without loss of generality, assume that $C \to \infty$, but possibly that $R \leq M$ for all $n$, where $M < \infty$. Then,

$$
\lim_{n \to \infty} \frac{1}{n} \| \frac{\partial V}{\partial \theta_i} \|^2 = \lim_{n \to \infty} \frac{C-1}{C} \Sigma (1-\frac{|k_1|}{C}) \Sigma (1-\frac{|k_2|}{R})
$$

\[< \lim_{C \to \infty} \frac{C-1}{C} \Sigma \Sigma |k_1| M \int_{k_{k}} \int_{k_{k}} \frac{\partial}{\partial \theta_i} \hat{c}(h;\theta) g_k(h) dh ]^2.
\]
Applying Kronecker's lemma (Fuller, 1976, p. 109) gives

\[
\lim_{n \to \infty} \frac{1}{n} \left\| \frac{\partial V}{\partial \Theta_i} \right\|^2 < \Sigma_{k=1}^\infty \Sigma_{k_2=-M}^M \int_{K_k} \frac{\partial}{\partial \Theta_i} \hat{C}(h; \theta) g_k(h) dh \] ^2,
\]

provided that \( \int_{K_k} \frac{\partial}{\partial \Theta_i} \hat{C}(h; \theta) g_k(h) dh \) is summable (over \( k \)). But, the absolute integrability of \( \int_{K_k} \frac{\partial}{\partial \Theta_i} \hat{C}(h; \theta) g_k(h) dh \) implies that

\[
\int_{K_k} \frac{\partial}{\partial \Theta_i} \hat{C}(h; \theta) g_k(h) dh
\]

is absolutely summable which in turn, implies that \( \int_{K_k} \frac{\partial}{\partial \Theta_i} \hat{C}(h; \theta) g_k(h) dh \) is summable. Hence, \( \lim_{n \to \infty} \frac{1}{n} \left\| \frac{\partial V}{\partial \Theta_i} \right\|^2 < \infty \), implying that \( \frac{1}{n} \left\| \frac{\partial V}{\partial \Theta_i} \right\|^2 = o(1) \), i.e., \( \left\| \frac{\partial V}{\partial \Theta_i} \right\| ^2 = o(n^{-1}) \). Thus, condition (iii) is satisfied with \( \delta = 1/2 \).

Q.E.D.

**Corollary 1**

Suppose, under RFLM-R, that all the conditions of Theorem 4.2 are satisfied, but that \( \mathcal{F}_X \) is observed at sites having Configuration I. Then, conditions (i)-(iii) of Theorem 4.1 are satisfied.
Proof:

The result follows immediately since Configuration I is a special case of Configuration II. Q.E.D.

The following corollary establishes the same result as Theorem 4.2 in the context of RFLM-P, and can be proved using arguments very similar to those used to prove Theorem 4.2.

**Corollary 2**

Under RFLM-P, if the conditions of Theorem 4.2 are satisfied, then conditions (i)-(iii) of Theorem 4.1 are satisfied.

Theorem 4.2 is essentially an extension of Theorem 3 of Mardia and Marshall (1984) from RFLM-P to RFLM-R (so that Corollary 2 is nearly equivalent to their theorem), except that the conditions of Theorem 4.2 are slightly less restrictive than those of Mardia and Marshall, who require that $R_{\infty}$ and $C_{\infty}$.

Which covariograms of weakly stationary random fields satisfy the conditions of Theorem 4.2? Clearly, any covariogram whose range exists, such as the spherical covariogram (1.3), is absolutely integrable. However, the second derivative of the spherical covariogram with respect to $\theta_2$ is not continuous at $\theta_2 = r$. The first- and second-order partial derivatives of the isotropic spherical covariogram with respect to $\theta_2$ are
Note that (4.2) is continuous everywhere but that (4.3) is discontinuous
at $\theta_2 = r$. [There has been some confusion regarding the continuity or
discontinuity of (4.2) and (4.3): since (4.2) is continuous, the
spherical covariogram is differentiable with respect to its parameters,
contrary to the assertion of Cressie (1985), while Mardia and Marshall
(1984, Section 5) erred by assuming that (4.3) is a continuous function.]
Thus, the conditions of Theorem 4.2 are not satisfied by the spherical
covariogram.

Two other covariograms that do not satisfy the conditions of
Theorem 4.2 (in this case, because they are not absolutely integrable)
are the oscillatory covariogram (1.4) and the covariogram

$$C(r; \theta_1, \theta_2) = \theta_1 \frac{1 - \exp(-\theta_2 r)}{r}$$

(where $\theta_1 > 0$ and $\theta_2 > 0$) introduced by Whittle (1962), which behaves
like a constant multiple of \( 1/r \) for large \( r \). However, the \( L^2 \)-
isotropic exponential and Gaussian covariograms (1.1) and (1.2) and
their \( L^1 \)-isotropic and geometrically anisotropic analogs defined in
Chapter 3 do satisfy the conditions, as does covariogram (1.6), which
behaves like a constant multiple of \( r^{1/2} \exp(-\theta r^2) \) for large \( r \).

In summary, under the conditions of Theorem 4.2 and subject to
conditions (iv) and (v) of Theorem 4.1, \( \hat{\phi} \) is consistent and
asymptotically normal with limiting covariance matrix \( B^{-1} \). It is very
easy to determine whether the conditions of Theorem 4.2 are satisfied.
However, conditions (iv) and (v) of Theorem 4.1 are not easily verified,
except in special cases.

Consider the following special case of a treatment-additive RFLM-P
(1.15) for which condition (v) can be easily verified. Suppose that \( d = 2 \)
and that the observational sites have Configuration II. Suppose further
that the mean function of \( \mathcal{Y} \) is

\[
\sum_{i=1}^{q} \alpha_i f_i(s) + \sum_{j=1}^{t} \tau_j g_j(s),
\]

where \( \sum_{i=1}^{q} \alpha_i f_i(s) \) is a first-order polynomial of the elements of \( g \) (i.e.,
a plane), and \( \{\alpha_i\} \), \( \{\tau_j\} \), and \( \{g_j(*)\} \) are defined as in Section 1.3.
Consider a sequence of layouts in which each treatment occurs the same
number of times in each row and column. Such a sequence of layouts can
easily be constructed by adding \( t \) rows and/or \( t \) columns to an initial layout in which each treatment occurs the same number of times in each row and column. For such a sequence, it can be shown that \((XX)^{-1}\)

converges to a null matrix as either \( R \) or \( C \) (or both) get large.

Consider now condition (iv) of Theorem 4.1. One RFIM-P for which (iv) can be verified is that for which \( d = 2 \), the observational sites have Configuration II, and the covariogram of \( \mathcal{F} \) is the \( L \)-isotropic exponential covariogram \( \hat{C}(x,y;\theta) = \theta_1 \exp(-\theta_2 (|x|+|y|)) \), where \( \theta_1 > 0 \) and \( \theta_2 > 0 \). Let \( n_1 \) and \( n_2 \) be alternative representations of \( C \) and \( R \), respectively. Due to the separability of this random field, \( V = P_1 \otimes P_2 \), where

\[
P_i = \sqrt{\theta_1} \begin{bmatrix}
1 & \rho_i & \rho_i^2 & \ldots & \rho_i^{n_i-1} \\
1 & \rho_i & & & \\
& \rho_i & \ddots & & \\
& & \ddots & \rho_i & \\
\text{symm} & & & 1
\end{bmatrix},
\]

and where \( \rho_i = \exp(-\theta_2 H_i) \). Consequently, \( V^{-1} = P_1^{-1} \otimes P_2^{-1} \), where by (3.8),

\[
P_i^{-1} = \frac{1}{\sqrt{\theta_1 (1-\rho_i)^2}} \begin{bmatrix}
1 & -\rho_i & -\rho_i & 1+\rho_i^2 \\
-\rho_i & 1 & 1+\rho_i^2 & \ddots \\
& 1+\rho_i^2 & \ddots & \ddots \\
& & \ddots & 1+\rho_i^2 & -\rho_i \\
& & & -\rho_i & 1
\end{bmatrix}.
\]
Furthermore, \( \frac{\partial V}{\partial \theta_2} = (1/\theta_1)V \), \( \frac{\partial \theta_1}{\partial \theta_2} \) is a symmetric \( n_1 \times n_1 \) Toeplitz matrix whose first row is \((0, -\rho_1, -2\rho_1, \ldots, -(n_1-1)\rho_1)\), and

\[
P_i^{-1} \frac{\partial \theta_1}{\partial \theta_2} = \begin{pmatrix}
-\rho_1 & \rho_1^2 & \cdots & \rho_1^{n_1-1} \\
0 & -\rho_1 & \cdots & \rho_1^{n_1-1} \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & -\rho_1 \\
-\rho_1 & \cdots & \cdots & \cdots & \cdots \\
\end{pmatrix},
\]

where \( c_{ij} = (1-\rho_2^2)\rho_1^i (j = 1, \ldots, n_1-2; i = 1, 2) \). Consequently,

\[
q_{11} = \text{tr}(V^{-1} \frac{\partial V}{\partial \theta_1} V^{-1} \frac{\partial V}{\partial \theta_1}) = \frac{n_2}{\theta_1},
\]

and

\[
q_{12} = \frac{1}{\theta_1} \text{tr}(V^{-1} \frac{\partial V}{\partial \theta_2}) = \frac{1}{\theta_1} \text{tr}(P_i^{-1} \frac{\partial \theta_1}{\partial \theta_2}) \text{tr}(P_i^{-1} \frac{\partial \theta_1}{\partial \theta_2}).
\]

\[
q_{12} = \frac{4\rho_1^2(n_1-1)\rho_2^2(n_2-1)}{\theta_1(1-\rho_1^2)(1-\rho_2^2)}.
\]

Now, let \( z_{jj} \) denote the \((j,j)\)th element of \( P_i^{-1} \frac{\partial \theta_1}{\partial \theta_2} P_i^{-1} \frac{\partial \theta_1}{\partial \theta_2} \). We have that
\[ z_{11} = \left( \frac{1}{1-\rho_1^2} \right)^2 (\rho_1^4 + \rho_1 c_{i1} + \rho_1^2 c_{i12} + \ldots + \rho_1 \sum_{i,n_1-2}^n + \rho_1 \sum_{i,n_1-2}^n + \rho_1 \sum_{i,n_1-2}^n) , \]

\[ z_{22} = \left( \frac{1}{1-\rho_1^2} \right)^2 (\rho_1^4 + \rho_1 c_{i1} + \rho_1^2 c_{i12} + \ldots + \rho_1 \sum_{i,n_1-3}^n + \rho_1 \sum_{i,n_1-3}^n + \rho_1 \sum_{i,n_1-3}^n) , \]

\[ z_{33} = \left( \frac{1}{1-\rho_1^2} \right)^2 (\rho_1^4 + \rho_1 c_{i12} + \rho_1^2 c_{i11} + \ldots + \rho_1 \sum_{i,n_1-4}^n + \rho_1 \sum_{i,n_1-4}^n + \rho_1 \sum_{i,n_1-4}^n) , \]

\[ \vdots \]

\[ z_{n-1,n-1} = \left( \frac{1}{1-\rho_1^2} \right)^2 (\rho_1^4 + \rho_1 c_{i,n_1-2} + \rho_1^2 c_{i,n_1-3} + \ldots + \rho_1 \sum_{i,n_1-4}^n + \rho_1 \sum_{i,n_1-4}^n + \rho_1 \sum_{i,n_1-4}^n) , \]

so that

\[ \text{tr}(p_1^2 \frac{3p_1}{\partial \theta_1} \frac{3p_1}{\partial \theta_2} \frac{3p_1}{\partial \theta_1} \frac{3p_1}{\partial \theta_2}) = \left( \frac{1}{1-\rho_1^2} \right)^2 [(4n_1-6)\rho_1^4 + 2\rho_1^2 + 2m_1-2 \sum_{j=1}^{n_1-2} \rho_1^{ij} c_{ij}^2 \sum_{j=1}^{n_1-3} \rho_1^{ij} c_{ij}^2 \sum_{j=1}^{n_1-4} \rho_1^{ij} c_{ij}^4] \]

\[ + 2 \sum_{j=1}^{n_1-3} (n_1-j-2) c_{ij}^2 \]

\[ = \left( \frac{1}{1-\rho_1^2} \right)^2 [(4n_1-6)\rho_1^4 + 2\rho_1^2 + 2m_1-2 \sum_{j=1}^{n_1-2} \rho_1^{ij} c_{ij}^2 \sum_{j=1}^{n_1-3} \rho_1^{ij} c_{ij}^2 \sum_{j=1}^{n_1-4} \rho_1^{ij} c_{ij}^4] \]

\[ + 2(1-\rho_1^2) \sum_{j=1}^{n_1-2} (n_1-j-2) c_{ij}^2] \]
Thus, by Kronecker's lemma

\[
\lim_{n_1 \to \infty} \frac{1}{n_1} \text{tr}(p_1^{-1} \frac{\partial p}{\partial x_1} p_1^{-1} \frac{\partial p}{\partial x_2}) = \left(\frac{1}{1-\rho_1^2}\right)^2 \left[4\rho_1^4 + 2(1-\rho_1^2)^2 \sum_{j=1}^{\infty} \rho_j^2 \right]
\]

\[
= 2\left(\frac{1}{1-\rho_1^2}\right)^2 \rho_1^2 (1 + \rho_1^2).
\]

Hence,

\[
\lim_{n_1, n_2 \to \infty} \frac{q_{12}}{q_{11}q_{22}} = \frac{\lim_{n_1, n_2 \to \infty} \{16(n_1-1)\rho_1^4(n_2-1)\rho_2^4/n_1n_2(1-\rho_1^2)^2 (1-\rho_2^2)^2\}}{4\left(\frac{1}{1-\rho_1^2}\right)^2 \rho_1^2 (1+\rho_1^2) \left(\frac{1}{1-\rho_2^2}\right)^2 \rho_2^2 (1+\rho_2^2)}
\]

\[
= \frac{4\rho_1^2 \rho_2^2}{(1+\rho_1^2)(1+\rho_2^2)}.
\]

Thus, the limit exists for all \( \theta \) and \( \Lambda \) is singular iff \( 4\rho_1^2 \rho_2^2 / [(1+\rho_1^2)(1+\rho_2^2)] = 1 \), which is easily shown to be impossible for \( \theta \in \Theta \). Therefore, (iv) is satisfied.

Combining the results of the previous two examples, we see that a two-dimensional treatment-additive RFLM-P in which the observational sites have Configuration II, the mean function of \( F_Y \) is that given by (4.4), and the covariogram of \( F_Y \) is the \( l^1 \)-isotropic exponential covariogram satisfies conditions (iv) and (v) of Theorem 4.1 and satisfies the conditions of Theorem 4.2, provided that any given treatment occurs with equal frequency in each row and in each column.
5. ALTERNATIVE APPROACHES TO ANALYZING SPATIAL EXPERIMENTS

The random field approach is by no means the first nonclassical approach ever proposed for accounting for the effects of spatial correlation in spatial experiments. In this chapter, we review several other approaches, and compare them to each other, to the random field approach, and to classical blocking methodology. In Section 5.1, we introduce some terminology and notation that will prove useful. In Section 5.2, we describe the essential features of each approach, showing how the method of analysis associated with each is appropriate for a corresponding member of a particular class of linear models. Relationships among different approaches are discussed in Section 5.3. In Section 5.4, we demonstrate some properties of the various estimators of treatment effects and provide sufficient conditions for those estimators to be unbiased. Finally, in Section 5.5 we compare the results produced by various approaches when applied to uniformity trial data.

5.1. Terminology and Notation

All of the methods pertaining to the design and analysis of spatial experiments described herein have been developed in the context of agricultural field experimentation. Subsequently, we adopt the terminology customarily used in conjunction with such experimentation. Thus, the experimental material (which is a region in two-dimensional space) is a "field," experimental units are called "plots," and responses are referred to as "yields." Variation among plots is attributed to differences in "fertility."
Each method to be reviewed in this chapter is essentially an application of generalized least squares to one of various special cases of the linear model

\[ G\hat{y} = GA\hat{a} + GT\gamma + Ge, \quad (5.1) \]

where \( \hat{y}, \hat{a}, T, \gamma, \) and \( e \) are defined as in the treatment-additive RFLM (1.15), \( G \) is an \( g \times n \) matrix whose elements are fixed and known, \( A \) is an \( n \times q \) matrix whose elements may or may not be fixed and known, \( E(Ge) = 0 \), and \( \text{Var}(Ge) = V \). The elements of \( V \) are assumed to be known only up to functions of an \( m \times l \) parameter vector \( \theta \in \Theta \), where \( \Theta \) is the subset of \( \mathbb{R}^m \) for which \( V \) is positive definite. It is assumed that \( n > q + t \). Note that the class (1.15) of treatment-additive RFLMs is a subclass of the class (5.1) of models for which \( g = n, G = I, \) and \( A = X_{\alpha}^\prime \).

It is important to distinguish between a model belonging to the class of models (5.1), whose sole purpose is to suggest a method of analysis, and the model believed to actually govern the plot yields of a particular spatial experiment. The former shall be referred to as an analysis-oriented model, and the latter we shall call the true model. Subsequently, we assume that the true model is

\[ \hat{y} = T\gamma + \varepsilon, \quad (5.2) \]
where $y, X, T, \alpha$, and $\tau$ are defined as in model (5.1), and $\epsilon$ is an $n \times 1$ random vector such that $E(\epsilon) = 0$ and $\text{Var}(\epsilon) = \Sigma$, where $\Sigma$ is an unknown matrix.

Let $\lambda$ denote a $t \times 1$ vector whose elements $\lambda_1, \ldots, \lambda_t$ satisfy
\[
\sum_{i=1}^t \lambda_i = 0.
\]
Often, the estimation of functions $\lambda^T \tau$, known as treatment contrasts, is of primary importance to the investigator conducting a spatial experiment. Let $\Lambda^T \tau$ denote a vector whose elements are treatment contrasts.

For several of the approaches to be described, the best linear unbiased estimator (BLUE) of an estimable treatment contrast $\lambda^T \tau$ under the analysis-oriented model associated with the approach is given by
\[
\lambda^T \tau, \quad \text{where } \lambda^T \tau \text{ is any solution to the equations}
\]
\[
T' (I - A(A'A)^{-1} A') \tau = T' (I - A(A'A)^{-1} A') y. \tag{5.3}
\]

Furthermore, the covariance matrix of the vector of BLUEs of the elements of $\Lambda^T \tau$ (assuming that those elements are estimable functions) is
\[
\theta_1 \Lambda^T [T' (I - A(A'A)^{-1} A') T]^{-1} \Lambda, \tag{5.4}
\]
which can be estimated unbiasedly by estimating $\theta_1$ by
\[
[y' (I - P_{AT}) y]/u, \tag{5.5}
\]
where $u$ is a scalar whose value depends on the approach.
Most of the methods to be discussed are flexible enough to be used in conjunction with a wide variety of spatial configurations and blocking schemes. Some methods, however, in the form in which they have been put forward, can only be used in conjunction with Configurations I or II, equireplicated designs, and particular blocking schemes, and it is unclear precisely how to extend these methods to a more general setting. Moreover, close relationships among some of the approaches have been demonstrated only in the context of certain blocking schemes. Consequently, we shall find it convenient to discuss some methods in the context of three particular blocking schemes associated with equireplicated designs. Subsequently, let \( r \) denote the number of times a treatment is replicated in an equireplicated design. Blocking Scheme I refers to the scheme in which \( r \) blocks of \( t \) contiguous plots lie end-to-end in a field whose plots are laid out according to Configuration I (see Figure 5.1). Blocking Scheme II is applicable to a field whose plots are laid out according to Configuration II and is a scheme whereby blocks lie side-by-side and consist of a single row (or column) of \( t \) plots. In some settings it may be important to distinguish the scheme in which rows serve as blocks from the scheme in which columns serve as blocks. In those settings, if rows serve as blocks the scheme is referred to as Blocking Scheme IIr;
Figure 5.1. Blocking Scheme I

Figure 5.2. Blocking Scheme (a), (b) Blocking Scheme IIc
if columns serve as blocks the scheme is referred to as Blocking Scheme IIc (see Figure 5.2). Necessarily, \( R = r \) and \( C = t \) in the former and \( R = t \) and \( C = r \) in the latter.

We shall frequently refer to certain "incidence" matrices. If an experiment is carried out in \( q \) blocks, the block incidence matrix is defined as the \( nxq \) matrix whose \((i,j)\)th element is equal to one if plot \( i \) belongs to block \( j \), but is otherwise equal to zero. If plots are laid out according to Configuration II, row and column incidence matrices can be defined. A row incidence matrix is an \( nxR \) matrix whose \((i,j)\)th element is equal to one if plot \( i \) lies in row \( j \), but is equal to zero otherwise. Similarly, a column incidence matrix is an \( nxC \) matrix whose \((i,j)\)th element is equal to one if plot \( i \) lies in column \( j \), but is equal to zero otherwise.

5.2. Description of Alternative Approaches

5.2.1. Classical approach: blocking and randomization

The classical approach taken to minimize the effects of spatial correlation in spatial experiments consists of blocking and randomization, i.e., dividing the field into groups of plots in such a way that plots within a block are highly correlated, and then randomly assigning treatments to plots within blocks. In some cases, blocks of homogeneous units of size \( t \) can be formed, so that a complete block design can be employed. However, it may not always be possible to form blocks of \( t \) homogeneous plots (particularly if \( t \) is large), in which case it is
necessary to resort to an incomplete block design. In either case, the blocking is carried out in the hope that the analysis of variance (ANOVA) corresponding to the block design will be more sensitive in detecting true treatment differences than the ANOVA corresponding to a completely randomized design in which blocking is not utilized.

It is possible for blocking to decrease, rather than increase, the sensitivity of the statistical analysis. This could happen, for example, in an experiment having Configuration II, if a large fertility gradient exists along rows but not along columns, and plots are assigned to blocks according to Blocking Scheme III.

As discussed in Chapter 1, for a block design it is the incorporation of the proper randomization scheme that justifies the use of the F-test associated with the ANOVA to detect treatment differences. However, the classical ANOVA corresponding to a randomized complete block design, hereafter referred to as Method RCB, is not necessarily the most efficient analysis of the experimental data even if the blocks are well chosen. Similarly, the usual intra-interblock analysis associated with a randomized incomplete block design, though it may often improve upon Method RCB, may be less efficient than other methods of analysis. The approaches described in the remainder of Section 5.2 represent attempts to improve upon Method RCB and the intra-interblock analysis.

The analysis-oriented linear model associated with Method RCB is a special case of model (5.1) in which \( g = n \), \( G = I \), \( A = A_b \), where \( A_b \)
is a block incidence matrix, \( q \) is the number of blocks, and \( V = \theta_1 I \), where \( \theta_1 > 0 \). We refer to this analysis-oriented model as Model RCB.

All treatment contrasts are estimable under Model RCB, and the BLUE of \( \lambda \) is \( \lambda \) where \( \lambda \) is any solution to equations (5.3) with \( A = A_b \). Furthermore, the covariance matrix of \( \lambda \) is given by (5.4) with \( A = A_b \), and can be estimated unbiasedly by putting \( u = (q-1)(t-1) \) and \( A = A_b \) into (5.5) and (5.6), respectively.

5.2.2. Row-column analysis

We indicated previously that a poor choice of blocks can render Method RCB ineffective. When the spatial configuration of the plots is Configuration II, a row-column analysis (see Pearce, 1980 or Hinz, 1985) can be more efficient than Method RCB. Given \( t, R, \) and \( C \), the row-column analysis is uniquely defined, and we refer to it as Method RC. Method RC is merely the application of ordinary least squares to a special case of model (5.1) in which \( g = n, G = I, A = (A_r, A_c) \), where \( A_r \) and \( A_c \) are the \( nxR \) and \( nxC \) row and column incidence matrices, respectively, \( \alpha = (\alpha_r, \alpha_c)' \), where \( \alpha_r \) and \( \alpha_c \) are \( Rx1 \) and \( Cx1 \) vectors, respectively, and \( V = \theta_1 I \) (\( \theta_1 > 0 \)). We refer to this model as Model RC.

When treatment contrasts are estimable under Model RC, the BLUE of \( \lambda \) is \( \lambda \) where \( \lambda \) is any solution to equations (5.3) with \( A = (A_r, A_c) \). The covariance matrix of \( \lambda \) is given by (5.4) with \( A = (A_r, A_c) \), and can be estimated unbiasedly by putting \( u = RC-R-C-t+2 \) and \( A = (A_r, A_c) \) into (5.5) and (5.6), respectively.
5.2.3. **Ordinary covariance analysis**

One method of analysis which can be considerably more efficient than an ANOVA in many types of experiments is an analysis of covariance (ANOCOVA). In order for an ANOCOVA to result in a more efficient analysis than an ANOVA in the context of spatial experiments, the value of the concomitant variable(s) associated with a given plot should reflect the inherent fertility of that plot relative to other plots. Two fundamentally different types of concomitant variables have been suggested. In Section 5.2.4, concomitant variables which are functions of neighboring plot yields are considered. Here we discuss "ordinary" covariates, i.e., variables which are not direct functions of the plot yields themselves. Let $K$ denote the number of ordinary covariates to be used in the ANOCOVA.

Two rather natural candidates for an ordinary covariate in spatial experiments have been suggested. One procedure for obtaining a single covariate, suggested by Yates (1936), is to include systematically-placed control plots in the field. One could then take the covariate associated with a treated plot to be a weighted average of the yields of the control plots. If positive spatial correlation between neighboring plots is expected, it would seem appropriate to choose the weights to be inversely proportional to the distances of the control plots from the treated plot. This idea could be extended to $K > 1$ covariates by, for example, constructing weighted averages of control plots in each of several directions from a treated plot.
A second way to obtain covariates is to construct functions of plot position within the field. For example, functions \( f_k(*): k = 1, \ldots, K \), defined on \( \mathbb{R} \), could be regarded as covariates. If \( f_k(s) = s^k (k = 1, \ldots, K) \), the ANOCOVA is referred to as the polynomial covariance technique of order \( K \), which we abbreviate as Method PC-K. Federer and Schlottfeldt (1954) applied Method PC-2 to a spatial experiment whose plots were assigned to blocks according to Blocking Scheme II and found it to be considerably more efficient than Method RCB.

The analysis-oriented model associated with either of these covariate-based methods belongs to the subclass of class (5.1) for which \( g = n, G = I, A = (A_1, A_2) \) where \( A_1 \) is an \( nxK \) matrix whose \( i \)-th row is comprised of the covariate(s) associated with the \( i \)-th observation and \( A_2 \) is an \( nx(q-K) \) matrix, \( \alpha = (\alpha_1, \alpha_2) \) where \( \alpha_1 \) is a \( Kx1 \) vector of parameters associated with the covariates and \( \alpha_2 \) is a \( (q-K)x1 \) vector of parameters, and \( V = \theta_1 I (\theta_1 > 0) \). When treatment contrasts are estimable under a model from this subclass, the BLUE of \( \hat{\lambda} \) is any solution to equations (5.3) with \( A = (A_1, A_2) \), and the covariance matrix of \( \hat{\lambda} \) is given by (5.4), with \( A = (A_1, A_2) \). This covariance matrix can be estimated unbiasedly by putting \( u = n - \text{rank}(A_1, A_2) \) and \( A = (A_1, A_2) \) into (5.5) and (5.6), respectively.

Both of these covariate-based methods of analysis have their disadvantages. The first method necessarily increases the size (and hence the cost) of the experiment, while the second method will prove useful only when the pattern of fertility as one plot coordinate varies is fairly consistent over all values of the other plot coordinate. The random field approach does not suffer from these defects.
5.2.4. The Papadakis analysis

As noted earlier, there is typically a tendency in spatial experiments for a plot yield to be high (low) when yields on neighboring plots are also high (low). This tendency leads, in a natural way, to the following idea: the value of any particular treatment should be judged by how well the plots receiving that treatment perform relative to their neighbors. Papadakis (1937) devised a method of analysis, now bearing his name, which is based on this idea. Papadakis suggested that spatial experiments should be analyzed by an ANOCOVA in which the covariate is constructed as follows:

1. Compute the treatment averages.

2. Obtain the "residual" for each plot by subtracting the appropriate treatment average from the yield of the plot.

3. Take the covariate for a plot to be the average of the residuals of that plot's neighbors.

There are several different ways in which the neighbors of a plot can be designated. Perhaps the most common designation of a plot's neighbors are adjacent plots, i.e., plots which share a boundary of nonzero length with that plot. If we apply this designation, for example, to a field whose plots are laid out according to Configuration I, then plots on either end of the field have only one neighbor while the remaining plots have two neighbors. Many other designations are possible, of course. Subsequently, we refer to plots that do not have a full complement of neighbors as border plots, and those plots that do have a full complement of neighbors are referred to as internal plots.
Multiple Papadakis covariates could be constructed by forming averages of residuals over disjoint subsets of a plot's neighbors; for example, in a field whose plots are laid out according to Configuration II, separate covariates for row neighbors and column neighbors could be formed.

The Papadakis method for estimating treatment effects is quite flexible because it can be used in conjunction with any spatial configuration and any blocking scheme and no specific assumptions about the pattern of spatial heterogeneity are necessary (though the method may perform better for some patterns than for others). Moreover, it is relatively easy to implement because nothing more than standard ANOCOVA calculations, which can be carried out by any of a number of statistical computing packages, are involved.

It is quite apparent, owing to the dual use of each plot yield as the dependent variable and in constructing the covariate, that the F-test for detecting treatment differences associated with a Papadakis ANOCOVA may be inappropriate. However, Bartlett (1938, 1978) examined some of the theoretical aspects of a Papadakis analysis and judged the associated F-test to be appropriate if two degrees of freedom are subtracted for the covariate rather than the usual one. Wilkinson et al. (1983) discovered that even when two degrees of freedom are subtracted for the covariate, the F-test tends to be conservative.

There is a considerable body of empirical evidence attesting to the increased efficiency of a Papadakis analysis over the ANOVA
corresponding to a completely randomized design and over Method RCB (see, e.g., Pearce and Moore, 1976; Kempton and Howes, 1981). For some experiments the error mean square is reduced 50% or more by a Papadakis analysis; very seldom is it increased.

The three-step prescription for obtaining the Papadakis covariate(s) is ambiguous on one point, namely, how to form the covariates for the border plots. Papadakis' own solution was to compute average plot residuals over the neighbors which do exist. Another way of dealing with border plots is to imagine the experiment as being laid out on the perimeter of a circle (for one-dimensional layouts) or on the surface of a cylinder or torus (for two-dimensional layouts) so that the layout has no borders. This idea has many interesting theoretical and computational implications (see, e.g., Martin, 1982), but would be objectionable to most practitioners because, when neighboring plots are positively correlated, the least desirable substitute for the missing neighbors of a border plot are plots on the opposite border of the field. Yet another possibility would be to modify the design by adding untreated plots to the boundaries of the field in such a way that each treated plot has a full complement of neighbors. The inclusion of these untreated plots may be practical for one-dimensional layouts such as Configuration I and, perhaps, for those two-dimensional layouts in which neighbors are defined in only one dimension, but when neighbors are defined in two dimensions, too many additional plots are required. For example, if plots are laid out according to Configuration II, \( R = C = 4 \), and adjacent plots are designated as neighbors, the size of the experiment
must be doubled if a full complement of neighbors is to be provided for each treated plot. Thus, when neighbors are defined in two dimensions, this approach is generally too wasteful of experimental material to be practical. Moreover, it requires that the experimenter consider, prior to conducting an experiment, the possibility of performing a Papadakis analysis.

To illustrate certain aspects of a Papadakis analysis, we shall restrict our attention to a single-covariate Papadakis analysis that ignores any blocking that may have been incorporated into the design. Subsequently, we refer to this particular Papadakis analysis as Method PAP. The corresponding model, referred to as Model PAP, is a special case of model (5.1) in which \( g = n, G = I, A = \tilde{z} \) where \( \tilde{z} \) is an \( n \times 1 \) vector of Papadakis covariates, \( g \) is a \( 1 \times 1 \) vector which we represent by \( \phi \), and \( V = \theta_1 I \ (\theta_1 > 0) \). Note that each element of \( \tilde{z} \) is estimable under model PAP provided that \( \tilde{z} \) does not lie in the column space of \( \tilde{T} \). Here,

\[
\tilde{z} = \frac{1}{2} N(\tilde{z} - \tilde{\tau}_{\text{OLS}}), \quad \tilde{\tau}_{\text{OLS}} = (T' T)^{-1} T' \tilde{y}
\]

is the ordinary least squares estimator of \( \tilde{\tau} \), and \( N = \{n_{ij}\} \) is the \( n \times n \) neighbor incidence matrix whose elements satisfy

\[
\begin{align*}
    n_{ij} &= \begin{cases} 
        1, & \text{if plots } i \text{ and } j \text{ are neighbors,} \\
        0, & \text{otherwise,}
    \end{cases} 
\end{align*}
\]

with possible modifications to account for how border plots are handled. Note that \( \tilde{z} = \frac{1}{2} N(I - P_T) \tilde{\tau} \) where \( P_T = T(T' T)^{-1} T' \) and that \( N \) is symmetric.
Model PAP differs from the usual model underlying an ANOCOVA in that the vector of covariates is an exact linear transformation of the vector of observations. In performing the Papadakis analysis this distinction is ignored and an ordinary ANOCOVA is carried out. Consequently, the Papadakis estimator of treatment effects is

\[ \hat{\tau}_{PAP} = [T'(I-P_z)T]^{-1}T'(I-P_z)y, \quad (5.7) \]

where

\[ P_z = z(z'z)^{-1}z' \]

\[ = [y'(I-P_T)N(I-P_T)y]^{-1}N(I-P_T)y (I-P_T)N, \quad (5.8) \]

provided that \( z \neq 0 \) and that \( T'(I-P_z)T \) is nonsingular. Expression (5.7) is the form of the Papadakis estimator of treatment effects given by Martin (1982). Alternatively, the Papadakis estimator can be expressed in the form

\[ \hat{\tau}_{PAP} = (T'T)^{-1}T(y-\hat{\phi}z), \quad (5.9) \]

where

\[ \hat{\phi} = [z'(I-P_T)z]^{-1}z'(I-P_T)y \]

\[ = 2[y'(I-P_T)N(I-P_T)y]^{-1}y'(I-P_T)y. \quad (5.10) \]

When treatments are replicated an equal number of times \( r \), expressions (5.9) and (5.10) reduce to
\[ \hat{\tau}_{PAP} = \frac{1}{r} \left( y - \hat{\phi} z \right) \]  

and

\[ \hat{\phi} = 2 \left[ y' \left( I - \frac{1}{r} TT' \right) N ( I - \frac{1}{r} TT' ) N ( I - \frac{1}{r} TT' ) y \right]^{-1} \]

\[ \cdot y' \left( I - \frac{1}{r} TT' \right) N ( I - \frac{1}{r} TT' ) y. \]

Expression (5.11) and a formula for \( \hat{\phi} \) when treatments are equireplicated were given previously by Draper and Faraggi (1985); however, their formula for \( \hat{\phi} \) is incorrect.

It is of interest to seek sufficient conditions for \( z \) to be nonnull and for \( T (I - P_{z})T \) to be nonsingular; the latter property is necessary and sufficient for all elements of \( z \) to be estimable under model PAP. Such conditions have not yet been investigated. The following series of lemmas establishes such conditions in the context of a field whose plots are laid out according to Configuration I and in which adjacent plots are designated as neighbors.

Two neighbor incidence matrices \( N_1 \) and \( N_2 \), corresponding to two different border plot adjustment schemes, are considered. The elements of \( N_1 \) correspond to the scheme wherein covariates for border plots are obtained by averaging over those neighbors which do exist; consequently, the \((1,2)\) and \((n,n-1)\) elements of \( N_1 \) equal two, the remaining elements on the first superdiagonal and the first subdiagonal equal one, and all remaining elements equal zero. The elements of \( N_2 \) correspond to the scheme wherein the strip of plots is regarded as lying on the perimeter of a
circle; hence, the \((1,n)\) and \((n,1)\) elements of \(N_2\), as well as all elements on the first super- and sub-diagonals, equal one, and the remaining elements equal zero. Note that \(N_2\) satisfies the definition of a symmetric regular circulant (see Graybill, 1983, p. 241).

**Lemma 5.1**

(i) If \(n\) is even, \(\text{rank}\{N_1(I-P_T)\} = n-t\); if \(n\) is odd, \(\text{rank}\{N_1(I-P_T)\} > n-t-1\).

(ii) If \(n \equiv 0 \mod 4\), \(\text{rank}\{N_2(I-P_T)\} \geq n-t-2\); otherwise, \(\text{rank}\{N_2(I-P_T)\} = n-t\).

**Proof of (i):**

A vector \(a = \{a_i\}\) belonging to \(\mathbb{N}(N_1)\) must satisfy the conditions

\[
\begin{align*}
a_2 &= 0, \\
a_1 + a_3 &= 0, \\
a_2 + a_4 &= 0, \\
&\vdots \\
a_{n-2} + a_n &= 0, \text{ and} \\
a_{n-1} &= 0.
\end{align*}
\]  

Conditions (5.12) can be re-expressed as

\[
a_i = -a_{i+2} \quad (i=1, \ldots, n-2),
\]  

\[(5.13a)\]
a_n-1 = 0 \quad (5.13c)

Conditions (5.13a) and (5.13b) imply that for any n, a_i = 0 for all even i. Further implications of conditions (5.13a)-(5.13c) depend on whether n is even or odd.

If n is even, then n-1 is odd, so that conditions (5.13a) and (5.13c) imply that a_i = 0 for all odd i. Thus, a_i = 0 for all i, implying that N_1 is nonsingular. The result for even n follows since rank(I-P_1) = n-t.

If n is odd, then n-1 is even, so that condition (5.13c) puts no restrictions on the elements of a beyond those already imposed by conditions (5.13a) and (5.13b). It is easy to show, using condition (5.13a), that N_1 has dimensionality equal to one. By Corollary 6.2 of Marsaglia and Styan (1974),

\[ \text{rank}\{N_1(I-P_1)\} = \text{rank}(I-P_1) - \dim \{C(I-P_1) \cap N(N_1)\} \]

\[ \geq n-t - \dim[N(N_1)] \]

\[ = n-t-1. \quad \text{Q.E.D.} \]

Proof of (ii):

A vector \( \tilde{a} = \{a_i\} \) belonging to \( \mathbb{N}(N_2) \) must satisfy the conditions
\[ a_n + a_2 = 0, \]
\[ a_1 + a_3 = 0, \]
\[ a_2 + a_4 = 0, \]
\[ \vdots \]
\[ a_{n-3} + a_{n-1} = 0, \]
\[ a_{n-2} + a_n = 0, \text{ and} \]
\[ a_{n-1} + a_1 = 0. \]

Conditions (5.14) can be re-expressed as

\[ a_i = -a_j \text{ if } i \mod 4 = (j+2) \mod 4, \]  \hspace{1cm} (5.15a)

\[ a_i = a_j \text{ if } i \mod 4 = j \mod 4, \]  \hspace{1cm} (5.15b)

\[ a_2 = -a_n, \text{ and} \]  \hspace{1cm} (5.15c)

\[ a_1 = -a_{n-1}. \]  \hspace{1cm} (5.15d)

(Note that condition (5.15a) implies condition (5.15b).)

The implications of these conditions depend on the value of \( n \mod 4 \) in a manner now to be described. If \( n \equiv 0 \mod 4 \), then condition (5.14c) places no restrictions on \( a \) beyond those imposed by condition (5.14a).
Similarly, condition (5.14d) places no restrictions on \( a \) beyond those imposed by condition (5.14b). Thus, the elements of \( a \) are actually subject only to condition (5.14a), from which it is easy to show that \( N(N_2) \) has dimensionality equal to two. If \( n \equiv 1 \mod 4 \), then conditions (5.14b) and (5.14c) imply that \( a_2 = -a_4 \), and conditions (5.14b) and (5.14c) imply that \( a_1 = -a_4 \). Thus, \( a_2 = a_4 \), which contradicts (5.14a) unless \( a_2 = a_4 = 0 \). But, \( a_2 = 0 \) implies that \( a_1 = 0 \) for all \( i \), so that \( N_2 \) is nonsingular when \( n \equiv 1 \mod 4 \). If \( n \equiv 2 \mod 4 \), then conditions (5.14a), (5.14b), and (5.14d) imply that \( a_1 = 0 \) for odd \( i \). Similarly, conditions (5.14a), (5.14b), and (5.14c) imply that \( a_1 = 0 \) for even \( i \). Thus, \( N_2 \) is nonsingular when \( n \equiv 2 \mod 4 \). If \( n \equiv 3 \mod 4 \), an argument very similar to that associated with the case \( n \equiv 1 \mod 4 \) establishes that \( N_2 \) is nonsingular. The rank of \( N_2(I-P_T) \) for each value of \( n \mod 4 \) can then be established by arguments similar to those used to prove (i).

Q.E.D.

Subsequently, let \( N^* \) denote a matrix which can be equal to either \( N_1 \) or \( N_2 \). Also, we abbreviate the phrase "with probability one" to "w.p. 1."

**Lemma 5.2**

Suppose that the support of the distribution of \( \varepsilon \) in the true model (5.2) is \( \mathbb{R}^n \). Then, \( z \neq 0 \) w.p. 1.
Proof:

Clearly, \( z = 0 \) iff \( y \in \mathcal{M}(N^*(I-P_T^*)) \). But the dimensionality of \( \mathcal{M}(N^*(I-P_T^*)) \) equals \( n - \text{rank}(N^*(I-P_T^*)) \), which is less than \( n \) by Lemma 5.1.

By hypothesis, the probability that \( y \) lies in a subspace of \( \mathbb{R}^n \) of dimensionality less than \( n \) is zero, and the result follows. Q.E.D.

**Lemma 5.3**

Suppose that the support of the distribution of \( \varepsilon \) in the true model (5.2) is \( \mathbb{R}^n \). Then, if \( n > 2t + 2 \), \( T(I-P_z)T \) is nonsingular w.p. 1.

Proof:

By Lemma 5.1, the dimensionality of \( \mathcal{C}(N^*(I-P_T^*)) \) is at least \( n - t - 2 \).

Thus, by hypothesis, \( z \) lies in a subspace of \( \mathbb{R}^n \) whose dimensionality is at least \( n - t - 2 \) w.p. 1. Now, using well-known results of linear model theory,

\[
\text{rank}[T(I-P_z)T] = \text{rank}[(I-P_z)T] = \text{rank}(T,z) - \text{rank}(z),
\]

which equals \( \text{rank}(T) \) unless \( z \in \mathcal{C}(T) \). But, if \( z \in \mathcal{C}(T) \), then \( z \) is restricted to a subspace of \( \mathbb{R}^n \) which has dimensionality \( t \). Hence,

\[
P(z \in \mathcal{C}(T)) = 0 \text{ if } n - t - 2 > t, \text{ i.e., if } n > 2t + 2. \quad \text{Q.E.D.}
\]

Although the Papadakis analysis was originally conceived as a method which would eliminate the need for blocking, there is no reason why it could not be used in conjunction with block designs. The extension of a
Papadakis analysis to an analysis-oriented model that includes block effects is straightforward and is not described here. In Section 5.5, this extension is applied to experiments whose plots were assigned to blocks according to Blocking Scheme II in the hope that including block effects in the analysis-oriented model will account for spatial heterogeneity in one dimension and the Papadakis covariate(s) will account for spatial heterogeneity in the other dimension.

Bartlett (1978) suggested iterating the Papadakis analysis, using the estimator of $\gamma$ after each iteration to form the vector of covariates for the next iteration. Thus, at the $i$th stage of this iterative process, an ANOCOVA is performed using the covariate $z^{(i)} = \frac{1}{2}N(y - \hat{\gamma}_P^{(i-1)})$ ($i = 1, \ldots$), where $\hat{\gamma}_P^{(i)}$ is the estimator of $\gamma$ at the $i$th stage, and $\hat{\gamma}_P = \hat{\gamma}_{OLS}$. This process continues until the estimator of $\gamma$ converges; conditions which ensure convergence when plots are laid out according to Configuration I and assigned to blocks according to Blocking Scheme I are given by Draper and Faraggi (1985).

5.2.5. Difference-based methods

Another class of methods by which spatial experiments can be analyzed are difference-based methods, i.e., methods for which estimators of treatment contrasts depend on the data only through certain differences among the observations. Two methods of analysis of this type are the nearest-neighbor (NN) analysis of Wilkinson et al. (1983) and the first-difference analysis of Besag and Kempton (1986).
The development of the NN analysis was motivated by a Monte Carlo randomization study of uniformity trial data carried out by Wilkinson et al. in which it was found that the F-test for detecting treatment differences in a noniterated Papadakis analysis is conservatively biased, while an iterated Papadakis analysis tends to produce too large an F-ratio. Moreover, whether iterated or not, a Papadakis analysis proved to be quite inaccurate (relative to Method RCB) in some situations. Wilkinson et al. attributed the loss of accuracy to the subtraction of treatment averages from the data when forming the Papadakis covariate. They suggested that a better way to adjust for nearest neighbors is to reduce the yields \( \{y_i\} \) to differences

\[
y_i'(b) = y_i - b\bar{y}_{N_i}
\]

(5.16)
called NN-adjusted yields, where \( b \) is a scalar to be determined and \( \bar{y}_{N_i} \) represents the average yield of the \( i \)th plot's neighbors. Note that if plots are laid out according to Configuration I and adjacent plots are designated as neighbors, then

\[
y_i'(1) = y_i - \frac{1}{2}(y_{i-1} + y_{i+1}).
\]

Suppose, for some integer \( i_o \) such that \( 2 \leq i_o \leq n-1 \), that \( y_{i_o-1}, y_{i_o}, \) and \( y_{i_o+1} \) were related according to the deterministic linear equation

\[
y_i = \tau [i] + u_i + v,
\]

where \( \tau [i] \) is the effect of the treatment applied to plot \( i \), and where \( u \) and \( v \) are arbitrary scalars. Then,

\[
y_{i_o}'(1) = \tau [i_o] + u_{i_o} + v - \frac{1}{2}[\tau [i_o-1] + u(i_o-1) + v + \tau [i_o+1] + u(i_o+1) + v]
\]
\[ \hat{\tau}_{[i_o]} = \tau_{[i_o]} - \frac{1}{2}(\tau_{[i_o-1]} + \tau_{[i_o+1]}) \]

which is free of \( u \) and \( v \). Thus, an estimator of \( \tau \) that depends on \( y \) only through \( \{y_i(1): i = 1, \ldots, n\} \) is approximately invariant to locally nearly linear trends, i.e., relatively unaffected by fertility trends (or other environmental trends) that are nearly linear over that portion of the length of the field spanned by a plot and its neighbors.

We describe the NN analysis in the context in which it was introduced by Wilkinson et al. (1983), namely, a field whose plots are laid out according to Configuration II and assigned to blocks according to Blocking Scheme II, and in which adjacent plots within the same block are designated as neighbors. It is assumed that an additional treated plot is adjoined to both ends of each block so that the two border plots in each block have their full complement of neighbors for NN adjustment. These additional plots are used solely to compute \( y_i(1) \) for all internal plots. The total number of plots is \( n = r(t+2) \). This particular application of the NN methodology is called a linear one-dimensional NN analysis by Wilkinson et al., and is most appropriate when long narrow plots are oriented in such a way that adjacent plots within the same block abut one another along their longer dimension, for it is in just such a case that the centroids of neighbors are in closer proximity and thus neighbors are more likely to be highly correlated.

The equations by which \( \tau \) is estimated in a NN analysis are
\[(I_{t-1} - \frac{1}{b} \hat{N}) T = (I_{t-1} - \frac{1}{b} \hat{N}) \hat{M} Y\]  

(5.17)

where $M = \text{diag}(0, 1, 1, \ldots, 1)$ is an $n \times n$ matrix, $\hat{M}^{*} = \text{diag}(0, 1, 1, \ldots, 1, 0)$ is a $(t+2) \times (t+2)$ matrix, $\hat{N} = \text{diag}(\tilde{N}, \tilde{N}, \ldots, \tilde{N})$ is an $n \times n$ neighbor incidence matrix, and $\tilde{N}$ is a $(t+2) \times (t+2)$ matrix whose elements on the first super- and sub-diagonals equal one and whose remaining elements equal zero. Equations (5.17) are derived not from an analysis of covariance using the difference given by (5.16) as a covariate, but rather by equating $\Sigma [y_{i}(b) - \hat{y}_{\tilde{N}}(b)]$, where $y_{i}(b)$ is the average of $y_{j}^{(b)}$ over all plots in the same block as plot $i$ and $\tilde{U}_{j}$ is the set of all internal plots receiving the $j$th treatment, to its expectation (regarding $b$ as fixed) under a certain "smooth trend plus independent errors" model defined and deemed appropriate by Wilkinson et al. Interestingly, despite their quite different origin, equations (5.17) are very similar to the estimation equations that result from applying a maximum likelihood approach to a certain conditional spatial autoregression model, as will be demonstrated in Section 5.3.

Two $NN$-estimators of $\tau$ are defined by Wilkinson et al.: the intra-$NN$ estimator, denoted $\hat{\tau}_{IN}$, and the extra-$NN$ estimator, denoted $\hat{\tau}_{EX}$. Both $\hat{\tau}_{IN}$ and $\hat{\tau}_{EX}$ are particular solutions to equations (5.17) which satisfy $\hat{\tau}_{EX}^{\prime} T = 0$; the distinction between them is due to the choice of $b$ in the equations. In an intra-$NN$ analysis, $b$ is set equal to one. In the more complicated extra-$NN$ analysis, $b$ is chosen according to a scheme which minimizes the residual variance of the $NN$-adjusted yields under the "smooth trend plus independent errors" model. The choice for $b$ is
\[ b_{opt} = \frac{\hat{\omega}}{\omega + 1} \]  
(5.18)

where

\[ \hat{\omega} = \frac{3}{2} (\text{RMS}_{RCB}/\text{RMS}_{In})^{-1} \]

\text{RMS}_{RCB} \text{ represents the residual mean square from Method RCB, and } \text{RMS}_{In} \text{ represents the estimate of residual variance from an intra-NN analysis given by } \hat{\omega} w/u, \text{ where }

\[ w = \left[ I_{t+2}^{-1} - \lambda_{t+2} - \lambda_{t+2} \right] \otimes I_{r} \cdot M \left[ (I-P_{T}) - \frac{1}{2} N_{T} \right] y \]

and \( u \) does not depend on the data (see Wilkinson et al., 1983, equations 24 and 33).

The computation of the covariance matrix of either \( \hat{\tau}_{In} \) or \( \hat{\tau}_{Ex} \) under the "smooth trend plus independent errors" model is very involved and is not described very clearly by Wilkinson et al.; therefore, these computations will not be described here.

The claim by Wilkinson et al. that a NN analysis tends to be considerably more efficient than a Papadakis analysis has been disputed by many investigators (e.g., Patterson, Howes, Ripley, and Martin) in the discussion following their 1983 paper. The criticisms center on the failure of Wilkinson et al. to penalize the NN analysis for requiring additional plots when comparing it to a Papadakis analysis. Thus, the gain in information per plot is not so great as is claimed. Indeed, the use of additional plots solely for NN adjustment seems wasteful.
Nevertheless, empirical results of Besag and Kempton (1986) suggest that a NN analysis is usually somewhat more accurate and efficient than a Papadakis analysis even when the former is penalized for requiring extra plots.

Despite the increased efficiency of a NN analysis, there are at least two reasons why most practitioners might prefer the Papadakis analysis. The first reason is that more complex experimental designs are required for the NN analysis to be appropriate. The appropriateness of the formulae for the covariance matrices of \( \hat{\tau}_{\text{in}} \) and \( \hat{\tau}_{\text{ex}} \) given by Wilkinson et al. requires that the design have a certain degree of NN balance (Williams, 1952). Procedures for constructing NN-balanced designs are not yet widely available. The second reason for preferring the Papadakis analysis to the NN analysis is the greater complexity of the NN analysis. Software capable of performing a NN analysis is not yet widely available, so practitioners wishing to use such an analysis face the formidable task of translating its somewhat cryptic description by Wilkinson et al. (1983) into computer code.

In addition to its practical limitations, the philosophy underlying the NN analysis is more radical than that behind the Papadakis analysis. Because of the need for additional plots and a degree of NN balance in the design, experiments must be designed with NN analysis in mind if such an analysis is to be appropriate. In fact, Wilkinson et al. suggested that NN-balanced designs and NN analyses replace classical designs and methods of analysis. On the other hand, most workers agree with Bartlett (1978), who views a Papadakis analysis, like any other
covariance technique, as an ancillary analysis performed subsequent to an ordinary ANOVA in an effort to reduce the residual mean square. From this point of view, orthodox randomized designs need not be abandoned even if one believes, prior to conducting the experiment, that a Papadakis analysis may be useful.

Another difference-based method of analysis proposed by Besag and Kempton (1986) is computationally straightforward and has the potential to be more flexible than a NN analysis with regard to blocking scheme and design. Subsequently, we refer to this method as Method BK. For concreteness we discuss Method BK as it applies to a field whose plots are laid out according to Configuration II and assigned to blocks according to Blocking Scheme IIr. This is a slightly more general setting than that considered by Besag and Kempton, but their method extends quite naturally to this setting.

Let \( \Delta_t \) denote the \( r(t-1) \times rt \) matrix which transforms the vector of plot yields into the vector of first difference within blocks, i.e.,

\[
\Delta_t^* = \text{diag}(\Delta_t^*, \Delta_t^*, ..., \Delta_t^*)
\]

where

\[
\Delta_t^* = \begin{bmatrix}
1 & -1 \\
1 & -1 \\
\vdots & \ddots & \ddots \\
1 & -1
\end{bmatrix}
\]

has dimensions \( (t-1) \times rt \). The analysis-oriented model corresponding to Method BK in this setting is a special case of model (5.1) in which
g = r(t-1), G = A^r, q = r, A is a block incidence matrix, a is a vector of block parameters, and V = \theta_1 I (\theta_1 > 0). Subsequently, this model is referred to as Model BK.

Let \( u = \Delta_T v \) and let \( F = \Delta_T T \). Under Model BK we have \( E(u) = F \eta \) (since \( \Delta_T A \eta = 0 \)) and \( \text{Var}(u) = \theta_1 I \). Now \( F \), unlike \( T \), does not have full column rank. In fact, each row of \( F \) has one element equal to 1, one element equal to -1, and \( t-2 \) elements equal to 0; thus, the columns of \( F \) sum to zero. It can be shown that the rank of \( F \) is \( t-1 \) and that all treatment contrasts are estimable from \( u \). The BLUE of \( \hat{\tau} \) under Model BK is \( \hat{\tau}_{BK} \), where \( \hat{\tau}_{BK} \) is any solution to the equations

\[
F' F \hat{\tau} = F' u. \tag{5.20}
\]

The covariance matrix of \( \hat{\tau}_{BK} \) is \( \theta_1 A' (F' F)^{-1} A \), which can be estimated by replacing \( \theta_1 \) by an estimate.

Note that \( \hat{\tau}_{BK} \) depends on \( v \) only through \( \Delta_T v \), the i-th element of which is twice the second difference \( y_i - \frac{1}{2} (y_{i-1} + y_{i+1}) \) except for those elements of \( \Delta_T v \) corresponding to border plots. Thus, \( \hat{\tau}_{BK} \), like \( \hat{\tau}_{In} \), is approximately invariant to locally nearly linear trends.

Besag and Kempton suggested that an estimator for \( \theta_1 \), like that for \( \tau \), should be at least approximately invariant to locally nearly linear trends. The "natural" estimator

\[
(u - F \hat{\tau}_{BK})' (u - F \hat{\tau}_{BK}) / (n-t-r)
\]
of $\theta_1$ is unbiased under Model BK but is not even approximately invariant to locally nearly linear trends; however, the estimator

$$
\hat{\theta}_1 = (u-F_{BK}^\prime) \Delta_{t-1} (u-F_{BK}^\prime) / u
$$

(5.21)

where $u$ is determined so that $\hat{\theta}_1$ is unbiased for $\theta_1$, does have this approximate invariance property. The appropriate value of $u$ is not available in the literature (since we are extending the Besag-Kempton approach to a different setting than the one in which it was presented), but is easily derived.

**Lemma 5.4**

The estimator $\hat{\theta}_1$ is unbiased for $\theta_1$ under Model BK when

$$
u = r(2t-4) - tr(\Delta_{t-1} (I-F(F)^\prime F)^\prime).
$$

**Proof:**

Since $u - F_{BK}^\prime = (I-F(F)^\prime F)^\prime u$, we have $E(u-F_{BK}^\prime) = 0$ and $\text{Var}(u-F_{BK}^\prime) = \theta_1 (I-F(F)^\prime F)^\prime$. Thus, the expectation of the numerator of (5.21) is $\theta_1 tr(\Delta_{t-1} (I-F(F)^\prime F)^\prime)$. Because $\Delta_{t-1}^* (I-F(F)^\prime F)^\prime$ is a $(t-1)\times(t-1)$ matrix whose first and (t-1)st diagonal elements equal unity and whose remaining diagonal elements equal two, we have

$$
tr(\Delta_{t-1}^* (I-F(F)^\prime F)^\prime) = r[2(t-3)+2] = r(2t-4).
$$

Q.E.D.
Besag and Kempton (1986) indicated how their method of analysis could be extended to accommodate an analysis-oriented model which is more general than Model BK but is still a special case of model (5.1). The model to which Method BK could be extended is equivalent to Model BK in every respect except that \( \text{Var}(A_t e) = \theta_1 I + \theta_2 A_t A_t', \) where \( \theta_1 > 0 \) and \( \theta_2 > 0. \) This model is a components of variance model; hence, \( \theta_1 \) and \( \theta_2 \) can be estimated by any one of several well-known variance component estimation procedures, after which \( \tau \) can be estimated using equations having the same form as the equations which result from applying generalized least squares to this components of variance model, but with the covariance matrix \( \theta_1 I + \theta_2 A_t A_t' \) replaced by its estimate.

5.2.6. Response surface methods

To conclude our review of methods for analyzing spatial experiments, we briefly mention a few techniques related to response surface methodology. Pearce (1980) fits a paraboloidal surface, i.e., a full second-order polynomial response surface in the two plot coordinates, to spatial experimental data using ordinary least squares; he views this technique, hereafter referred to as Method PS, as a competitor to a more general row-column analysis when plots are laid out according to Configuration II, particularly if the number of rows and columns is large. The analysis-oriented model corresponding to Method PS is a special case of model (5.1) for which \( g = n, G = I, q = 6, A = A_{PS} \) where \( A_{PS} \) is a matrix whose \( i \)th row is \( (1, r[i], c[i], r[i]^2, c[i]^2, r[i] c[i]) \) (where \( r[i] \) and \( c[i] \) are the row and column, respectively, to which plot \( i \) belongs), and
\[ V = \theta_1 \mathbf{I} \quad (\theta_1 > 0). \] Subsequently, we refer to this model as Model PS. If treatment contrasts are estimable under model PS, the BLUE of \( \lambda \) is \( \hat{\lambda}_{PS} \), where \( \hat{\lambda}_{PS} \) is given by any solution to equations (5.3) with \( A = A_{PS} \). The covariance matrix of \( \lambda \) is given by (5.4) with \( A = A_{PS} \), and can be estimated unbiasedly by putting \( u = n-t-5 \) and \( A = A_{PS} \) into (5.5) and (5.6), respectively.

A response surface model which is nonlinear in its parameters was suggested for spatial experimental data by Draper and Guttman (1980).

### 5.3. Relationships Among Approaches

In spite of their different origins, many of the estimation procedures described in the previous section are closely related to one another and also to applications of generalized least squares or maximum likelihood methods to various spatial models. In this section, we explore these relationships.

Atkinson (1969) was apparently the first to discover a relationship between the Papadakis method of analysis and ML estimation. For a field whose plots are laid out according to Configuration I and assigned to blocks according to Blocking Scheme I, Atkinson demonstrated that \( \hat{\gamma}_{PAP} \) is a very good approximation to the unique ML estimator when: (1) for the Papadakis analysis, adjacent plots are designated as neighbors and plots are regarded as lying on the perimeter of a circle (this implies that the neighbor incidence matrix is \( N_2 \)); (2) the model to which ML estimation is applied is the special case of model (5.1) in which \( g = n, G = I, q = 0, \)
(where \( \theta_1 > 0 \) and \( |\theta_2| < 1 \), and \( \varepsilon \) is multivariate normal.

The relationship between the Papadakis approach and ML estimation can also be demonstrated in the more important two-dimensional setting. From expression (5.9),

\[
\hat{\mathrm{IPAP}} = (T'T)^{-1}[T'y - \frac{1}{2} \hat{\phi}T^t N(y - \hat{T}_{OLS})],
\]  

(5.22)

with \( \hat{\phi} \) given by (5.10). Now, consider the application of ML to a special case of model (5.1) in which \( g = n, G = I, q = 0, \varepsilon \) is multivariate normal, \( \Sigma = \theta_1(I-\theta_2N)^{-1} \), where \( \theta_1 > 0 \) and \( \theta_2 > 0 \), and \( N \) is any neighbor incidence matrix. By Theorem 1.3, such a covariance matrix could arise from the particular conditional autoregression (CAR) model for which \( C = \theta_2 N \). A ML estimator \( \hat{T}_{\mathrm{CAR}} \) of \( T \) is any solution to the equations

\[
T'(I-\theta_2N)T = T'(I-\hat{\theta}_2N)y,
\]  

(5.23)

where \( \hat{\theta}_2 \) is a ML estimator of \( \theta_2 \) (Draper and Faraggi, 1985).

Equations (5.23) can be easily rearranged into the form

\[
\hat{T}_{\mathrm{CAR}} = (T'T)^{-1}[T'y - \theta_2 T^t N(y - T_{\mathrm{OLS}})],
\]  

(5.24)
which bear a striking resemblance to equations (5.22). On this basis, Ripley (1978) pointed out that if \( \phi \) was fixed and equal to 202, then (5.22) could be viewed as the first iteration in an iterative process for solving equations (5.23).

The suggestion by Bartlett (1978) that the Papadakis analysis be iterated, together with the similarity of equations (5.22) and (5.24), have prompted several workers to explore the relationship between \( \hat{\tau}_{\text{PAP}}^{(i)} \) and \( \hat{\tau}_{\text{CAR}} \). Draper and Faraggi (1985) essentially showed that \( \hat{\tau}_{\text{PAP}}^{(i)} \rightarrow \hat{\tau}_{\text{CAR}} \) as \( i \rightarrow \infty \) under the following conditions: (1) plots are laid out according to Configuration I and regarded as lying on the perimeter of a circle; (2) adjacent plots are designated as neighbors (together with (1), this implies that \( N = N_2 \)); (3) the design is such that each treatment occurs adjacent to every other treatment the same number of times; (4) \( 0 < \theta_2 < \frac{1}{2} \).

Martin (1978, 1982) proved a similar result applicable to Configuration II.

The NN-estimators \( \hat{\tau}_{\text{IN}} \) and \( \hat{\tau}_{\text{EX}} \) are also quite similar to \( \hat{\tau}_{\text{CAR}} \), as observed by Ripley (1983). Suppose that the configuration of the field, the blocking scheme, the neighbor designation, and the design are those which are appropriate for the linear one-dimensional NN analysis described in Section 5.2.5. Suppose further that deviations of yields from their block averages follow the particular CAR model in which \( C = \frac{1}{2} bN \). Then, the equations which result when generalized least squares is applied to the deviations are

\[
T'(I - \frac{1}{2} bN)T = T'(I - \frac{1}{2} bN)[I_{t+2} - 1]T_{t+2} (I_{t+2} - 1) \otimes I_{\tau_{\text{NN}}}.
\]
These equations differ from equations (5.17) only in that $M$ is absent (so that, in contrast to the NN analysis, every plot yield is included directly in the analysis) and in that block effects are accounted for in a different manner (here deviations of the raw yields, rather than the NN-adjusted yields, from their block averages are formed).

The estimator $\hat{\tau}_{BK}$ obtained by the Besag-Kempton approach is closely related to $\hat{\tau}_{In}$ (and therefore is closely related to $\hat{\tau}_{CAR}$ and $\hat{\tau}_{PAP}$ through relationships already described). Suppose that the experimental setting is that which is appropriate for the linear one-dimensional NN analysis, and that adjacent plots within the same block are designated as neighbors. Then,

$$\Delta_{t+2} = 2(I - \frac{1}{2}N - \frac{1}{2}v) ,$$

where $v = \{v_i\}$ is an $n \times 1$ vector such that $v_i = 1$ if plot $i$ is a border plot, $v_i = 0$ otherwise. Now, equations (5.20), when extended to this setting, can be written as

$$T'\Delta_{t+2}T = T'(I - \frac{1}{2}N - \frac{1}{2}v)y .$$

Therefore, aside from how border plots and block effects are treated, the Besag-Kempton approach is equivalent to the intra-NN approach. This
near-equivalence was noted by Besag and Kempton, and helps to explain why the empirical results of their first-difference approach were very similar to the results of an intra-NN analysis.

In summary, several ostensibly different approaches to the analysis of spatial experiments are remarkably similar, being closely related to generalized least squares as it applies to a CAR with covariance matrix \( \Theta_1 (I-\Theta_2 N)^{-1} \) for various choices of \( N \). The random field approach also leads to what is essentially a generalized least squares estimation procedure for \( \gamma \) (cf. Chapter 2). It is therefore of interest to determine whether an RFLM exists for which the covariance matrix of \( y \) is equal (or very nearly equal) to \( \Theta_1 (I-\Theta_2 N)^{-1} \). This will clarify the relationship of the random field approach to other approaches.

Consider a random field whose mean function is identically equal to zero and whose covariogram is the \( L^1 \)-isotropic exponential covariogram

\[
\hat{C}(x,y;\theta) = \theta_1 \exp\{-\theta_2 (|x|+|y|)\} \quad (\theta_1 > 0, \theta_2 > 0).
\]

Note that this random field is separable. If plots are laid out according to the special case of Configuration II in which \( H_1 = H_2 = H \), then according to Theorem 3.4, the covariance matrix of the observations under the RFLM-P corresponding to this random field is \( V = \Theta_1 P \otimes Q \), where
\[ P = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{C-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \rho & \rho^2 & \cdots & \rho^{C-1} \\ \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{R-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \rho & \rho^2 & \cdots & \rho^{R-1} \\ \end{bmatrix} \]

and \( \rho = \exp(\theta^2 H) \). Hence, \( V^{-1} = \frac{1}{\theta^2} P^{-1} \otimes Q^{-1} \), where

\[
\begin{bmatrix}
1 & -\rho & \cdots & 0 \\
\rho & 1+\rho^2 & -\rho & \cdots \\
0 & \rho & 1+\rho^2 & \ddots & \vdots \\
& & \ddots & \ddots & \ddots \\
& & & \rho & 1+\rho^2 & -\rho \\
0 & \cdots & \cdots & \cdots & \cdots & 1
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
1 & -\rho & \cdots & 0 \\
\rho & 1+\rho^2 & -\rho & \cdots \\
0 & \rho & 1+\rho^2 & \ddots & \vdots \\
& & \ddots & \ddots & \ddots \\
& & & \rho & 1+\rho^2 & -\rho \\
0 & \cdots & \cdots & \cdots & \cdots & 1
\end{bmatrix}
\]

[see formula (3.8)]. Now, consider the matrices
and

\[
\hat{P}^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix}
1 + \rho^2 & -\rho \\
-\rho & 1 + \rho^2 \\
& & \ddots & \ddots \\
& & & -\rho & 1 + \rho^2 \\
& & & & & & 1 + \rho^2
\end{bmatrix}
\]

and

\[
\hat{Q}^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix}
1 + \rho^2 & -\rho \\
-\rho & 1 + \rho^2 \\
& & \ddots & \ddots \\
& & & -\rho & 1 + \rho^2 \\
& & & & & & 1 + \rho^2
\end{bmatrix}
\]

obtained by replacing the first and last diagonal elements of \(P^{-1}\) and \(Q^{-1}\) with \(1 + \rho^2\), and set \(\eta = \rho/(1 + \rho^2)\). Let \(N_3\) denote the neighbor incidence matrix corresponding to the scheme in which adjacent plots in both dimensions are designated as neighbors. Then, we have

\[
\left( \frac{1 - \rho^2}{1 + \rho^2} \right)^2 \hat{P}^{-1} \otimes \hat{Q}^{-1} = I - \eta N_3.
\]

Now, \(\hat{P}^{-1} \otimes \hat{Q}^{-1}\) differs from \(P^{-1} \otimes Q^{-1}\) only at those entries which correspond to border plots. Therefore, apart from border plot considerations, ML estimation of the parameters of this RFLM is equivalent to solving equations (5.23) with \(N = N_3\).
5.4. Properties of Treatment Parameter Estimators

Assume that the true model governing the plot yields of a spatial experiment is model (5.2). Suppose, however, that for the purpose of suggesting an analysis, the model adopted for $y$ is a member of the class of analysis-oriented models (5.1). Suppose that every treatment contrast is estimable in model (5.2). Suppose further that every treatment contrast is estimable w.p. 1 in any particular case of model (5.1) to be considered.

Subsequently, we say that an estimator $w(y)$ of $\lambda^\top T$ is odd if $w(-y) = -w(y)$ for all $y \in \mathbb{R}^n$, and we say that $w(y)$ is a location equivariant estimator of $\lambda^\top T$ if $w(y + \lambda k) = w(y) + \lambda k$ for all $k \in \mathbb{R}^t$ and all $y \in \mathbb{R}^n$. All of the estimators of treatment contrasts discussed in the previous sections of this chapter are odd and location equivariant w.p. 1, as is now proved. (Part (i) of the theorem is easily demonstrated so a formal proof is not given.)

**Theorem 5.1**

(i) The linear estimators $\lambda^\top \hat{T}_{RCB}$, $\lambda^\top \hat{T}_{RC}$, $\lambda^\top \hat{T}_{COV}$, $\lambda^\top \hat{T}_{IN}$, $\lambda^\top \hat{T}_{Ex}$, and $\lambda^\top \hat{T}_{PS}$ are odd and location equivariant.

(ii) The Method PAP estimator $\lambda^\top \hat{T}_{PAP}$ is odd and location equivariant w.p. 1 if $\lambda \in \mathbb{R}[T(I-P_Z)T]$ w.p. 1.

(iii) The extra-NN estimator $\lambda^\top \hat{T}_{Ex}$ is odd and location equivariant w.p. 1 if $\lambda \in \mathbb{R}[(I - k)T M(I - \frac{1}{2} \hat{b}^\top \hat{N})T]$ w.p. 1.
(iv) The estimators $\hat{\lambda} \sim \tau_{RF}$ and $\hat{\lambda} \sim \tau_{RF}$, obtained by applying the methods of ML and REML, respectively, to a treatment-additive RFIM, are odd and location equivariant, as is $\hat{\lambda} \sim \tau_{CAR}$.

**Proof of (ii):**

Recall that $\hat{\tau}_{PAP}$ is any solution to

$$T'(I-P_T)T = T'(I-P_T)y,$$

where $P_T = [y(I-P_T)NN(I-P_T)y]^{-1}N(I-P_T)y'(I-P_T)N$. Put $w(y) = \hat{\tau}_{PAP}$. Note that the elements of $P_T$ are even functions of $y$ and are translation invariant. Thus, $w(-y) = -w(y)$, so that $\hat{\tau}_{PAP}$ is odd. Furthermore, by hypothesis there exists (w.p. 1) a vector $\tau$ such that $\tau' (I-P_T)T = \lambda'$. Therefore,

$$w(y+Tk) = \lambda' [T(I-P_T)T]^{-1} T(I-P_T)(y+Tk)$$

$$= \tau' T(I-P_T)[T(I-P_T)T]^{-1} T(I-P_T)Tk + w(y)$$

$$= \lambda' k + w(y),$$

showing that $\hat{\tau}_{PAP}$ is location equivariant w.p. 1. Q.E.D.

**Proof of (iii):**

Recall that $\hat{\tau}_{Ex}$ is any solution to equations (5.17), with $b = b_{opt}$ given by (5.18). Put $w(y) = \hat{\tau}_{Ex}$. From (5.19), RMS$_{RCB}$ and RMS$_{In}$ are
clearly even and translation invariant functions of $y$. Consequently,

$b$ is an even and translation invariant function of $y$. Thus, $w(\cdot y) = -w(y)$, so that $\lambda_{\text{Ex}}^*$ is odd. By hypothesis, there exists (w.p. 1) a vector $\mathbf{q}$ such that $\mathbf{q}^T (I - \frac{1}{2} M(I - \frac{1}{2} \mathbf{b}^\text{opt})^T M(I - \frac{1}{2} \mathbf{b}^\text{opt})^T = \lambda'$. An argument very similar to that used in establishing the location equivariance of $\lambda_{\text{PAP}}^*$ can be used to establish that $\lambda_{\text{Ex}}^*$ is location equivariant w.p. 1.

Q.E.D.

Proof of (iv):

The ML estimator $\lambda_{\text{RF}}^*$ associated with a treatment-additive random field linear model is given by the second part of any solution to the equations

$$X \mathbf{v}^{-1} \mathbf{y} \begin{bmatrix} \alpha & -\mathbf{v}^T \\ -\mathbf{v} & 1 \end{bmatrix} = X \mathbf{v}^{-1} \mathbf{y},$$

where $\hat{\mathbf{v}} = V(\hat{\theta})$ and $\hat{\theta}$ is a value of $\theta \in \Theta$ that maximizes

$$L_1^*(\theta; y) = -\frac{1}{2} \log|V| - \frac{1}{2} \mathbf{v}^T \mathbf{v}_X.$$

The REML estimator $\lambda_{\text{RF}}^*$ is obtained in a similar fashion, but $\tilde{\mathbf{v}} = V(\tilde{\theta})$ is used instead of $\hat{\mathbf{v}}$, where $\tilde{\theta}$ is a value of $\tilde{\theta} \in \Theta$ that maximizes

$$L_1(\tilde{\theta}; y) = L_1^*(\tilde{\theta}; y) - \frac{1}{2} \log|X^* \mathbf{v}^{-1} X^*|.$$
Matrices in the above expressions for $L^*_1$ and $L_1$ were defined in Chapters 1 and 2. It was noted by Kackar and Harville (1981) that $L^*_1$ and $L_1$ are even, translation invariant functions of $\gamma$; by the definition of ML and REML estimates (cf. Section 2.2), this implies that $\hat{\Theta}$ and $\bar{\Theta}$ are even and translation invariant. Therefore, both $\lambda_{RF}^\wedge$ and $\lambda_{RF}^\wedge$ are odd and location equivariant. Essentially the same argument can be used to establish that $\lambda_{CAR}^\wedge$ is odd and location equivariant. Q.E.D.

We now prove a theorem that is a special case of a more general theorem given by Jeske and Harville (1986), and use it to establish the unbiasedness of the nonlinear estimators of $\lambda_\wedge^T$. The unbiasedness [under model (5.2)] of the linear estimators listed in part (i) of Theorem 5.1 can be easily verified (assuming that their expectations exist).

We say that the distribution of a random vector $x$ is symmetric about a vector $x_\wedge$ if $(x-x_\wedge)$ has the same distribution as $-(x-x_\wedge)$.

**Theorem 5.2**

If the distribution of $\xi$ in model (5.2) is symmetric about $0$, and if $w(\eta)$ is any odd, location equivariant estimator of $\lambda_\wedge^T$, then the distribution of $w(\eta) - \lambda_\wedge^T$ is symmetric about $0$.

**Proof:**

Define $z = J^\wedge \eta$, where $J$ is an $n \times (n-t)$ matrix such that $J^T = 0$ and $\operatorname{rank}(J) = n-t$. Also define $z_0 = v - \tilde{r}^T \eta$, where $v = \lambda_\wedge^T$ and $\tilde{r}$ is any $n \times 1$ vector such that $\tilde{r}^T = \lambda_\wedge$. Since
and since \( z \) has a symmetric distribution about 0, \( (z_0, z) \) also has a symmetric distribution about 0. Furthermore,

\[
\begin{bmatrix}
  z_0 \\
  z
\end{bmatrix} = \begin{bmatrix}
  v - x' \\
  y
\end{bmatrix} = - \begin{bmatrix}
  x' \\
  y'
\end{bmatrix} \varepsilon ,
\]

and for any two scalars \( v_1 \) and \( v_2 \) and any two \( n \times 1 \) vectors \( y_1 \) and \( y_2 \),

\[
v_1 - v_2 = \frac{y_1}{z} - \frac{y_2}{z} \quad \text{and} \quad J \frac{y_1}{z} = J \frac{y_2}{z}
\]
desire that \( v_2 = v_1 + \lambda \frac{y_2}{z} \) and

\[
y_2 = y_1 + Tk
\]
for some \( k \); thus, \( (z_0, z) \) is a maximal invariant with respect to transformations from \( \mathbb{R}^{n+1} \) to \( \mathbb{R}^{n+1} \) of the form \( g(v, y) = (v + \lambda k, y + Tk) \). Let \( t(v, y) = w(y) - v \). Then,

\[
t(v + \lambda k, y + Tk) = w(y + Tk) - (v + \lambda k)
\]

\[
= w(y) - v
\]

\[
= t(v, y).
\]

Thus, \( t(v, y) \) is invariant with respect to transformations of the general form \( g(v, y) = (v + \lambda k, y + Tk) \). It follows that \( t(v, y) \) can be re-expressed as a function of \( z_0 \) and \( z \), say \( h(z_0, z) \) (see, e.g., Theorem 1 of Section 5.6 of Ferguson, 1967). Since \( w(y) \) is an odd estimator,
\[ h(-z_0, z) = t(-v, -y) = w(-y) + v = -w(y) + v = -h(z_0, z). \]  

Theorem 2.1 of Wolfe (1973), together with the symmetry of the distribution of \((z_0, z)^\prime\) about 0 and relation (5.26), imply the result. Q.E.D.

**Corollary**

Suppose that the distribution of \(\varepsilon\) in model (5.2) is symmetric about 0. Then, under model (5.2):

(i) If \(\lambda \in \mathbb{R}[T(I-P_Z)^T] w.p. 1\), and if \(E(\lambda_\sim T_{PAP})\) exists, \(\lambda_\sim T_{PAP}\) is unbiased.

(ii) If \(\lambda \in \mathbb{R}[T(I-I_{P_Z}) M(I-\frac{1}{Z_{opt}} N)] T w.p. 1\), and if \(E(\lambda_\sim T_{Ex})\) exists, \(\lambda_\sim T_{Ex}\) is unbiased.

(iii) The estimators \(\lambda_\sim T_{RF}, \lambda_\sim T_{RF}, \) and \(\lambda_\sim T_{CAR}\) are unbiased provided that their expectations exist.

**Proof:**

The results follow directly from Theorems 5.1 and 5.2. Q.E.D.

5.5. An Empirical Comparison of Methods of Analysis

It would seem to be useful to compare methods of analyzing spatial experiments not only analytically but also with respect to how well they
perform when applied to data from actual spatial experiments. In this section we present results which suggest that: (1) several previously proposed methods may often be superior to the classical analysis; (2) a random field approach may often be superior to other approaches.

Methods of analysis which involve the construction of an ANOVA table (e.g., Method RGB, Method RC, or an uniterated Papadakis analysis) have been compared using actual experimental data by Pearce (1980), Kempton and Howes (1981), and Hinz (1985), among others. Typically, such methods are compared on the basis of the magnitude of the residual mean squares from their corresponding ANOVA tables or, alternatively, on the basis of the magnitude of the coefficients of variation. While such studies may indicate which method of analysis leads to the greatest increase in precision, they typically provide little (if any) information pertaining to the improvement of point estimates of treatment contrasts because the true values of those contrasts are unknown in actual experiments. Furthermore, for actual experimental data it is difficult to conceive of a basis for comparing methods in which an ANOVA table is not constructed (e.g., the first-difference method of Besag and Kempton (1986) or a random field approach) to methods in which an ANOVA table is constructed.

An alternative way to compare methods of analysis of spatial experiments is via a randomization study of uniformity trial data. In this approach, some number \( h \) of arrangements of "dummy" treatment labels are considered. The labels within each arrangement are assigned at random, subject to possible blocking restrictions. For each randomization, the
yields of the uniformity trial are analyzed by each method of analysis, with the analysis being carried out as though the yields were derived from an actual spatial experiment. For each method of analysis, certain quantities (two of which are defined in the following paragraph) that reflect the accuracy and precision of estimates of treatment contrasts are computed for each arrangement and are then averaged over all h arrangements.

In a randomization study of uniformity trial data, we know, since the treatments are not real, that any treatment contrast equals zero. Thus, there is a basis on which to compare the accuracy of point estimates of treatment contrasts for different methods of analysis. Furthermore, any method can be compared to any other method, provided that the quantities which reflect the accuracy and precision of the treatment contrasts are well-defined and can be computed for both methods.

By Theorem 5.1, the estimator of \( \lambda' \) obtained by each method reviewed in this chapter is location equivariant (w.p. 1). Thus, if the effects of treatments are additive, as they are in model (5.2), the results of a randomization study of uniformity data are relevant to actual spatial experiments.

Let \( \gamma_{k,l} \) denote the txl vector whose kth element is 1, whose lth element is -1, and whose remaining elements are zeroes. Put \( \gamma_{k,l} = \lambda' \). The \( \{\gamma_{k,l}\} \) are special cases of treatment contrasts called treatment differences.

Randomization studies of uniformity trial data were conducted by Kempton and Howes (1981, Section 2.2), Wilkinson et al. (1983, Section
1.1), Besag (1983), and Binns and Jui (1985). Several different measures of accuracy and precision of treatment parameter estimation were used by these authors. In our randomization study, we adopted essentially the same measures as Besag (1983).

Following Besag, we define

\[
\text{Emp} = \frac{2}{t(t-1)} \sum_{k=1}^{t} \sum_{l=k+1}^{t} (\hat{\gamma}_{kl})^2
\]

(5.27)

and

\[
\text{Pre} = \frac{2}{t(t-1)} \sum_{k=1}^{t} \sum_{l=k+1}^{t} \text{Var}(\hat{\gamma}_{kl})
\]

(5.28)

where the \{\hat{\gamma}_{kl}\} are the estimates of the treatment differences \{\gamma_{kl}\} associated with a given method of analysis, and \text{Var}(\hat{\gamma}_{kl}) is the estimated variance of \hat{\gamma}_{kl} for that method of analysis. (Expressions (5.27) and (5.28) differ slightly in form from those given by Besag; however, the two pairs of formulae are equivalent.) If \hat{\gamma}_{kl} is an unbiased estimator, then Emp (for "Empirical") estimates the variance of estimated treatment differences, and Pre (for "Predicted") is the average estimated variance of estimated treatment differences (over all \binom{t}{2} possible treatment pairs). Clearly, a small value of Emp for a particular arrangement reflects high accuracy in estimating treatment contrasts for that arrangement, while Pre is inversely related to the precision of those estimates.

Let \overline{\text{Emp}}_h and \overline{\text{Emp}} represent the average of Emp over the h chosen arrangements and the average of Emp over the population of all possible
arrangements, respectively. Define $\bar{\text{Pre}}_h$ and $\bar{\text{Pre}}$ in a similar manner. Besag (1983) defined a method of analysis "to be valid in a randomization framework" if $\bar{\text{Emp}}$ is approximately equal to $\bar{\text{Pre}}$ for that method. He argued further that if a method is to be valid in a randomization framework, then $\bar{\text{Emp}}_h$ and $\bar{\text{Pre}}_h$ should be approximately equal if $h$ is reasonably large, say, at least twenty. Thus, if $\bar{\text{Pre}}_h$ is substantially smaller (larger) than $\bar{\text{Emp}}_h$ for a particular method, it suggests that the method gives variance estimates of estimated treatment differences which are overly optimistic (pessimistic).

In order to investigate the merits of the random field approach relative to other methods of analyzing spatial experiments, we conducted a randomization study of uniformity trial data considerably larger in scope than any other yet published. Three data sets were chosen from the literature for study and are displayed in Tables 5.1, 5.2, and 5.3. The spatial configuration of all three data sets is Configuration II. The data in Table 5.1 are yields of tea from a uniformity trial carried out on Malawi by Laycock (1955). The data in Table 5.2 are yields of barley (expressed as deviations from an overall mean) from a uniformity trial conducted in the United Kingdom in 1978; these data have been condensed (by summation to 4-row plot totals within each column) from a 28x7 array displayed by Kempton and Howes (1981) to the 7x7 array of Table 5.2. The data in Table 5.3 are yields of wheat from a subset of a uniformity trial conducted in Idaho by Wiebe (1935); the 9x6 array of data in Table 5.3 was obtained from rows 49 to 120 and columns 1 to 6 of the entire
125x12 array of plot yields given by Wiebe by summation to form 8-row plot totals within each column.

Table 5.1. Laycock (1955) data

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>28.5</td>
<td>27.1</td>
<td>30.4</td>
<td>27.6</td>
<td>25.1</td>
<td>23.2</td>
</tr>
<tr>
<td>32.7</td>
<td>20.5</td>
<td>22.3</td>
<td>24.8</td>
<td>22.1</td>
<td>22.8</td>
</tr>
<tr>
<td>31.7</td>
<td>23.0</td>
<td>23.9</td>
<td>29.4</td>
<td>22.5</td>
<td>29.0</td>
</tr>
<tr>
<td>26.7</td>
<td>21.9</td>
<td>23.0</td>
<td>25.3</td>
<td>26.9</td>
<td>29.4</td>
</tr>
<tr>
<td>25.8</td>
<td>21.4</td>
<td>20.0</td>
<td>23.7</td>
<td>31.1</td>
<td>37.3</td>
</tr>
<tr>
<td>25.1</td>
<td>20.2</td>
<td>21.2</td>
<td>32.2</td>
<td>27.4</td>
<td>34.7</td>
</tr>
<tr>
<td>28.1</td>
<td>25.3</td>
<td>21.4</td>
<td>25.1</td>
<td>28.5</td>
<td>35.7</td>
</tr>
<tr>
<td>30.6</td>
<td>37.0</td>
<td>52.7</td>
<td>41.9</td>
<td>43.9</td>
<td>45.1</td>
</tr>
<tr>
<td>41.4</td>
<td>49.7</td>
<td>49.2</td>
<td>47.4</td>
<td>47.8</td>
<td>43.7</td>
</tr>
</tbody>
</table>
Table 5.2. Kempton and Howes (1981) data condensed to 4-row plot totals

<table>
<thead>
<tr>
<th></th>
<th>119</th>
<th>62</th>
<th>-60</th>
<th>-107</th>
<th>-225</th>
<th>-243</th>
<th>-43</th>
</tr>
</thead>
<tbody>
<tr>
<td>113</td>
<td>-17</td>
<td>37</td>
<td>-51</td>
<td>-170</td>
<td>-115</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>-52</td>
<td>-5</td>
<td>-23</td>
<td>28</td>
<td>-97</td>
<td>-13</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>-150</td>
<td>45</td>
<td>-34</td>
<td>-33</td>
<td>-39</td>
<td>129</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>-104</td>
<td>-25</td>
<td>-90</td>
<td>118</td>
<td>-61</td>
<td>154</td>
<td>277</td>
<td></td>
</tr>
<tr>
<td>-193</td>
<td>-106</td>
<td>26</td>
<td>143</td>
<td>-37</td>
<td>157</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>-71</td>
<td>-7</td>
<td>17</td>
<td>46</td>
<td>-15</td>
<td>-42</td>
<td>124</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3. Subset of Wiebe (1935) data condensed to 8-row plot totals

<table>
<thead>
<tr>
<th></th>
<th>5515</th>
<th>5435</th>
<th>5995</th>
<th>4670</th>
<th>4400</th>
<th>4880</th>
</tr>
</thead>
<tbody>
<tr>
<td>5555</td>
<td>5905</td>
<td>6210</td>
<td>5080</td>
<td>4585</td>
<td>4750</td>
<td></td>
</tr>
<tr>
<td>5415</td>
<td>6210</td>
<td>6605</td>
<td>5380</td>
<td>4700</td>
<td>4950</td>
<td></td>
</tr>
<tr>
<td>5560</td>
<td>6435</td>
<td>6810</td>
<td>5795</td>
<td>4890</td>
<td>5310</td>
<td></td>
</tr>
<tr>
<td>5275</td>
<td>6410</td>
<td>6690</td>
<td>5500</td>
<td>4855</td>
<td>5210</td>
<td></td>
</tr>
<tr>
<td>5275</td>
<td>5665</td>
<td>5305</td>
<td>4975</td>
<td>4895</td>
<td>4990</td>
<td></td>
</tr>
<tr>
<td>4995</td>
<td>5760</td>
<td>4870</td>
<td>4795</td>
<td>4940</td>
<td>5035</td>
<td></td>
</tr>
<tr>
<td>4802</td>
<td>5865</td>
<td>5320</td>
<td>5330</td>
<td>4600</td>
<td>4665</td>
<td></td>
</tr>
<tr>
<td>3905</td>
<td>4450</td>
<td>4865</td>
<td>5180</td>
<td>4090</td>
<td>4755</td>
<td></td>
</tr>
</tbody>
</table>
The three data sets display features which appear to be common in a large proportion of small to moderately-sized spatial experiments, though they do not encompass the entire range of possible fertility patterns. The Laycock data exhibit a strong gradient along columns in the last three rows, but do not exhibit a strong gradient along rows. A plot of yield versus spatial location for the Kempton and Howes data is somewhat saddle-shaped, with the plots of lowest yield located near the lower left and upper right corners of Table 5.2 and plots of highest yield occurring near the upper left and lower right corners; thus, there is not a fertility gradient along columns which is consistent over rows, nor vice versa. The subset of the Wiebe data displayed in Table 5.3 appears to have a patch of relatively high fertility in the upper left center of the table.

In all three data sets, there appears to be a high positive correlation between adjacent plots. An ANOVA corresponding to a model whose mean structure includes only classification effects for rows and columns was performed for each of the data sets. The results are given in Table 5.4 and reflect the features noted above.

In our study we sought to compare the random field approach to other approaches using the classical randomized blocks ANOVA as a standard. To this end, for each of the three data sets we generated 100 arrangements of dummy treatment labels by randomization subject to Blocking Scheme IIr and then subject to Blocking Scheme IIc. In this way, 100 randomized complete block "pseudo-experiments" were carried
Table 5.4. Row-column ANOVA's of the data sets in Tables 1-3

(a) Laycock data:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>8</td>
<td>3102</td>
<td>17.2</td>
</tr>
<tr>
<td>Columns</td>
<td>5</td>
<td>179</td>
<td>1.6</td>
</tr>
<tr>
<td>Residual</td>
<td>40</td>
<td>901</td>
<td></td>
</tr>
</tbody>
</table>

(b) Kempton-Howes data:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>6</td>
<td>55463</td>
<td>0.9</td>
</tr>
<tr>
<td>Columns</td>
<td>6</td>
<td>235806</td>
<td>3.9</td>
</tr>
<tr>
<td>Residual</td>
<td>36</td>
<td>359908</td>
<td></td>
</tr>
</tbody>
</table>

(c) Wiebe data:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>8</td>
<td>6.8x10^6</td>
<td>6.5</td>
</tr>
<tr>
<td>Columns</td>
<td>5</td>
<td>9.9x10^6</td>
<td>15</td>
</tr>
<tr>
<td>Residual</td>
<td>40</td>
<td>5.2x10^6</td>
<td></td>
</tr>
</tbody>
</table>

out for each of six data set X blocking scheme combinations that, using obvious notation, we refer to as Series LAY, Series LAY, Series KH, Series KH, Series W, and Series W. For each series, values of $\overline{\text{Emp}}_{100}$ and $\overline{\text{Pre}}_{100}$ corresponding to each of nine methods of analysis were computed.

The nine methods of analysis compared were the classical randomized complete blocks ANOVA (method RCB), the row-column analysis (Method RC),
the second-order polynomial covariance analysis (Method PC-2), an
uniterated single-covariate Papadakis analysis based on an analysis-
oriented model without block effects (Method PAP), an uniterated
single-covariate Papadakis analysis based on an analysis-oriented
model including block effects (Method PAPB), generalized least
squares estimation of the parameters of a CAR (Method CAR), the first-
difference analysis of Besag and Kempton (1986) as extended to
Configuration II and Blocking Scheme II (Method BK), the paraboloidal
surface method (Method PS), and a random field approach (Method RF).

The covariate employed by Methods PAP and PAPB was the average of
residuals over those neighbors which did exist, where neighbors were
designated as adjacent plots within rows for Blocking Scheme IIr and as
adjacent plots within columns for Blocking Scheme IIC. We followed the
custom, initiated by Bartlett (1938), of associating two degrees of
freedom with the covariate in a single-covariate Papadakis analysis.

The CAR used by Method CAR was that for which $C = \frac{1}{2} \bar{N}$, where $\bar{N}$ is the neigh-
bor incidence matrix corresponding to the neighbor designation used in
conjunction with Methods PAP and PAPB; consequently, as was demonstrated
in Section 5.3, Method CAR is closely related to an iterated Papadakis
analysis (with $\phi = 1$) and to the intra-NN analysis. In the random field
approach, REML estimation was used and the model was taken to be the
special case of RFLM-P whose mean function is a paraboloidal surface
and whose covariogram is the $L^1$-geometrically anisotropic exponential
covariogram
\[ C(x,y;\theta_1,\theta_2,\theta_3) = \theta_1 \exp\{-\theta_2 |x| - \theta_3 |y|\} \]  
\[ (\theta_1 > 0, \theta_2 > 0, \theta_3 > 0). \]

This covariogram was chosen over other anisotropic covariograms for two reasons, both of which are consequences of its separability:

1. The necessary computations are less burdensome;
2. The actual plot dimensions \( H_1 \) and \( H_2 \) need not be known (since (5.29) can be reparametrized, by putting \( \hat{\theta}_2 = H_1 \theta_2 \) and \( \hat{\theta}_3 = H_2 \theta_3 \), to the form

\[ C(x,y;\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3) = \theta_1 \exp\{-\hat{\theta}_2 \frac{|x|}{H_1} - \hat{\theta}_3 \frac{|y|}{H_2}\} \]

so that plots can be regarded as square). In Method RF, \( \text{Var}(\hat{\gamma}_{k\ell}) \) was estimated by

\[ \lambda_k \lambda_\ell (X'V^{-1}X)^{-1} \lambda_k \lambda_\ell, \]

which has the same form as the variance of the generalized least squares estimator of \( \gamma_{k\ell} \) if \( V \) was known but with \( V \) replaced by its REML estimate \( \tilde{V} \). Because \( V \) is estimated, it is to be expected that the estimator of \( \text{Var}(\hat{\gamma}_{k\ell}) \) given by (5.30) will tend to be overly optimistic, i.e., smaller than the true variance of \( \hat{\gamma}_{k\ell} \) under this RFIM.

The extra-NN method was not included in this randomization study because it requires additional plots and because, as noted previously, the prescription given by Wilkinson et al. (1983) for obtaining \( \text{Var}(\hat{\gamma}_{k\ell}) \) by this method is unclear. Moreover, the randomization study of Besag and Kempton (1986) suggests that the less complicated intra-NN
analysis performs as well as the extra-NN analysis. Due to their close relationship, results for Method CAR and the intra-NN analysis should be very similar, so the intra-NN analysis was not included in this study either.

Table 5.5 displays $\bar{\text{Emp}}_{100}$ and $\bar{\text{Pre}}_{100}$ (expressed, respectively, as percentages of $\text{Emp}_{100}$ and $\text{Pre}_{100}$ corresponding to Method RGB) for each of the 9 methods of analysis. With few exceptions, the nonclassical methods of analysis were more accurate than Method RGB. The superiority of the nonclassical methods as a whole was most vividly demonstrated for Series LAY$_c$; this is not surprising because Table 5.4 suggests that the choice between rows or columns to serve as the blocks of an actual experiment would have been most critical for the field used by Laycock, i.e., columns would have been a poor choice for blocks but rows would have been a very good choice.

Methods RC and PC-2 were the least successful of the alternatives to Method RCB. In fact, in three series (LAY$_r$, KH$_c$, and W$_c$), Method RC was not as accurate as Method RCB and Method PC-2 was about equally as accurate as Method RCB.

Methods PAP, PAPB, and CAR were generally more accurate than Method RCB. Method PAP was always as accurate or more accurate than Method PAPB, suggesting that the addition of block parameters to Model PAP may often serve no purpose. Generally, the accuracy of Method CAR was about the same as that of Method PAP. For Method PAP (and Method PAPB), $\bar{\text{Pre}}_{100}$ was always somewhat larger than $\bar{\text{Emp}}_{100}$; in contrast,
Table 5.5. $\overline{Emp}_{100}$ and $\overline{Pre}_{100}$ for 9 methods of analysis

<table>
<thead>
<tr>
<th></th>
<th>RCB</th>
<th>RC</th>
<th>PC-2</th>
<th>PAP</th>
<th>PAPB</th>
<th>CAR</th>
<th>BK</th>
<th>PS</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAY_r</td>
<td>$\overline{Emp}_{100}$</td>
<td>100</td>
<td>106</td>
<td>101</td>
<td>85</td>
<td>99</td>
<td>88</td>
<td>80</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>$\overline{Pre}_{100}$</td>
<td>100</td>
<td>106</td>
<td>95</td>
<td>97</td>
<td>104</td>
<td>74</td>
<td>73</td>
<td>110</td>
</tr>
<tr>
<td>LAY_c</td>
<td>$\overline{Emp}_{100}$</td>
<td>100</td>
<td>34</td>
<td>34</td>
<td>42</td>
<td>42</td>
<td>35</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>$\overline{Pre}_{100}$</td>
<td>100</td>
<td>34</td>
<td>34</td>
<td>50</td>
<td>57</td>
<td>39</td>
<td>36</td>
<td>31</td>
</tr>
<tr>
<td>KH_r</td>
<td>$\overline{Emp}_{100}$</td>
<td>100</td>
<td>79</td>
<td>85</td>
<td>76</td>
<td>80</td>
<td>74</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>$\overline{Pre}_{100}$</td>
<td>100</td>
<td>87</td>
<td>90</td>
<td>81</td>
<td>95</td>
<td>63</td>
<td>58</td>
<td>61</td>
</tr>
<tr>
<td>KH_c</td>
<td>$\overline{Emp}_{100}$</td>
<td>100</td>
<td>117</td>
<td>94</td>
<td>52</td>
<td>54</td>
<td>49</td>
<td>41</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>$\overline{Pre}_{100}$</td>
<td>100</td>
<td>124</td>
<td>97</td>
<td>57</td>
<td>67</td>
<td>43</td>
<td>34</td>
<td>92</td>
</tr>
<tr>
<td>$w_r$</td>
<td>$\overline{Emp}_{100}$</td>
<td>100</td>
<td>47</td>
<td>74</td>
<td>71</td>
<td>74</td>
<td>74</td>
<td>78</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>$\overline{Pre}_{100}$</td>
<td>100</td>
<td>44</td>
<td>73</td>
<td>84</td>
<td>90</td>
<td>57</td>
<td>74</td>
<td>65</td>
</tr>
<tr>
<td>$w_c$</td>
<td>$\overline{Emp}_{100}$</td>
<td>100</td>
<td>110</td>
<td>97</td>
<td>75</td>
<td>84</td>
<td>76</td>
<td>69</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>$\overline{Pre}_{100}$</td>
<td>100</td>
<td>114</td>
<td>99</td>
<td>84</td>
<td>98</td>
<td>66</td>
<td>63</td>
<td>167</td>
</tr>
</tbody>
</table>

$\overline{Pre}_{100}$ corresponding to Method CAR was almost always smaller than $\overline{Emp}_{100}$. These results, which are consistent with findings of Wilkinson et al. (1983) and Besag and Kempton (1986), suggest that variance estimates of estimated treatment contrasts obtained by an uniterated Papadakis analysis tend to be conservative whereas analogous estimates obtained by an iterated Papadakis analysis tend to be over-optimistic.
Method BK was usually as accurate or more accurate than all other methods save Method RF, and, as was the case for Method CAR, \( \overline{\text{PRE}}_{100} \) was usually smaller than \( \overline{\text{EMP}}_{100} \). This is somewhat at odds with results reported by Besag and Kempton (1986), who obtained excellent agreement between \( \overline{\text{PRE}}_{20} \) and \( \overline{\text{EMP}}_{20} \) for Method BK.

For some series (e.g., LAY, KH, and W), Method PS was as accurate as Method BK, but for the remaining series Method PS was not even as accurate as Method PAP. Furthermore, the precision of Method PS was sometimes very poor (e.g., Series W).

The most accurate method of this study was Method RF. For some series (e.g., W and W), Method RF was vastly superior to all other methods. Usually, \( \overline{\text{PRE}}_{100} \) was slightly smaller than \( \overline{\text{EMP}}_{100} \), but this is not altogether surprising because of the way that \( \text{Var}(\gamma_{kl}) \) was estimated in Method RF.

The actual values of the REML estimates \( \tilde{\theta}_2 \) and \( \tilde{\theta}_3 \) of the parameters \( \theta_2 \) and \( \theta_3 \) of the assumed RFLM are of some interest, as they indicate the strength of the spatial correlation. The averages of these estimates over the 100 treatment arrangements are displayed in Table 5.6. Note the good agreement between the average values of either \( \tilde{\theta}_2 \) or \( \tilde{\theta}_3 \) corresponding to the two blocking schemes employed for a given data set.

In summary, the results of this randomization study, while not conclusive (due to the relatively few series considered), suggest at least two things. First, there are often more successful methods of analysis than the classical ANOVA corresponding to a randomized blocks
Table 5.6. Average values of REML estimates of $\hat{\theta}_2$ and $\hat{\theta}_3$ over 100 treatment arrangements

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\theta}_2$</th>
<th>$\hat{\theta}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAY_r</td>
<td>0.51</td>
<td>0.36</td>
</tr>
<tr>
<td>LAY_c</td>
<td>0.51</td>
<td>0.33</td>
</tr>
<tr>
<td>KH_r</td>
<td>0.28</td>
<td>0.66</td>
</tr>
<tr>
<td>KH_c</td>
<td>0.26</td>
<td>0.66</td>
</tr>
<tr>
<td>W_r</td>
<td>0.42</td>
<td>0.82</td>
</tr>
<tr>
<td>W_c</td>
<td>0.43</td>
<td>0.83</td>
</tr>
</tbody>
</table>

design. Second, the random field approach can be more successful than all other nonclassical methods of analysis currently known and being used. Certainly the random field approach deserves further attention.
6. REFERENCES


7. ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my major professor, David Harville, for the privilege to work under his guidance and for his encouragement during the completion of this dissertation. I also thank Noel Cressie, whose expertise in spatial statistics was invaluable. Indeed, I am grateful to all other professors in the Statistics Department from whom I took coursework, especially Dean Isaacson, Glen Meeden, Ken Koehler, Krishna Athreya, Wayne Fuller, and Oscar Kempthorne; together they made my stay at Iowa State University a challenging but very enjoyable experience.

I feel privileged to have been associated with professors and graduate students in the A. E. S. group, especially fellow Voyageurs David F. Cox and Paul Hinz. May the wind always be at their back and their frying pan filled with a mess of walleye. I am sure that the statistical techniques and philosophy which they imparted to me will stand me in good stead in the years to come, and I hope that I can someday attain the same level of warmth and concern for students which they always display. I am also grateful for the many fruitful discussions with graduate students Thomas (Tsung-Hua) Lin, Mark Bryan, Steve Miller, and Jorge Morel.

Thanks are due to Darlene Wicks for her eager typing of this dissertation.

Finally, I thank God, my parents Dean and Wendy, my sister Dee, and
my best friend Bridget Tirol for the unfailing love and support which I received from them.

This research was supported in part by Office of Naval Research Contract N0014-85-K-0418.