ESTIMATION OF SPATIAL VARIABILITY
PART III: SOME MODELS FOR SPATIAL AND SPATIO-TEMPORAL PROCESSES

by

Michael Stein
Stanford University

TECHNICAL REPORT NO. 75
AUGUST 1984

STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH (SIMS)

PREPARED UNDER A GRANT TO SIMS FROM

ENVIRONMENTAL PROTECTION AGENCY (EPA)
SLOAN FOUNDATION
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATION OF SPATIAL VARIABILITY
PART III: SOME MODELS FOR SPATIAL AND SPATIO-TEMPORAL PROCESSES

by

Michael Stein
Stanford University

TECHNICAL REPORT NO. 75
AUGUST 1984

STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH (SIMS)

PREPARED UNDER A GRANT TO SIMS FROM

ENVIRONMENTAL PROTECTION AGENCY (EPA)
SLOAN FOUNDATION
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# TABLE OF CONTENTS

## I.0 INTRODUCTION

I.1 Spatial and Spatio-Temporal Processes ............................... 1

I.1.1 Some problems in spatial analysis ............................... 1
I.1.2 The basic model ............................................. 1
I.1.3 More about variograms and interpolation ...................... 4

I.2 Other Approaches to Interpolation and Analysis of Spatial Data. 8

I.3 Summary of Contents .............................................. 12

I.3.1 Estimation of variograms ..................................... 12
I.3.2 Regression and co-kriging .................................. 18
I.3.3 A model for spatial processes observed repeatedly
      in time .................................................... 19
I.3.4 Sequential siting ........................................... 21
I.3.5 Other results ................................................. 21

I.4 Notation ................................................................ 23

## II.0 NON-PARAMETRIC VARIOGRAM ESTIMATION

II.1 Statement of Problem .............................................. 25

II.2 A Generalization of Matheron's Estimator ....................... 26

II.3 Conditions for Existence of Solution to Unbiasedness
    Constraints ..................................................... 30

II.3.1 Necessary and sufficient condition for a solution .......... 30
II.3.2 The case of linear drift ................................... 31

II.4 The Case of Uncorrelated Pairs .................................. 32

II.4.1 An $s^2$-like estimator of $\gamma(d)$ ....................... 32
II.4.2 Using estimators based on residuals when pairs are
      correlated ................................................. 34
II.4.3 Repeated observations in time ................................ 34

II.5 Irregularly Spaced Data .......................................... 37

II.6 Choosing a Form for the Mean Function ........................... 38

II.7 Misspecification of the Relationship Between Covariates and
    the Dependent Variable ........................................ 40

II.7.1 Alternatives to the regression model ....................... 40
II.7.2 Misspecification of regression equation .................... 41

II.8 Application to Landsat Data ...................................... 44

## III.0 PARAMETRIC VARIOGRAM ESTIMATION

III.1 Introduction ..................................................... 67

III.2 Single Scalar Parameter ........................................ 70

III.2.1 Maximum likelihood and modified maximum
       likelihood estimators ....................................... 71
APPENDIX B: Use of the Estimated Covariance Structure Instead of the True Covariance Structure .................. 160

APPENDIX C: Proof of Theorem III.1 ....................... 168

APPENDIX D: An Example Where the Maximum Likelihood Estimator of the Covariance Structure is Not Determined by the Variogram Model .......................... 170

APPENDIX E: Proofs of Results from Section III.4 ............ 172

APPENDIX F: Some Notes on Spatio-Temporal Variograms ....... 176

References ..................................................... 181
<table>
<thead>
<tr>
<th>TABLE OF FIGURES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.1: Plot of visible frequencies at Site 1</td>
<td>45</td>
</tr>
<tr>
<td>II.2: Plot of visible frequencies at Site 2</td>
<td>46</td>
</tr>
<tr>
<td>II.3: Plot of infrared frequencies at Site 1</td>
<td>47</td>
</tr>
<tr>
<td>II.4: Plot of infrared frequencies at Site 2</td>
<td>48</td>
</tr>
<tr>
<td>II.5: Unbiased non-parametric semi-variogram estimates for visible frequencies</td>
<td>50</td>
</tr>
<tr>
<td>II.6: Unbiased non-parametric semi-variogram estimates for infrared frequencies</td>
<td>51</td>
</tr>
<tr>
<td>II.7: Unbiased non-parametric semi-variogram estimates by subregion - Site 1 - Visible frequencies</td>
<td>53</td>
</tr>
<tr>
<td>II.8: Unbiased non-parametric semi-variogram estimates by subregion - Site 1 - Infrared frequencies</td>
<td>54</td>
</tr>
<tr>
<td>II.9: Unbiased non-parametric semi-variogram estimates by subregion - Site 1 - Visible frequencies</td>
<td>55</td>
</tr>
<tr>
<td>II.10: Unbiased non-parametric semi-variogram estimates by subregion - Site 1 - Infrared frequencies</td>
<td>56</td>
</tr>
<tr>
<td>II.11: Interpolators based on four nearest neighbors</td>
<td>59</td>
</tr>
<tr>
<td>II.12: Interpolators based on eight nearest neighbors</td>
<td>59</td>
</tr>
<tr>
<td>II.13: Prediction of interpolation error of eight nearest neighbor &quot;full model&quot; interpolator</td>
<td>61</td>
</tr>
<tr>
<td>II.14: Comparison of semi-variogram estimators</td>
<td>64</td>
</tr>
<tr>
<td>III.1: Set of sites in which second pair of points provides little additional information about $\gamma(d)$</td>
<td>75</td>
</tr>
<tr>
<td>III.2: Set of sites in which second pair of points &quot;doubles&quot; information about $\gamma(d)$</td>
<td>75</td>
</tr>
<tr>
<td>III.3: Pairs of covariance functions in $\mathbb{R}^2$ which do and do not satisfy Equation (4l)</td>
<td>87</td>
</tr>
<tr>
<td>III.4: Parametric estimates of $\theta$ by subregion</td>
<td>106</td>
</tr>
<tr>
<td>III.5: Comparison of semi-variogram estimators</td>
<td>106</td>
</tr>
<tr>
<td>III.6: Performance of non-parametric semi-variogram estimator</td>
<td>108</td>
</tr>
<tr>
<td>III.7: Rescaled non-parametric estimates of $\gamma_v^*(1)$ by subregion</td>
<td>108</td>
</tr>
<tr>
<td>III.8: Performance of non-parametric semi-variogram estimator</td>
<td>110</td>
</tr>
<tr>
<td>III.9: Performance of non-parametric semi-variogram estimator</td>
<td>110</td>
</tr>
<tr>
<td>III.10: Parametric estimates of $\theta$ by subregion</td>
<td>111</td>
</tr>
<tr>
<td>III.11: Comparison of semi-variogram estimators</td>
<td>113</td>
</tr>
</tbody>
</table>
Note: For the purpose of making technical reports of manageable proportions, this thesis has been divided into three parts. In each part, the entire introductory chapter is included. Part I contains Chapter II, "Non-Parametric Variogram Estimation." Part II contains Chapter III, "Parametric Variogram Estimation," and Appendices A through E. Part III, entitled "Some Models for Spatial and Spatio-Temporal Processes," contains Chapters IV-VI and Appendix F.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>III.2.2</td>
<td>Confidence intervals</td>
<td>71</td>
</tr>
<tr>
<td>III.2.3</td>
<td>Implications of the previous section</td>
<td>73</td>
</tr>
<tr>
<td>III.3</td>
<td>Estimation of Parameters: General Case</td>
<td>76</td>
</tr>
<tr>
<td>III.3.1</td>
<td>Maximum likelihood and modified maximum likelihood estimators</td>
<td>76</td>
</tr>
<tr>
<td>III.3.2</td>
<td>Generalized minimum norm quadratic estimators</td>
<td>79</td>
</tr>
<tr>
<td>III.3.3</td>
<td>Another parametric estimator</td>
<td>83</td>
</tr>
<tr>
<td>III.4</td>
<td>Asymptotics</td>
<td>85</td>
</tr>
<tr>
<td>III.4.1</td>
<td>A consistency result which does not depend on the locations of the observations</td>
<td>85</td>
</tr>
<tr>
<td>III.4.2</td>
<td>A discussion of asymptotic properties with multi-parameter models</td>
<td>89</td>
</tr>
<tr>
<td>III.5</td>
<td>Effects of Misspecification of the Mean Function</td>
<td>93</td>
</tr>
<tr>
<td>III.6</td>
<td>Local Parametric Estimators</td>
<td>95</td>
</tr>
<tr>
<td>III.7</td>
<td>Choosing a Parametric Model for the Variogram</td>
<td>100</td>
</tr>
<tr>
<td>III.8</td>
<td>Analysis of Landsat Data</td>
<td>104</td>
</tr>
<tr>
<td>IV.0</td>
<td>A COMPARISON OF CO-KRIGING AND REGRESSION MODELS</td>
<td></td>
</tr>
<tr>
<td>IV.1</td>
<td>Introduction</td>
<td>115</td>
</tr>
<tr>
<td>IV.2</td>
<td>Comparison of Probabilistic Properties of the Two Models</td>
<td>116</td>
</tr>
<tr>
<td>IV.3</td>
<td>Asymptotic Behavior of Estimators Based on the Regression Model When it is Not Valid</td>
<td>120</td>
</tr>
<tr>
<td>IV.3.1</td>
<td>Model with repeated observations in time</td>
<td>120</td>
</tr>
<tr>
<td>IV.3.2</td>
<td>Mean changing in time</td>
<td>122</td>
</tr>
<tr>
<td>IV.3.3</td>
<td>Mean fixed in time</td>
<td>125</td>
</tr>
<tr>
<td>IV.3.4</td>
<td>Interpolation based on the regression model</td>
<td>129</td>
</tr>
<tr>
<td>V.0</td>
<td>KRIGING WITH REPEATED OBSERVATIONS IN TIME AND AN UNKNOWN, VARYING SPATIAL MEAN FUNCTION</td>
<td></td>
</tr>
<tr>
<td>V.1</td>
<td>Description of Model</td>
<td>132</td>
</tr>
<tr>
<td>V.2</td>
<td>Interpolation</td>
<td>134</td>
</tr>
<tr>
<td>V.2.1</td>
<td>Mean changing in time</td>
<td>134</td>
</tr>
<tr>
<td>V.2.2</td>
<td>Mean fixed in time</td>
<td>138</td>
</tr>
<tr>
<td>V.3</td>
<td>Estimation of Time Trends</td>
<td>142</td>
</tr>
<tr>
<td>V.4</td>
<td>Estimation of the Variograms</td>
<td>145</td>
</tr>
<tr>
<td>VI.0</td>
<td>THE EFFECT OF SEQUENTIAL SITING ON ESTIMATION</td>
<td></td>
</tr>
<tr>
<td>VI.1</td>
<td>Description of the Problem</td>
<td>148</td>
</tr>
<tr>
<td>VI.2</td>
<td>A Theorem on the Effect of Sequential Siting</td>
<td>148</td>
</tr>
<tr>
<td>APPENDIX A:</td>
<td>An Asymptotic Result for Estimation of the Variogram in a One-Dimensional Process</td>
<td>154</td>
</tr>
</tbody>
</table>
CHAPTER I: Introduction

I.1 Spatial and Spatio-Temporal Processes

I.1.1 Some problems in spatial analysis

Analysis of processes which take on values in space or in space and time is a rapidly growing field in statistics. Statistical models of spatial phenomenon are used in many fields, including meteorology, air and soil pollution, forestry, agriculture, geology, and satellite imagery. For example, we may be interested in estimating mineral concentrations in the ground based on some set of core samples. Or we may be interested in identifying ground features based on photographs from satellites. Or we may be interested in changes in acid deposition in a region from one year to the next. We will consider the properties and uses of some possible stochastic models of spatial and spatio-temporal phenomena.

I.1.2 The basic model

In this work, we will focus mainly on problems which relate to doing point interpolation of a random field. That is, having observed some quantity \( z(\cdot) \) at \( x_1, x_2, \ldots, x_n \), we want to estimate \( z(x_0) \). While \( x \) will usually just indicate a place, in theory, it could also indicate a time, or a place and a time. The basic model for \( z(\cdot) \) will be

\[
  z(x) = m(x) + e(x) \tag{1}
\]

The mean function, \( m(\cdot) \), will have some specific structure, depending on the problem. It is often just a constant. The disturbance function, \( e(\cdot) \), will be a stochastic process with some sort of stationarity condition. The usual assumption about \( e(\cdot) \) is
\[ \text{Ee}(x) = 0 \text{ for all } x, \text{ and} \]
\[ \frac{1}{2} \text{E}\{(e(x) - e(x'))^2\} = \gamma(x - x'). \] (2)

The function \( \gamma(\cdot) \) is known as the semi-variogram; the "semi" referring to the "\( \gamma \)" in Equation (2). The disturbance function is stationary in the sense that its mean is independent of \( x \), and \( \gamma(\cdot) \) depends only on \( x-x' \). In some situations, we may also assume that \( e(\cdot) \) is an isotropic process; that is, that \( \gamma(\cdot) \) depends only on \(|x-x'|\), the distance between the two points.

The model described by (1) and (2) is commonly used in geostatistics. Basic references in geostatistics are Matheron (1971) and Journel and Huijbregts (1978). This model has also been used to describe spatial distributions of acid deposition (Eynon and Switzer (1983)), petroleum reserves (Davis (1981)), and forest characteristics (Matern (1960)). One of the notable features of this model is the use of the semi-variogram instead of the covariance function. If the covariance function, defined as

\[ \text{Ee}(x) e(x') = C(x - x') \] (3)

exists, then we have the relationship

\[ \gamma(r) = C(0) - C(r). \] (4)

The simplicity of this relationship indicates why the semi-variogram, \( \gamma(\cdot) \), is commonly used instead of the variogram, \( 2\gamma(\cdot) \). Matern (1960) contains an extensive discussion on the properties of stationary and isotropic covariance functions. One reason for using variograms instead of covariances is that it is possible to have situations in which \( \gamma(\cdot) \) is finite but covariances are infinite. For example, a possible form for \( \gamma(\cdot) \) is \( \gamma(r) = b|r| \) for all vectors \( r \). In this case, \( C(r) = \infty \) for all
Even when the covariance function is finite, the semi-variogram may be easier to estimate, as we shall see in I.1.3. In particular, we do not need to estimate $\text{Var} \ z(x) = C(0)$ to estimate $\gamma(\cdot)$.

One model for $z(\cdot)$ that will be used extensively throughout this work is

$$z(x) = a_0 + \mathbf{a}' \mathbf{f}(x) + e(x),$$

(5)

where the disturbance term, $e(\cdot)$, has mean zero and a stationary variogram (Equation (2)). The mean function is $a_0 + \mathbf{a}' \mathbf{f}(x)$, where $\mathbf{f}(x)$ is a known, vector-valued function and $a_0$ and $\mathbf{a}$ are unknown coefficients.

Because of the form of the mean function, Equation (5) will be referred to as the "regression model." In this work, $\mathbf{f}(\cdot)$ will usually be considered to be a vector of observed covariates. For example, if $z(\cdot)$ represents the level of some pollutant, $\mathbf{f}(\cdot)$ may represent weather characteristics. Or if $z(\cdot)$ represents average annual temperature, $\mathbf{f}(\cdot)$ may represent altitude. In what is known as "universal kriging," the mean function is assumed to be a smooth unknown function that can be modelled, at least locally, as a low order polynomial (Delfiner (1975)). This situation creates different problems than when $\mathbf{f}(\cdot)$ is made up of observed covariates, and for the most part, we will not discuss universal kriging in this work.

Now, let us consider the problem of estimating $z(x_0)$ based on $z(x_1), \ldots, z(x_n)$ and the $\mathbf{f}(\cdot)$ field. Suppose we restrict ourselves to linear unbiased estimators of $z(x_0)$. Such estimators are of the form

$$\hat{z}(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha z(x_\alpha),$$

(6)

where the unbiasedness condition, $Ez(x_0) = E\hat{z}(x_0)$, implies that
\[ \sum_{\alpha=1}^{n} \lambda_{\alpha} = 1 \text{, and} \]
\[ \sum_{\alpha=1}^{n} \lambda_{\alpha} \hat{f}(x_{\alpha}) = \hat{f}(x_{0}). \]

We can evaluate the mean squared error of a linear unbiased interpolator in terms of the semi-variogram. We obtain (Matheron (1971)),

\[ E[(\hat{z}(x_{0}) - z(x_{0}))^2 = 2 \sum_{\alpha=1}^{n} \lambda_{\alpha} \gamma(x_{\alpha} - x_{0}) - \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_{\alpha} \lambda_{\beta} \gamma(x_{\alpha} - x_{\beta}). \]

If the variogram is specified, then we can use this equation to find the linear unbiased estimator of \( z(x_{0}) \) with minimum variance, also known as the best linear unbiased estimator (BLUE). This problem can be easily solved using Lagrange multipliers (Journel and Huijbregts (1978), pp. 316-320) to obtain the BLUE of \( z(x_{0}) \). This method of interpolation is known as "kriging" in geostatistics. One of the desirable properties of kriging is that we obtain an expression for the variance of the error of our interpolator along with the interpolator. Goldberger (1962) considered essentially the same problem with applications to econometrics. If the disturbance field, \( e(\cdot) \), is Gaussian, then the BLUE of \( z(x_{0}) \) will also be the uniformly minimum variance unbiased estimator (UMVUE) of \( z(x_{0}) \). Of course, there may be biased and/or non-linear estimators with smaller mean squared error than the BLUE. However, to evaluate biased linear estimators we need to specify \( \text{Var} \ z(x) = C(0) \), and to evaluate non-linear estimators, we need to know more about the distribution of the disturbance field than its covariance structure. Therefore, in this work, we will only consider linear unbiased estimators.

I.1.3 More about variograms and interpolation

In this section, we will briefly review some of the basic facts about variograms and their relationship to kriging interpolators.
It is well known that a covariance function must be positive definite (Matern (1960)); i.e., for any set of sites \( x_1, \ldots, x_n \) and weights \( \lambda_1, \ldots, \lambda_n \), the covariance function,

\[
C(x, x') = E[z(x) - Ez(x)](z(x') - Ez(x'))
\]

must satisfy

\[
\sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C(x_\alpha, x_\beta) \geq 0. \tag{9}
\]

Similarly, a semi-variogram

\[
\gamma(x, x') = \frac{1}{2} E \left[ (z(x) - Ez(x)) - (z(x') - Ez(x')) \right]^2
\]

must satisfy the slightly weaker condition

\[
\begin{cases}
\text{if } \sum_{\alpha=1}^{n} \lambda_\alpha = 0, \text{ then } \\
- \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta \gamma(x_\alpha, x_\beta) \geq 0
\end{cases} \tag{10}
\]

(Matheron (1971)). The function \(-\gamma(\cdot, \cdot)\) is said to be "conditionally positive definite."

In order to introduce some of the properties of variograms and kriging, we will consider the simple case of an isotropic variogram and a mean function which is an unknown constant, or \( m(x) \equiv m \). First, we should note that in many instances, while the local average of a process may remain relatively constant over short distances, it may vary considerably over longer distances. To protect against this possibility, we can make our interpolator of \( z(x_0) \) depend only on \( z(x_\alpha)'s \) for which \( |x_\alpha - x_0| \) is relatively small. We should note that in some circumstances, it may not matter whether slow variations in the \( z(\cdot) \) field are modelled as changes in the mean field or as variations in the disturbance field over long distances. However, one of our main principles
of point interpolation will be that our procedure should try to be independent of any modelling of the process over long distances. The reasoning behind this principle is that interpolation is essentially a problem of modelling the local variations in a spatial process; thus, we should focus our modelling efforts on understanding the local behavior of the process, and not worry about what happens over long distances. This reasoning may not be valid if there are strong periodicities in the process. In this case, observations far away from $x_o$ may give us substantial information about $z(x_o)$ which is not contained in nearby observations.

The semi-variogram is a measure of local variation, and thus its use fits in with the above principle. To estimate the semi-variogram over a short distance $d$, we can take all $n_d$ pairs of points separated by $d$ and use

$$
\hat{\gamma}(d) = (2n_d)^{-1} \sum_{|x_i-x'_i|=d} (z(x_i) - z(x'_i))^2.
$$

This simple, unbiased estimator was suggested by Matheron (1971). Note that this estimator does not depend on the unknown constant mean, $m$, in any way. Also, if $m$ actually varies slowly in space, but does not vary significantly over the distance $d$, then $\hat{\gamma}(d)$ will not be greatly affected. On the other hand, in order to estimate $C(d)$, the covariance at distance $d$, we need to have an estimate of $m$, since

$$
C(d) = E(z(x)-m)(z(x')-m), \text{ where } |x-x'|=d.
$$

Therefore, to estimate $C(d)$, we need to estimate the global mean, which is a quantity we do not need to know in order to compute the variance of linear unbiased estimators. So we see that $\gamma(\cdot)$, and not $C(\cdot)$, is the appropriate function
simple calculation shows that the BLUE is given by
\[ \hat{z}(0) = \frac{1}{2} rz(h) + (1 - \frac{1}{2} r)z(2h). \] (12)

If \( r=1 \), which will occur if \( \gamma(\cdot) \) exhibits a pure nugget effect, then we get
\[ \hat{z}(0) = \frac{1}{2} (z(h) + z(2h)). \]

If \( 1 < r < 2 \), then \( z(2h) \) will get a smaller, but still positive, weight than \( z(h) \). For \( h \) small, this behavior for \( \gamma(\cdot) \) will occur if \( \gamma'(0^+) > 0 \) and \( \gamma''(0^+)< 0 \). A commonly used example is the "exponential" semi-variogram,
\[ \gamma(d) = c_1 (1 - e^{-c_2 d}). \]

If \( r=2 \), then \( \hat{z}(0) = z(h) \); that is, \( z(2h) \) gets no weight in estimating \( z(0) \). This phenomenon will occur if \( \gamma(d) = c_1 d \). If \( 2 < r < 4 \), then \( z(2h) \) will get a negative weight. Examples of when \( 2 < r < 4 \) for small \( h \) are
\[ \gamma(d) = c_1 (1 - e^{-c_2 d^2}), \text{ and} \]
\[ \gamma(d) = c_1 d^\alpha, \quad 1 < \alpha < 2. \]

In these cases, there is no nugget effect, \( \gamma'(0^+) = 0 \), and \( \gamma(\cdot) \) is differentiable. Therefore, it is not surprising that we get a negative weight for \( z(2h) \), since if \( z(2h) - z(h) \) is positive, we would guess that \( z(h) - z(0) \) is also positive. If \( r=4 \), then \( z(0) = 2z(h) - z(2h) \); that is, we can estimate \( z(0) \) without error. This situation will occur if \( \gamma(d) = c_1 d^2 \). A quadratic variogram is an extreme case, as it is equivalent to all of the observations lying on a single line with probability one. It is not possible to have \( r > 4 \), as \( -\gamma(\cdot) \) will then not be conditionally positive definite (Equation (10)).

The behavior of \( \gamma(d) \) as \( d \) approaches infinity is another basic feature of a semi-variogram. As we expect the relationship between \( z(x) \)
and \( z(x') \) to weaken as \( |x-x'| \) increases, we will usually assume that \( \gamma(\cdot) \) is monotonically increasing in \( d \). If
\[
\lim_{d \to \infty} \gamma(d) = \infty,
\]
then we also have \( C(d) = \infty \) for all \( d \). If
\[
\lim_{d \to \infty} \gamma(d) = \gamma(\infty) < \infty,
\]
then \( \gamma(\infty) \) is known as the "sill value;" it just equals \( \text{Var}(z(x)) = C(0) \). The distance at which \( \gamma(\cdot) \) stops increasing is known as the "range" of the semi-variogram. Observations at sites farther apart than the range are uncorrelated.

I.2 Other Approaches to Interpolation and Analysis of Spatial Data

While the model described in I.1.2 is extensively used in geological applications, it is not used so much in other fields where statistical models of spatial processes are needed. Spatial statistics have been used in a variety of fields other than geology, including geography (Cliff and Ord (1975)), archaeology (Hodder and Orton (1976)), and agriculture (Bartlett (1975)). Recent books on spatial statistics include those by Ripley (1981), Bennett (1979), Bartlett (1975) and Cliff and Ord (1981). We will briefly discuss two often used models for spatial processes.

A model which has been frequently used in geography and ecology is one in which the conditional distribution of each observation given all other observations is considered (Besag (1974)). If our process is assumed to be Gaussian, then it is sufficient to specify the conditional mean and variance of each observation. If \( x_1, \ldots, x_n \) represent the random values of a spatial process at \( n \) sites and \( x_1, \ldots, x_n \) their
to estimate if all we want to do is evaluate the variance of linear unbiased interpolators.

Since we are interested in the local behavior of the \( z(\cdot) \) field, it follows that we are most interested in \( \gamma(\cdot) \) over short distances. We see, by definition, that \( \gamma(0) = 0 \). However, we may have

\[
\gamma(0^+) = \lim_{d \to 0} \gamma(d) > 0.
\]

In geostatistical parlance, the size of this discontinuity at the origin is known as the "nugget effect." It may be due to microstructure in the \( z(\cdot) \) field, that is, variations in the process over very short distances, or due to measurement error, or both. If \( \gamma(d) \) is some constant for all distances greater than zero, then we have what is known as a "pure nugget effect." Alternatively, we could say \( z(\cdot) \) is a white noise process. In this case, the process exhibits no spatial structure, and our interpolator will be the same at all places at which there is not an observation: the average of all observations.

Another important feature of the semi-variogram over short distances is \( \gamma'(0^+) \). In models used in practice, \( \gamma'(0^+) \) is almost always well-defined and finite. Let us consider a process \( z(\cdot) \) with constant mean and an isotropic semi-variogram which has no nugget effect. Then, if \( \gamma'(0^+) > 0 \), \( z(\cdot) \) will be continuous but not differentiable. If \( \gamma'(0^+) = 0 \), then \( z(\cdot) \) will be differentiable.

The following simple example will illustrate the impact of the behavior of \( \gamma(\cdot) \) at the origin on kriging interpolators. Suppose \( z(\cdot) \) is a process in \( \mathbb{R}^1 \) with fixed, unknown mean and stationary variogram. We observe \( z(h) \) and \( z(2h) \), where \( h \) is "small," and wish to estimate \( z(0) \). We will find the BLUE of \( z(0) \) as a function of \( r = \gamma(2h)/\gamma(h) \). A
respective observed values, then one model for the conditional means and
variances is
\[
E(X_i | x_j, j \neq i) = E(x_i) + \sum_{j \neq i} \beta_{i,j}(x_j - E(x_j)), \text{ and} \tag{14}
\]
\[
\text{Var}(X_i | x_j, j \neq i) = \sigma_i^2 \tag{15}
\]
"(Besag (1975)). If we did not actually observe \( x_n \), then we could, in
theory, use this model to estimate it (by the modelled conditional expec-
tation) and obtain an expression for the error of this estimator (by the
modelled conditional variance). Thus, this model can be used to produce
point interpolators along with estimates of the variance of the interpo-
lation error. If we are only interested in values of the process on a
regular grid, then there are natural ways to model the \( \beta_{i,j} \)'s and \( \sigma_i \)'s.
For example, if the data sites form a lattice, then it is reasonable to
assume all of the \( \sigma_i \)'s are equal except for those sites near the outer
edge of the lattice. However, in this work, we are interested in modelling processes which are not necessarily observed at sites on a grid,
and even if the observation sites are on a grid, we want to be able to
do point interpolation at sites not on the grid. Some work has been
done on applying the model in (14) and (15) to irregularly spaced data
(Besag (1975)), but the models for the \( \beta_{i,j} \)'s and \( \sigma_i \)'s, while qualita-
tively reasonable, are somewhat arbitrary. The kriging formulation,
in which we try to measure the degree of association in a process at all
distances (through the variogram), seems more natural when we are inter-
ested in the value of the process at all sites, and not just those on a
regular grid.

Another model which can be used to do point interpolation with
spatial data is splining. Wegman and Wright (1983) contains a review of
the uses of splines in statistics. "Smoothing" splines try to find a smooth surface which pass near the observations. For example, suppose we observe \( z(x_1), \ldots, z(x_n) \), where \( x = (x^1, x^2) \) is in \( \mathbb{R}^2 \). Then the "thin plate" smoothing spline will be a function \( f(*) \) which minimizes

\[
\frac{1}{n} \sum_{j=1}^{n} (f(x_j) - z(x_j))^2 + \lambda J_m(f), \quad \text{where}
\]

\[
J_m(f) = \sum_{k=1}^{m} \int_{\mathbb{R}^2} \left( \frac{\partial^m f}{\partial x^k \partial x^l} \right)^2 \, dx^1 \, dx^2.
\]  

The parameters \( \lambda \) and \( m \) are usually chosen by a cross-validation procedure (Wahba (1979)). As splining is usually used, no attempt is made to predict the variance of \( f(x_o) - z(x_o) \) at a specific place \( x_o \), which is something the kriging formulation allows us to do. While kriging and splining appear to be rather different approaches to the problem of point interpolation, they in fact can produce equivalent interpolators. It turns out that some splines give the same estimation surfaces as particular cases of "universal kriging," which is described by Delfiner (1975). A specific example of such an equivalence is given by Dubrul (1983); Wegman and Wright (1983) also mention some equivalences between splining and universal kriging.

We could also consider doing point interpolation using methods of non-parametric regression not specifically aimed at spatial data. Such methods include nearest neighbor estimators, local linear estimators (see Stone (1977) and the references therein), and projection pursuit (Friedman and Stuetzle (1981)). While these methods may produce reasonable interpolators, they do not, in general, allow us to estimate the variance of the interpolation error at a particular place.
I.3 Summary of Contents

I.3.1 Estimation of variograms

To calculate the variance of the error of a linear unbiased interpolator of $z(x_o)$, it is necessary to specify the variogram (see Equation (8)). However, variograms usually cannot be specified a priori, but must be estimated form the data. The problem of estimating the variogram is addressed in Chapters II and III. We will use the regression model given in Equation (5). To avoid confusion, we will usually assume the variogram is isotropic, although the procedures considered can be easily modified to apply to stationary but anisotropic variograms.

In Chapter II, we will consider non-parametric estimators of the variogram. The estimators will be non-parametric in the sense that $\gamma(\cdot)$ is estimated at a specific distance $d$ in a way that does not require any assumptions about $\gamma(\cdot)$ at distances other than $d$. Also, since we do not need to estimate the regression coefficients directly to use unbiased linear interpolators, we will seek variogram estimators which do not depend on explicitly estimating the regression coefficients. Matheron's estimator of $\gamma(d)$, given in (II), satisfies these properties when the mean is a constant. However, it cannot be used if the mean function is not a constant, as $z(x)-z(x')$ will no longer have mean zero. When the mean is given by a regression equation, an unbiased estimator of $\gamma(d)$ which does not depend on estimating the regression coefficients is given in II.2. It is based on a weighted sum of squares of all observed values of $z(x)-z(x')$ which satisfy $|x-x'|=d$. This estimator cannot be used when the mean function includes an arbitrary polynomial in $x$ of degree greater than zero (see II.3); thus it is not
applicable to what is commonly called "universal kriging" by geostatisticians (Delfiner (1975)). In II.4, we consider how we can take advantage of an assumption that some pairs of points separated by $d$ are uncorrelated with other pairs. This situation can occur with a space-time process with repeated observations in time which are sufficiently separated in time so that temporal uncorrelatedness can be assumed.

One of the main reasons for wanting to estimate $\gamma(\cdot)$ non-parametrically is that it allows us to plot $\hat{\gamma}(d)$ as a function of $d$. These plots can serve as guides to the structure of $\gamma(\cdot)$ without making any a priori assumptions (beyond stationarity or isotropy) as to what that structure should be. Therefore, such plots are very useful in choosing a parametric model for the variogram which fits the data. In Section II.6, we discuss how these plots can also be used to help choose a subset of regressors from a set of potential regressors.

One of the assumptions of the regression model is that the parametric form of the mean function is correctly specified. In some circumstances, there may be physical reasons for believing a specific parametric model for the mean function. However, parametric regression functions are often used even when the relationship between the variables is assumed to be only approximately linear. In spatial processes, there is the added problem that the relationship between the covariates and the variable of interest may change in space. In II.7, the effects of misspecification of the mean function on both linear unbiased interpolators and on non-parametric variogram estimators are considered.

To illustrate some of the ideas and methods of this chapter, some computations are done using Landsat data in II.8. In these data, integrated surface reflectance in four different frequency bands is obtained.
at each of 400 different pixels in a 16 x 25 rectangular region. The analysis focuses on two frequencies at a time, as the frequencies can be naturally separated into pairs of visible and infrared. The results indicate that much of the spatial variation in the energies at one frequency can be explained by the other frequency, and that if the first frequency had a missing value at a pixel, estimation of the missing value can be improved by using the second frequency rather than just taking an average of nearby values of the first frequency. The unbiased non-parametric estimator of the semi-variogram given in II.2 appears to be a quite noisy estimator. With these data, it seems that we are better off using an estimator based on residuals which will be slightly biased, but has lower variance.

In Chapter III, parametric methods of estimation of the variogram are considered. One reason parametric estimators are needed is to obtain an estimator of \( \gamma(\cdot) \) which yields a permissible semi-variogram, i.e., we want \( \hat{\gamma}(\cdot) \) to be conditionally positive definite (see Equation (10)). If the plot of a non-parametric estimator of \( \gamma(d) \) v. \( d \) is smoothed, then the resulting function will not, in general, be a permissible semi-variogram. It is difficult to specify a class of variograms non-parametrically, but easy to specify a parametric class of variograms. Hence, the plot of \( \hat{\gamma}(d) \) v. \( d \) is almost always fit by choosing a parametric class of semi-variograms which seems to be appropriate based on the plot, and then estimating the parameters by some means. Within the geostatistical literature, the parameters are usually estimated by ad hoc procedures, or by fitting the plot of \( \hat{\gamma}(d) \) v. \( d \) by (weighted) least squares. The question of when such procedures provide adequate estimators of the
variogram is quite unclear, although there must be situations in which more efficient estimators of the parameters would be of practical value. Therefore, in Chapter III, we will consider applying some of the traditional methods of estimation to the problem of estimating the parameters in a variogram model. A substantial literature exists concerning the estimation of parameters in a covariance matrix; Harville (1977) contains a review of this problem. Methods discussed in this chapter include maximum likelihood (ML), modified or restricted maximum likelihood (MML), and minimum norm quadratic estimators. Such methods can be much more efficient than ad hoc or least squares fits to the non-parametric estimator.

In Section III.2, the problem of estimating a semi-variogram which is specified up to a scalar multiple is considered. That is, we assume

$$\gamma(\cdot) = \theta \gamma_0(\cdot),$$

(18)

where $\gamma_0(\cdot)$ is specified and $\theta$ is to be estimated. When the disturbance field is Gaussian, the modified maximum likelihood estimator (MMLE) is the UMVUE (III.2.1). Also, it is shown to be the same as the maximum likelihood estimator (MLE) times a constant in this special case (III.2.1).

Based on the MMLE of $\theta$, we can obtain exact confidence intervals for $z(x)$, where $x$ is a place at which $z(\cdot)$ is not observed (III.2.2). The distribution of the MMLE of $\theta$ does not depend on the locations of the observation sites as long as they are distinct. Thus, the modified maximum likelihood estimator will converge to the correct value (assuming the model in (18) is correct) as the number of observations gets large, even if the observations are all contained in a finite region.

In contrast, the efficiency of non-parametric estimators is highly
dependent on the locations of the observation sites (see III.2.3) and are much harder to use when the sites are irregular in space (see II.5).

In Section III.3, parametric estimators of the semi-variogram are considered for more general parametric models. In III.3.1, the MLE and MMLE are given. In III.3.2, what Rao (1979) calls the MINQE(U,I) (minimum norm quadratic estimator which is unbiased and invariant) is reviewed. Minimum norm estimators are only defined for models where the covariance matrix is linear in the parameters. In this section, a natural generalization of the MINQE(U,I) is given for general parametric models of covariance structures. This generalized MINQE(U,I), while no longer having an explicit solution as does the MINQE(U,I), is still easier to compute than the MLE or MMLE. Both the MINQE(U,I) and its generalization depend on some a priori estimate of the parameters. This a priori estimate can be obtained by using an eyeball fit to the plot of the non-parametric estimator of the semi-variogram versus distance.

In Section III.4, asymptotic properties of these estimators are considered. The asymptotics are difficult, since the joint distribution of the observations depends not only on their number but also on their locations. However, in III.4.1, a consistency result, independent of the locations (as long as they are distinct), is shown for a very special case of the generalized MINQE(U,I). We assume

$$\gamma(\cdot) = \gamma_1(\cdot) + \theta \gamma_2(\cdot),$$  \hspace{1cm} (19)

where $\theta$ is the parameter to be estimated and $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ are specified and satisfy the condition that $\gamma_2 - \alpha \gamma_1(\cdot)$ is a permissible semi-variogram for some $\alpha > 0$. In III.4.2, heuristic arguments are given for asymptotic covariances of the MMLE and MINQE(U,I). When
there is only one parameter, we show that the hypothesized asymptotic variance of the MMLE is less than the hypothesized asymptotic variance of the generalized MINQE(U,I), with equality if and only if the a priori estimate of the parameter used to compute the generalized MINQE(U,I) is equal to the true value of the parameter.

In III.5, the effects of misspecification of the mean function on parametric estimators are considered. We see that, in an important way, the parametric estimators are more sensitive to misspecifications of the mean function than the non-parametric estimators. However, perhaps the main reason ML, MML, and minimum norm quadratic estimators have largely been ignored by geostatisticians is the computational and numerical problems that occur when these methods are used with large data sets. One way to alleviate these problems is to compute the parametric estimator in a set of subregions, and then obtain a global estimator of the parameters by taking some sort of average of the local estimators (III.6). This method has the additional benefit of protecting against the effect of some types of misspecification of the mean function. In particular, by doing estimation locally, we protect ourselves against slow spatial variations of the regression coefficients. Local estimators can also be used as a way to detect non-stationarities in the variogram. The greater efficiency of parametric estimators should allow for greater sensitivity in picking up non-stationarities than non-parametric estimators. In III.7, strategies for picking parametric models for the variogram are considered. In III.8, the local parametric estimator of a variogram specified up to a scalar multiple (Equation (18)) is tested using the Landsat data. Specifically, the parameter is first estimated within each of 20 4x5 pixel subregions, and these local estimates are averaged to
obtain a global estimate of the parameter. These estimates are easy and quick to compute, as the congruence of the subregions greatly reduces the calculations. The degree to which the parametric estimator provides information about the variogram not contained in non-parametric estimators depends on the strength of local correlations and on the locations of the observations. In any case, computing the parametric estimator does no harm, and if it gives a much different variogram than an eyeball fit to a non-parametric estimator, we would at least know that the non-parametric estimator cannot be trusted.

1.3.2 Regression and co-kriging

In Chapters II and III, covariates are considered as fixed regressors of the variable of interest. An alternative method of using covariates is to consider them as a jointly stationary stochastic process. That is, we consider \((f(\cdot)', z(\cdot))\) to be a stationary, vector-valued stochastic process. We could then estimate all covariances between the variables, and then obtain an interpolator by using the BLUE based on these estimated covariances. This technique of interpolation is known as co-kriging, see Journel and Huijbregts (1978), pp. 324-26. In Chapter IV, the regression and co-kriging models are compared. While the co-kriging model will be inappropriate for situations where the covariates cannot reasonably be modelled as a stationary stochastic process, the co-kriging model does not require that the covariates be observed at specific sites and will produce better interpolators than the regression model in some circumstances. The main result of this chapter is that if the co-kriging model is correct, then the regression model will only strictly make sense if there exists a vector \(a\) such that
\[ \text{Cov}(\tilde{f}(x), z(x') - \alpha' \tilde{f}(x')) = 0 \text{ for all } x, x'. \] (20)

This condition can be interpreted as meaning that the covariate and disturbance fields are uncorrelated. If (20) holds, then we roughly have that all of the information about \( z(x) \) contained in the \( \tilde{f}(\cdot) \) field can be obtained from \( \tilde{f}(x) \), which is an assumption the regression model implicitly makes.

Because of the simplicity and familiarity of the regression model, we may decide to use it even when there is no \( \tilde{\alpha} \) which satisfies (20). Suppose we observe many independent replications of \( z(\cdot) \) and \( \tilde{f} \) at a fixed set of sites. For example, the replications may be observations at different times. Suppose we use the regression model, and estimate the regression coefficients by a generalized least squares procedure. Then, in general, what our estimators of the regression coefficients converge to will depend on both the weighting matrix used in the generalized least squares procedure and on the locations of the fixed sites (IV.3.2 and IV.3.3). However, if there is an \( \tilde{\alpha} \) which satisfies (20), then the estimators of the regression coefficients will converge to that \( \tilde{\alpha} \) for any weighting matrix and set of sites. Even when there is no such \( \tilde{\alpha} \), we may still be able to produce reasonable (although not optimal) interpolators and accurate estimates of the variance of the interpolation error using the regression model (IV.3.4).

I.3.3 A model for spatial processes observed repeatedly in time

In Chapter V, we consider spatio-temporal processes which can be modelled as

\[ z(x, t) = m(x) + \mu(t) + e(x, t). \] (21)

We assume that this process is observed at a fixed set of sites at times
$t_1, \ldots, t_k$. The sequence $\mu(t_1), \ldots, \mu(t_k)$ is taken to be an arbitrary, deterministic time series and $e(x,t)$ is assumed to be uncorrelated from one time to another. Two problems related to this model are considered. In V.2.1, we consider estimation of $z(x_o, t_\alpha')$, where $\alpha$ is in $\{1, \ldots, k\}$ and $x_o$ is not one of the fixed observation sites. Clearly, we need to impose some structure on $m(x)$. The approach taken here is to assume $m(x)$ has a fixed mean and a stationary variogram. Using this model, the BLUE of $z(x_o, t_\alpha')$ is given. An asymptotic result is given as $k$, the number of observations in time, tends to infinity. This asymptotic expression can be easily interpreted as having a term which interpolates the $m(\cdot)$ field and a second term which interpolates the $e(\cdot, t_\alpha)$ field. In V.2.2, similar results are developed when $\mu(t)$ is assumed to be fixed in time.

The second problem considered in Chapter V is the estimation of the average change in the $z(\cdot, \cdot)$ field from time $t_1$ to $t_2$ in some specified region $R$. The main idea of this section is that we should estimate this quantity by computing $z(x_i, t_1) - z(x_i, t_2)$ for $i = 1, \ldots, k$, and then estimate the average value of this "difference" field in the region $R$. By using this procedure, we only have to model the difference field, which is independent of $m(x)$, so that the problem of modelling $m(x)$ is eliminated. Alternatively, we could separately estimate the average of $z(\cdot, t_1)$ and $z(\cdot, t_2)$ in $R$, take the difference of these estimators to obtain our estimator of the average difference, and assume the estimators of the average in $R$ within each time are uncorrelated to obtain the variance of the estimator of the average difference. This procedure essentially assumes that $m(x)$ is a constant, which will cause us to overestimate the variance of the estimator of the average difference if $m(x)$, in
fact, varies in $x$. The situation is similar to that of doing a paired t-test rather than a two-sample t-test.

I.3.4 Sequential sitting

In some situations, the locations of all of the observations of a spatial process are not determined before any observations are taken. Instead, an initial sample may be taken, and further locations are chosen based on the previous observations. A common example occurs when it is important to estimate $z(*)$ well in regions where it takes on large values, a problem which is crucial in pollution monitoring. After an initial set of observations are taken, further observations are taken near those sites with relatively high values of $z(*)$. Using the regression model in Equation (5), let us define a contrast as a linear combination of the data which has expectation zero for all possible values of the regression coefficients. We will assume the disturbance term is Gaussian and the variogram has been specified. In Chapter VI, we show that as long as the locations of future observations are functions of the contrasts of previous observations, the distribution of the error of a kriging interpolator will be unaffected by the fact that the locations of the observations were not all selected before any observations were taken.

I.3.5 Other results

In Appendix A, we consider the problem of estimating $\gamma'(0)$ for a Gaussian process in an interval of $\mathbb{R}^1$ when $\gamma(*)$ exhibits linear behavior at zero. In particular, a simple, non-parametric estimator of $\gamma'(0)$ is compared to the UMVUE when $\gamma(*)$ can be specified up to a multiplicative constant. As the observations in the interval get dense, the first two terms in an asymptotic expansion of the mean squared error
of the non-parametric estimator are given. As a corollary, we obtain that the non-parametric estimator has the same asymptotic efficiency as the optimal parametric estimator.

In Appendix B, we consider the effect of using estimated covariances instead of actual covariances to produce interpolators and estimators of regression coefficients. In particular, consider the regression model given in Equation (5), and assume the disturbance term is Gaussian. Let us assume the covariance structure is estimated based only on the contrasts of the data, which is true for the MLE, MMLE, and MINQE(U, I). If we are doing interpolation or estimating a regression coefficient, then the variance of a "pseudo-UMVUE" which uses the estimated covariance structure as if it were the actual covariance will be greater than the variance of the true UMVUE using the correct covariance structure. This result is not trivial since the pseudo-UMVUE need not be unbiased. In fact, what is actually proved in this appendix is the stronger result that the conditional variance of the pseudo-UMVUE given all contrasts is equal to the variance of the UMVUE.

In Appendix F, we show some results about spatio-temporal variograms. One result, Theorem F.3, is a basic result about the relationship between variograms and covariance functions which does not appear to have ever been proved before. This theorem says that if a spatial semi-variogram, \( \gamma(\cdot) \), has a finite sill, \( L \), then \( L - \gamma(\cdot) \) is a permissible spatial covariance function; that is, it is a positive definite function (see Equation (9)).
I.4 Notation

We now review some of the notational conventions to be used in this work.

In denoting vectors and matrices, the subscripts \( i \) and \( j \) will have special meanings. That is, \((a_i)\) will denote a column vector whose \( i \)'th element is \( a_i \). And, \((a_{ij})\) will denote a matrix whose element in the \( i \)'th row and \( j \)'th column will be \( a_{ij} \). This notation is used extensively in Chapter III.

For a matrix \( B \), we will define \( \mathcal{M}(B) \) to be the set of all matrices of full rank whose rows form a basis for the left null-space of \( B \). Thus, if \( B \) has \( n \) rows and is of rank \( k < n \), then \( A \in \mathcal{M}(B) \) means \( AB = 0 \), and \( A \) has \( n-k \) rows and rank \( n-k \). This notation is frequently used in Chapters III and VI where it is important to consider the contrasts of a random vector. If \( Z \) is a random vector with mean \( F'\mathbf{g} \), where \( F \) is a known matrix of regressors and \( \mathbf{g} \) a vector of unknown regression coefficients, then a contrast is a linear combination of the components of \( Z \) that has an expected value of zero for all possible values of \( \mathbf{g} \). If \( C \in \mathcal{M}(F') \), then \( CZ \) forms a basis for all contrasts. That is, if \( y'Z \) is a contrast, then \( y'Z = \omega'CZ \) for some \( \omega \).

Other notation:

- \( \text{BLUE} \) = best (minimum variance) linear unbiased estimator.
- \( \text{Cov}(\mathbf{x}, y') = E[(\mathbf{x} - E\mathbf{x})(y - Ey)'] \).
- \( e_n \) or \( e_n^\prime \) = a vector of 1's of length \( n \). The \( n \) will be suppressed if it is clear from context.
- \( I_n \) = the identity matrix of order \( n \).
MINQE(U, I) = minimum norm quadratic estimator which is unbiased ("U") and invariant ("I") with respect to changes in the regression coefficients. See III.3.2.

MLE = maximum likelihood estimator.

MMLE = modified maximum likelihood estimator. Also known as restricted maximum likelihood estimator. See III.3.1.

N(μ, Σ) = multivariate normal distribution with mean vector μ and covariance matrix Σ.

0 = a matrix of all zeroes.

Positive definite (p.d.): A square, symmetric matrix B is said to be positive definite if \( x' B x \geq 0 \) for all \( x \) with equality only if \( x = 0 \). A function \( f(x, x') \) is said to be positive definite if

\[
\sum_{i,j=1}^{n} \omega_i \omega_j f(x_i, x_j) \geq 0
\]

for all \( x_1, \ldots, x_n \), all \( \omega_1, \ldots, \omega_n \), and all \( n \).

Positive semi-definite (p.s.d.): A square, symmetric matrix is positive semi-definite if \( x' B x \geq 0 \) for all \( x \). We also define, for square, symmetric matrices \( P \) and \( Q \),

\( P \preceq Q \) if \( P - Q \) is p.s.d.

UMVUE = uniformly minimum variance unbiased estimator.

Var \( X = \text{Variance of } X = E(X - EX)^2 \).

Finally, we will use the following standard definitions:

\( x_n = o(y_n) \) means \( x_n / y_n \to 0 \) as \( n \to \infty \), and

\( x_n = O(y_n) \) means \( x_n / y_n \) remains bounded as \( n \to \infty \).
CHAPTER IV: A Comparison of Co-Kriging and Regression Models

IV.1 Introduction

In Chapters II and III, covariates were considered to be fixed variables which affect the mean of the variable of interest (see Equation (1) of Chapter II). In Section II.6.1, other ways to model the relationship between these variables were also considered. In this chapter, we will look more carefully at models in which the covariates are also assumed to be stochastic. In particular, letting $z(\cdot)$ be the variable of interest and $\mathbf{f}(\cdot) = \{f_1(\cdot) \ldots f_p(\cdot)\}$' the covariates, we will assume $\mathbf{z}(\cdot)$ is a jointly stationary, vector-valued random field. That is,

$$
E\left[ \begin{pmatrix} z(x) \\ f(x) \end{pmatrix} \right] = \begin{pmatrix} \mu_x \\ \mu_f \end{pmatrix}
$$

(1)

$$
\begin{aligned}
\frac{1}{2} E \left[ (z(x) - z(x'))^2 \right] &= \gamma_{zz}(x - x'), \\
\frac{1}{2} E \left[ (z(x) - z(x'))(f_r(x) - f_r(x')) \right] &= \gamma_{f_r z}(x - x'), \text{ and} \\
\frac{1}{2} E \left[ (f_r(x) - f_r(x'))(f_s(x) - f_s(x')) \right] &= \gamma_{f_r f_s}(x - x').
\end{aligned}
$$

(2)

The $\gamma_{f_r z}(\cdot)$'s and $\gamma_{f_r f_s}(\cdot)$'s for $r \neq s$ are known as "cross-variograms." If all of the variograms and cross-variograms are specified, then best linear unbiased interpolation of the $z(\cdot)$ field based on this model is known as co-kriging (Journel and Huijbregts, 1978, p. 324). That is, if we observe $z(x_1), \ldots, z(x_n)$, and $f_r(x_{rl}), \ldots, f_r(x_{rn})$, $r = 1, \ldots, p$, then an unbiased estimator of $z(x_o)$ will be of the form

$$
\hat{z}(x_o) = \sum_{\alpha=1}^{n} \gamma_{\alpha} z(x_\alpha) + \sum_{r=1}^{p} \sum_{j=1}^{n_r} \gamma_{rj} f_r(x_{rj}),
$$

(3)

where

$$
\sum_{\alpha=1}^{n} \gamma_{\alpha} = 1, \quad \text{and} \quad \sum_{j=1}^{n_r} \gamma_{rj} = 0 \quad \text{for all } r.
$$

(4)
Then, \( \text{Var}(z(x_o) - \hat{z}(x_o)) \) can be evaluated by the formulas in (2), and thus, the best linear unbiased estimator of \( z(x_o) \) can be obtained.

There are some basic differences between this model and the "regression" model for the covariates used in the previous two chapters. By considering the covariates as fixed regressors, there is no need to model them as a stochastic process as there is in the co-kriging formulation. Therefore, the regression model involves estimating only one variogram, that of the \( z(\cdot) \) field, and can be used in situations where a stochastic model for one or more of the covariates is unreasonable. However, because the covariates are not stochastically modelled, in order to estimate \( z(x_o) \) from \( z(x_1), \ldots, z(x_n) \) using the regression model, we must know \( \hat{z}(x_o), \hat{z}(x_1), \ldots, \hat{z}(x_n) \). In the co-kriging model, we do not need to observe any of the \( \hat{f}_r \)'s at any particular place, as we can see in Equation (3).

IV.2 Comparison of Probabilistic Properties of the Two Models

Let us suppose that the model for \( z(\cdot) \) and \( \hat{z}(\cdot) \) given by Equations (1) and (2) is correct. Then, under what circumstances, and in what sense, can the regression model for \( z(\cdot) \) and \( \hat{z}(\cdot) \) given by Equation (1) in Chapter II also be correct? In this section, we will assume the joint distribution of the \( z(\cdot) \) and \( \hat{z}(\cdot) \) fields has been specified. We need to define a linear functional of \( \hat{z}(\cdot) \) "centered" around \( x \), which we will denote by \( L_x(\hat{z}) \). That is, if \( L_\theta(\hat{z}) \) is some linear functional, then \( L_x(\hat{z}) \) replaces \( \hat{z}(z) \) in \( L_\theta(\hat{z}) \) by \( \hat{z}(x+z) \) for all \( z \). For example, if
\[ \mathcal{L}_o(\xi) = \int \mathbf{y}(z)' \xi(z) \, dz, \]

then

\[ \mathcal{L}_x(\xi) = \int \mathbf{y}(z)' \xi(x+z) \, dz. \]

Or, if

\[ \mathcal{L}_o(\xi) = f'_{\xi}(x), \]

then

\[ \mathcal{L}_x(\xi) = f'_{\xi}(x). \]

The obvious way to try to get from the co-kriging model to the regression model is to condition on the \( \xi(\cdot) \) field. If the \( z(\cdot) \) and \( \xi(\cdot) \) fields are jointly stationary throughout \( \mathbb{R}^k \), then conditioned on the value of \( \xi(\cdot) \) everywhere in \( \mathbb{R}^k \), the \( z(\cdot) \) field will still be Gaussian,

\[
E[z(x)|\xi(\cdot)] = \mathcal{L}_x(\xi),
\]

and

\[
\text{Cov}[z(x), z(x')|\xi(\cdot)] = C^*(x-x'),
\]

where \( C^* \) is a stationary covariance function. We will not attempt to prove these somewhat obvious results. They make use of two important properties of Gaussian random variables: Equation (5) uses the fact that a conditional expectation will be linear in the conditioning random variables, and Equation (6) uses the fact that conditional covariances are independent of the conditioning random variables. If the joint distributions of the \( z(\cdot) \) and \( \xi(\cdot) \) fields are such that either of these properties does not hold, then we cannot obtain the regression model from the co-kriging model by conditioning on \( \xi(\cdot) \).

For the regression model to be correct conditioned on \( \xi(\cdot) \), we need the additional condition

\[
E[z(x)|\xi(\cdot)] = \mathcal{L}_x(\xi) = a_o + a'\xi(x),
\]

for some constants \( a_o \) and \( a \). Equivalently,
\text{Cov}\{\xi(x'), z(x) - \bar{a}'\xi(x)\} = 0, \text{ for all } x, x'. \quad (8)

If there is a vector \( \bar{a} \) such that (8) holds, then we can call \( z(\cdot) - \bar{a}_0 - \bar{a}'\xi(\cdot) \) the "disturbance" field. It is the same as the \( e(\cdot) \) field defined in (1) of Chapter II. Then Equation (8) says that the covariate field and disturbance field are uncorrelated.

Informally, all information about \( z(x) \) contained in the \( \xi(\cdot) \) field can be obtained from \( \bar{a}'\xi(\cdot) \). This condition is not wholly unreasonable, and may be approximately true in many problems. For example, if \( z(\cdot) \) measures temperature and the covariate is altitude, then the effect of all altitudes in a region on temperature at \( x \) may be mostly explained by the altitude at \( x \).

Let us consider the problem of estimating \( z_\star = z(x_\star) \) using the co-kriging model, based on \( z = (z(x_1) \ldots z(x_n))' \) and the \( \xi(\cdot) \) field. Define \( \xi_\star = \xi(x_\star) \), and \( F = (\xi(x_1) \ldots \xi(x_n)) \). If all the means and covariances are specified, we would estimate \( z_\star \) by its conditional expectation given \( z \) and the \( \xi(\cdot) \) field, which would be the optimal estimator. If condition (8) is satisfied, we have

\[
\hat{z}_\star = \mathbb{E}\{z_\star | z, \xi(\cdot)\} \\
= \mathbb{E}\{z_\star | z - F'\bar{a}, \xi(\cdot)\} \\
= \xi_\star'\bar{a} + \mathbb{E}\{z_\star - \xi_\star'\bar{a} | z - F'\bar{a}, \xi(\cdot)\} \\
= \xi_\star'\bar{a} + \mathbb{E}\{z_\star - \xi_\star'\bar{a} | z - F'\bar{a}\},
\]

since by (8), \( z_\star - \xi_\star'\bar{a} \) and \( z - F'\bar{a} \) are independent of \( \xi(\cdot) \). Thus, using the formula for conditional expectation of Gaussian random variables (Anderson (1958), p. 34),

\[
\hat{z}_\star = \xi_\star'\bar{a} + \frac{\mu_z - \mu_{\xi'}\bar{a}}{\text{Cov}\{z_\star - \xi_\star'\bar{a}, (z - F'\bar{a})'\}\text{Cov}\{(z - F'\bar{a}, (z - F'\bar{a})'\}^{-1} (z - F'\bar{a} - (\mu_z - \mu_{\xi'}\bar{a}) e). \quad (9)
\]
Thus, the conditional expectation depends on \( \hat{f}(\cdot) \) only through \( \hat{f}_* \) and the matrix \( F \). The only parameters and covariances needed to obtain \( \hat{z}_* \) are \( \alpha \), \( \alpha_0 \), which equals \( \mu_z - \mu_f' \alpha \), and \( \text{Cov}(z(x) - \hat{f}(x)' \alpha, z(x') - \hat{f}(x')' \alpha) \). These are exactly the parameters and covariances to be specified in the regression model. Thus, if these parameters and covariances are specified, \( \hat{f}_* \) and \( F \) are observed, and condition (8) holds, then the regression model will give the conditional expectation of \( z_* \) given \( z \) and the entire \( \hat{f}(\cdot) \) field. Therefore, if (8) holds and the data requirements of the regression model are met, it can be used instead of the co-kriging model and the same interpolators will be obtained.

Let us now consider what happens if the fields are not jointly Gaussian, but for simplicity, we still choose to use linear estimators. If the disturbance and covariate fields are uncorrelated (Equation (8)), there is still a correspondence between the co-kriging and regression models, even though we can no longer get from the first to the second model by conditioning on the \( \hat{f}(\cdot) \) field. Let us assume all means and covariances are specified, and only consider linear estimators which are unbiased under the co-kriging formulation. Then the estimator \( \hat{z}_* \), given in (9), is the BLUE of \( z_* \) based on \( z \) and the \( \hat{f}(\cdot) \) field. Thus, if the data requirements of the regression model are met (\( F \) and \( \hat{f}_* \) are observed), then we get the BLUE under the co-kriging model when we find the BLUE under the regression model. We no longer can interpret \( \alpha \) and \( \alpha_0 \) in terms of conditional expectations. However, they still have interpretations as regression coefficients. That is,

\[
\alpha \text{ minimizes } \text{Var}(z(x) - \beta' \hat{f}(x)) \text{ among all vectors } \beta , \quad (10)
\]

and

\[
Ez(x) = \mu_z = E(a_o + \alpha' \hat{f}(x)) = a_o + \alpha' \mu_f. \quad (11)
\]
So, in general, if there is a vector \( \alpha \) such that the disturbance and covariate fields are uncorrelated, and the \( f(\cdot) \) field is observed at all observation and interpolation sites, the regression model can be substituted for the co-kriging model. If Equation (8) is not satisfied for some \( \alpha \), then the regression model will not be compatible with the co-kriging model. If we could specify the linear functional \( \mathcal{L} \) which minimizes

\[
\text{Var}(z(x) - \mathcal{L}_X(\hat{f}))
\]

then we could define \( g(x) = \mathcal{L}_X(\hat{f}) \), and we would have

\[
\text{Cov}(z(x) - g(x), g(x')) = 0 \quad \text{for all } x, x'.
\]

Therefore, if we can specify \( \mathcal{L} \), we can reduce the problem to one in which (8) holds, and still use the regression model as long as \( f(\cdot) \) is observed densely enough so that \( g(\cdot) \) can be calculated. However, because of the simplicity of the regression model using \( f(\cdot) \), we may want to use it even when (8) does not hold. In the next section, under two related models, we investigate the properties of estimators using the regression model when it is not equivalent to co-kriging.

IV.3 Asymptotic Behavior of Estimators Based on the Regression Model When it is Not Valid

IV.3.1 Model with repeated observations in time

As noted in Chapter III, the asymptotic behavior of spatial models is complicated by its dependence not only on the number of observations, but on where the observations are. To a certain extent, we can avoid this problem by considering a process which is observed at a fixed set of sites repeatedly in time, with the process being independent from one time to another. Then we can obtain asymptotic results as the number of times becomes large. We will assume the co-kriging model is valid, and
see how estimators of the quantities in the regression model behave as the number of observations in time tends to infinity. In IV.3.2 and IV.3.3, we will see that, unless there is an $\alpha$ which satisfies (8), the values to which a generalized least squares estimator of the regression coefficients converges cannot strictly be interpreted as regression coefficients. However, in IV.3.4, we will see that the regression model will yield interpolators whose expected squared error can be properly evaluated using the regression model, even though the interpolators will not be optimal.

We now describe the basic model of this section. Suppose we observe $z(x,t)$ and $f(x,t)$ at times $t = t_1, ..., t_k$. The $z(\cdot, \cdot)$ field is observed at $x_1, ..., x_n$ at each of these times, and the $f(\cdot, \cdot)$ field is observed densely. We will assume

$$
\begin{align*}
Ez(x,t) &= \mu_z(t), \\
Ef(x,t) &= \mu_f(t), \\
\text{Cov}(z(x,t), z(x', t_j)) &= \delta_{ij} C_{zz}(x-x'), \\
\text{Cov}(z(x,t), f_\alpha(x', t_j)) &= \delta_{ij} C_{zf_\alpha}(x-x'), \text{ and} \\
\text{Cov}(f_\alpha(x,t), f_\beta(x', t_j)) &= \delta_{ij} C_{f_\alpha f_\beta}(x-x'),
\end{align*}
$$

(12)

where

$$
\delta_{ij} = \begin{cases} 
1, & i=j, \\
0, & i \neq j.
\end{cases}
$$

(13)

Under these assumptions, we will see what will happen if we instead base the interpolator on the model

$$
z(x,t) = \mu(t) + \alpha' f(x,t) + e(x,t).
$$

(14)

We consider $\mu_k = (\mu(t_1) \ldots \mu(t_k))'$ and $\alpha$ fixed but unknown, and
\[ E_e(x, t) = 0 \]
\[ \text{Cov}(e(x, t_i), e(x', t'_j)) = C_{ee}(x - x'). \] (15)

In IV.3.2, each component of \( \mu_k \) will be allowed to be arbitrary. In IV.3.3, all components of \( \mu_k \) will be assumed to be equal.

### IV.3.2 Mean changing in time

Let us define
\[ Z_\alpha = (z(x_1, t_\alpha) \ldots z(x_n, t_\alpha))^\prime, \ \alpha = 1, \ldots, k, \text{ and } \]
\[ F_\alpha = (f(x_1, t_\alpha) \ldots f(x_n, t_\alpha)), \ \alpha = 1, \ldots, k. \] (16)

Assuming \( \mu_k \) is an arbitrary, unknown vector, a reasonable estimator of \( (\hat{a}' \mu_k') \) is
\[
\begin{pmatrix}
\hat{a}_k \\
\hat{\mu}_k
\end{pmatrix} = \left[ \begin{pmatrix}
F_1 & \ldots & F_k \\
e' & 0 & \vdots \\
0 & e' & W
\end{pmatrix} \cdot \begin{pmatrix}
F_1 & e & 0 \\
W & F_k & e \\
e' & 0 & e
\end{pmatrix} \right]^{-1}
\]
\[
\times \begin{pmatrix}
F_1 & \ldots & F_k \\
e' & 0 & \vdots \\
0 & e' & W
\end{pmatrix} \cdot \begin{pmatrix}
Z_1 \\
\vdots \\
Z_k
\end{pmatrix}. \] (17)

If the regression model is valid and \( C_{ee}(\cdot) \) is specified, then letting
\[ W = (C_{ee}(x_i - x_j))^{-1} \] yields the BLUE of \( (\hat{a}' \mu_k'). \) If \( C_{ee}(\cdot) \) is not specified, we may define \( W \) using some a priori guess of the covariance structure, or we may just decide to use the simple choice of \( W = I. \)

Throughout the remainder of this chapter, we will use the convention that any index which appears more than once in a product of terms is to be summed over. Then,
Using the inversion formula for a partitioned matrix, we obtain

\[
\hat{a}_k = \left( F_\alpha W_\alpha' - (e'W_\alpha)^{-1} F_\alpha W_\alpha W_\alpha' \right)^{-1} \left( F_\alpha W Z_\alpha - (e'W_\alpha)^{-1} F_\alpha W e' \right) \left( F_\alpha W_\alpha' \right)^{-1} \left( e' W_\alpha \right),
\]

where we define

\[
W = W - (e'W_\alpha)^{-1} W e' W,
\]

or

\[
w_{\alpha\beta}^* = w_{\alpha\beta} - \left( \sum_r w_{rs} \right)^{-1} \left( \sum_r w_{sr} \right),
\]

where \( W = (w_{ij}) \), \( W^* = (w_{ij}^*) \). Thus, all the rows and columns of \( W^* \) sum to zero. If the co-kriging model, given in (12) and (13) is valid, then under sufficient conditions on the \( z(\cdot, \cdot) \) and \( f(\cdot, \cdot) \) processes, using \( W e = 0 \),

\[
\hat{a}_k \to [E(F_1 W_1^*)]^{-1} E(F_1 W_1 Z_1) \quad \text{in probability as } k \to \infty.
\]

Let us define

\[
a_\alpha = [E(F_1 W_1^*)]^{-1} E(F_1 W_1 Z_1).
\]

Again using \( W e = 0 \), we can show

\[
a_\alpha = \left( w_{\alpha\beta}^* C_{f_i f_j} (x_\alpha - x_\beta) \right)^{-1} \left( w_{\alpha\beta}^* c_{f_i z} (x_\alpha - x_\beta) \right).
\]

When \( W = I \), we obtain

\[
a_\alpha = \left[ \sum_{\alpha, \beta} \left[ C_{f_i f_j} (0) - C_{f_i f_j} (x_\alpha - x_\beta) \right] \right]^{-1} \left[ \sum_{\alpha, \beta} \left[ C_{f_i z} (0) - C_{f_i z} (x_\alpha - x_\beta) \right] \right]
\]

\[
= \left( \sum_{\alpha, \beta} \gamma_{f_i f_j} (x_\alpha - x_\beta) \right)^{-1} \left( \sum_{\alpha, \beta} \gamma_{f_i z} (x_\alpha - x_\beta) \right),
\]

where \( \gamma_{f_i f_j} (\cdot) \) and \( \gamma_{f_i z} (\cdot) \) are defined in (2), and the relationships
between the semi-variograms and covariances are given by

\[
\begin{align*}
\gamma_{f_1z}(r) &= C_{f_1z}(0) - \frac{1}{2}[C_{f_1z}(r) + C_{f_1z}(-r)], \\
\gamma_{f_1f_j}(r) &= C_{f_1f_j}(0) - \frac{1}{2}[C_{f_1f_j}(r) + C_{f_1f_j}(-r)].
\end{align*}
\]

We see from (22) that, in general, what \( \hat{\mu}_k \) converges to depends not only on where the observations are, but also on \( W^* \). If the regression formulation were correct, the expected value of the regression coefficients would not depend on either the locations of the observations or \( W^* \). In fact, if there is an \( a \) which satisfies

\[
\text{Cov}\{\tilde{f}(x', t), z(x, t) - a'\tilde{z}(x, t)\} = 0 \text{ for all } x, x', \text{ and } t, \quad (27)
\]

which is essentially a restatement of (8), then we can easily show that \( a_x = a \) for all \( W^* \) for which the inverse in (24) exists. That is, if \( a \) satisfies (27), then

\[
\hat{a}_k \rightarrow a \text{ in probability} \quad (28)
\]

if \( \{w_{\alpha\beta} C_{ij}(x_{\alpha} - x_{\beta})\} \) is non-singular. If there is no \( a \) which satisfies (27), then the problem we have is essentially that of missing regressors. That is, if \( \tilde{f}(x) \) does not provide all of the information about \( z(x) \) contained in the \( \tilde{f}(\cdot) \) field, then we are missing regressors \( (\tilde{f}(\cdot) \) at places other than \( x \) \) which are not orthogonal to the regressors in the model \( (\tilde{f}(x)) \), which we know will affect the expected value of the regression coefficients for \( \tilde{f}(x) \) (Chatterjee and Price (1977), pp. 215-217).

In many situations, we are interested in estimating \( \hat{\mu}_k \). In particular, it is often important to estimate \( \mu(t_m) - \mu(t_{m'}) \), which can be thought of as the difference in the mean level of \( z(\cdot) \) from \( t_m \) to \( t_{m'} \), after adjusting for the effect of the covariates. A similar problem is
discussed in V.3. From Equation (18), we can show that

$$\hat{\mu}_k = (e'W_e)^{-1}(e'W(Z_i - F_i'\hat{a}_k)),$$  \hspace{1cm} (29)

where $\hat{a}_k$ is defined in (19). Thus, as $k \to \infty$, we have, in probability,

$$\hat{\mu}(t_m) - \hat{\mu}(t_{m'}) \to (e'W)^{-1}e'W[(Z_m - F_m'\hat{a}_*) - (Z_{m'} - F_{m'}'\hat{a}_*)],$$  \hspace{1cm} (30)

where $\hat{a}_*$ is given by (23) or (24). Therefore, even as the number of observations in time becomes large, this estimator of $\mu(t_m) - \mu(t_{m'})$ will depend on $\hat{a}_*$, which in turn depends on $\hat{W}^*$ and the locations of the observations. For example, consider the problem of estimating changes in time in mean air pollution levels after correcting for the effect of weather on pollution. Suppose we use only the weather at $(x,t)$ to "predict" the pollution at $(x,t)$, which is what is done by the regression model given in (14). Then if non-local weather has an additional effect on the pollution at $(x,t)$ which is not explained by the local weather, our estimators of the change in mean levels over time will be biased, and the bias will depend on the locations of the sites as well as the weighting matrix, $W$, used to compute the estimators of the regression coefficients. Therefore, unless $\hat{a}_*$ satisfies (27), we cannot say that the effect of the covariates on the mean level of $z(\cdot, \cdot)$ has been eliminated in our estimator of the mean level.

IV.3.3 Mean fixed in time

In this section, we will use the regression model given in (14), except that we will assume that the $\mu(\cdot)$ term does not depend on time. That is, we will consider the model

$$z(x,t) = \mu + \xi'(x,t) + e(x,t),$$  \hspace{1cm} (31)

where $e(\cdot, \cdot)$ satisfies (15), when the correct model for the process is
given by the co-kriging model defined by (12) and (13). As before, we observe \( z(\cdot, \cdot) \) at \( x_1, \ldots, x_n \) for times \( t_1, \ldots, t_k \), and assume \( f(\cdot, \cdot) \) is known for times \( t_1, \ldots, t_k \). Analogous to (17) of the previous section, a reasonable estimator of \( (\hat{\xi}', \hat{\mu}) \) is

\[
\begin{pmatrix}
\hat{\xi}_k \\
\hat{\mu}_k
\end{pmatrix} = 
\begin{pmatrix}
F' \\
e'
\end{pmatrix}
\begin{pmatrix}
W' \\
e'
\end{pmatrix}^{-1} 
\begin{pmatrix}
F' \\
e'
\end{pmatrix}
\begin{pmatrix}
W' \\
e'
\end{pmatrix} \begin{pmatrix}
Z_1 \\
z_k
\end{pmatrix},
\]

(32)

where \( F_{\alpha} \) and \( Z_{\alpha} \) are defined in (16), and \( W \) is a weighting matrix.

Let us define

\[
A^B = \sum_{\alpha=1}^k A_{\alpha} W_{B\alpha},
\]

where we understand \( e_{\alpha} = e \) for all \( \alpha \) (so \( e^* e = ke^* We \)). Then we can show

\[
\begin{pmatrix}
\hat{\xi}_k \\
\hat{\mu}_k
\end{pmatrix} = 
\begin{pmatrix}
F'F' \\
e'*e'
\end{pmatrix}
\begin{pmatrix}
F*e \\
e*e
\end{pmatrix}^{-1} 
\begin{pmatrix}
F*Z \\
e'*Z
\end{pmatrix}
\]

\[
= \begin{pmatrix}
(F*F')^{-1}(F*Z) + \eta^{-1}(F*F')^{-1}(F*e) \\
(e'*F')(F*F')^{-1}(F*Z) - e'*Z \\
- \eta^{-1}(e'*F')(F*F')^{-1}(F*Z) + \eta^{-1}e'*Z
\end{pmatrix}
\]

(33)

where

\[
\eta = e'*e - (e'*F')(F*F')^{-1}(F*e),
\]

(34)

using the formula for the inverse of a partitioned matrix. Note that we can write

\[
\hat{\xi}_k = (F*F')^{-1}(F*Z) - (F*F')^{-1}(F*e) \hat{\mu}_k.
\]

(35)

Let us consider some special cases. For some constants \( \mu \) and \( \alpha \), define

\[
R_{\alpha} = Z_{\alpha} - \mu e - F_{\alpha}' \hat{\xi}_k.
\]

(36)
Then (33) can be written as

\[
\begin{pmatrix}
\hat{a}_k \\
\hat{\mu}_k
\end{pmatrix} = \begin{pmatrix}
a + (F * F')^{-1}(F * R) - (F * F')^{-1}(F * e)(\hat{\mu}_k - \mu) \\
\mu + \eta^{-1}[e' * R - (e' * F')(F * F')^{-1}(F * R)]
\end{pmatrix}.
\]  \hspace{1cm} (37)

If \( a \) satisfies (27) and \( E R_{\alpha} = 0 \) for all \( \alpha \), \( \) then \( E F_{\alpha} W R_{\alpha} = 0 \) for all \( \alpha \).

Therefore, under sufficient conditions on the \( f(\cdot, \cdot) \) and \( z(\cdot, \cdot) \) fields, we would have

\[
\begin{align*}
\frac{1}{k} \sum_{\alpha=1}^{k} F_{\alpha} W R_{\alpha} & \to 0 \text{ in probability as } k \to \infty, \text{ and} \\
\frac{1}{k} \sum_{\alpha=1}^{k} e' W R_{\alpha} & \to 0 \text{ in probability as } k \to \infty.
\end{align*}
\]  \hspace{1cm} (39)

As a consequence, based on (37), we would expect that

\[
\begin{pmatrix}
\hat{a}_k \\
\hat{\mu}_k
\end{pmatrix} \to \begin{pmatrix}
a \\
\mu
\end{pmatrix} \text{ in probability as } k \to \infty.
\]  \hspace{1cm} (40)

When the disturbance and covariate fields are uncorrelated and the expectation of \( z(x,t) - a' f(x,t) \) is independent of \( x \) and \( t \) (see (38)), the regression model in (31) with \( \mu(\cdot) \) fixed in time can be substituted for the co-kriging model in (12) and (13). Also, the natural estimators for the regression coefficients, given in Equation (37), converge to the "correct" value for any \( W \) which is positive definite.

As a second special case, let us assume

\[
\begin{align*}
E f(x,t) & = \mu_f \text{ for all } x \text{ and } t, \text{ and} \\
E z(x,t) & = \mu_z \text{ for all } x \text{ and } t.
\end{align*}
\]  \hspace{1cm} (41)

Under sufficient conditions on \( f(\cdot, \cdot) \) and \( z(\cdot, \cdot) \), based on (33), we would have
\[
\hat{\mu}_k + \mu_+ \quad \text{in probability, where}
\]
\[
\mu_+ = \left(1 - e' \text{We}_{\text{z}_f} \left(\text{E}(\text{F}_1 \text{WF}_1')^{-1} \text{E}_f\right)^{-1} \text{E}(\text{F}_1 \text{WF}_1') \text{E}(\text{F}_1 \text{WZ}_1)\right).
\]

We would also have, based on (35),
\[
\hat{\alpha}_k + \alpha_+ \quad \text{in probability, where}
\]
\[
a_+ = \left(\text{E}(\text{F}_1 \text{WF}_1')^{-1} \text{E}(\text{F}_1 \text{WZ}_1) - \text{E}(\text{F}_1 \text{WF}_1')^{-1} \mu_f (e' \text{We}) \mu_+ \right).
\]

We see, in general, that \(\mu_+\) and \(\alpha_+\) depend on \(W\) and the locations of the observations. Of course, if \(\mu_z(\cdot)\) and \(\mu_f(\cdot)\) do not depend on \(t\) (Equation (41)), and there is an \(a\) which satisfies (27), then \(\alpha_+ = a\) for any positive definite \(W\) and any set of locations. If
\[
W = \begin{pmatrix}
w_1 \\
\vdots \\
w_n
\end{pmatrix}
\]

is diagonal, and
\[
w_\cdot = \sum_{\alpha=1}^n w_\alpha, \quad \text{then}
\]
\[
\text{E}_1 \text{WF}'_1 = w_\cdot \left[\left(\text{C}_{f_1 f_j}^{f_1 f_j}(0) + \mu_f \mu_f'\right)\right], \quad \text{and}
\]
\[
\text{E}_1 \text{WZ}_1 = w_\cdot \left[\left(\text{C}_{f_1 z}(0) + \mu_f \mu_z\right)\right].
\]

By straightforward calculations, which are made easier if we start from the first line of (33), we can conclude
\[
\begin{pmatrix}
\alpha_+ \\
\mu_+
\end{pmatrix} = \begin{pmatrix}
\left(\text{C}_{f_1 f_j}^{f_1 f_j}(0) + \mu_f \mu_f'\right) & \mu_f' \\
\mu_f & 1
\end{pmatrix}^{-1} \begin{pmatrix}
\left(\text{C}_{f_1 z}(0) + \mu_f \mu_z\right) \\
\mu_z
\end{pmatrix}
\]
\[
= \begin{pmatrix}
\left(\text{C}_{f_1 f_j}^{f_1 f_j}(0)\right)^{-1} \left(\text{C}_{f_1 z}(0)\right) \\
\mu_z - \mu_f \alpha_s
\end{pmatrix}.
\]

Therefore, if \(W\) is diagonal and the means of the \(z(\cdot, \cdot)\) and \(f(\cdot, \cdot)\) fields are constant under the co-kriging model, then \(\alpha_+\) and \(\mu_+\) are the
"pointwise" regression coefficients. That is, $\alpha_+$ and $\mu_+$ are the values of $\alpha$ and $\mu$ which minimize

$$E\left(z(x,t) - \mu - \alpha \xi(x,t)\right)^2$$

(49)

This result does not depend on the locations of the sites or on the $w_i$'s. However, we should be wary of interpreting the pointwise regression coefficients as the "correct" regression coefficients. While the pointwise regression coefficients will minimize (49) among all $\mu$ and $\alpha$, the regression equation will only yield all of the information about $z(x,t)$ contained in the covariate field if the pointwise regression coefficients satisfy (27); that is, if $z(\cdot, t) - \alpha_+ \xi(\cdot, t)$ and $\xi(\cdot, t)$ are uncorrelated fields for any $t$, where $\alpha_+$ is defined as in (43).

IV.3.4 Interpolation based on the regression model

Even if the regression model is not correct, we may still want to use it to produce interpolators. Consider the regression model in (14), in which $\mu(\cdot)$ varies in time. Assume we are given a fixed vector $\alpha_k$, where we hope $z(\cdot, \cdot) - \alpha \xi(\cdot, \cdot)$ has "less" spatial variation than $z(\cdot, \cdot)$. For example, $\alpha_+ \xi$ may be an a priori estimate of $\alpha$, or if $k$, the number of observations in time, is very large, $\alpha_+ \xi$ may be defined as in (23), where we assume $k$ is sufficiently large so that the difference between $\hat{\alpha}_k$ (Equation (19)) and $\alpha_+ \xi$ is negligible. If the regression model as defined in (14) and (15) were correct and $\alpha_+ \xi$ equaled the true value of the $\alpha$, then the BLUE of $z(x_o, t_\alpha)$ would be of the form

$$\hat{z}(x_o, t_\alpha) = \alpha_\xi(x_o, t_\alpha) + \sum_{j=1}^{n} \lambda_j \left(z(x_j, t_\alpha) - \alpha_\xi(x_j, t_\alpha)\right), \text{ with } \sum_{j=1}^{n} \lambda_j = 1.$$  

(50)

Suppose we use an estimator of this form even when the regression model is not correct, or $\alpha_+ \xi$ does not equal the true value of $\alpha$. Define
\begin{align}
    z_*(x,t) &= z(x,t) - x_0' \xi(x,t), \quad \text{and} \\
    \frac{1}{2} E\left( z_*(x,t) - z_*(x',t) \right)^2 &= \gamma_*(x-x').
\end{align}

The semi-variogram \( \gamma_*(\cdot) \) will be stationary whenever \( z(\cdot,\cdot) \) and \( \xi(\cdot,\cdot) \) satisfy the co-kriging model as defined in (12) and (13). Then, if \( \hat{z}(x_0,t_\alpha) \) is of the form given in (50),
\begin{equation}
    E(\hat{z}(x_0,t_\alpha) - z(x_0,t_\alpha))^2 = 2 \sum \gamma_*(x_0 - x_j) - \sum \gamma_*(x_i - x_j).
\end{equation}

Thus, if we can obtain an unbiased estimator of \( \gamma_*(\cdot) \), we can obtain an unbiased estimator of the expected squared error of an interpolator of the form given in (50). The natural estimator of \( \gamma_*(r) \),
\begin{equation}
    \hat{\gamma}_*(r) = (2kn_r)^{-1} \sum_{x_i-x_i'=r} \sum_{j=1}^k \left( z_*(x_i,t_j) - z_*(x_i',t_j) \right)^2,
\end{equation}

where the outer sum is over all pairs of observation sites satisfying \( x_i-x_i'=r \), and \( n_r \) is the number of such pairs, is an unbiased estimator of \( \gamma_*(r) \). However, even if \( \gamma_*(\cdot) \) is specified and the \( \lambda_j \)'s are chosen to minimize the expected squared error in (53), the resulting estimator will not, in general, be the BLUE of \( z(x_0) \) under the co-kriging model.

An estimator of \( \gamma_*(\cdot) \) which is not a function of the \( z_*(\cdot,\cdot) \) field may not converge to the desired value. For example, suppose \( \xi(\cdot,\cdot) \) has one component, and we estimate \( \gamma_*(x,x') \) by
\begin{equation}
    \hat{\gamma}(x,x') = (2(k-1))^{-1} \sum_{j=1}^k \left( z(x,t_j) - z(x',t_j) - \hat{a}[f_1(x,t_j) - f_1(x',t_j)] \right)^2,
\end{equation}

where
\begin{equation}
    \hat{a} = \frac{\sum_{j=1}^k (z(x,t_j) - z(x',t_j))(f_1(x,t_j) - f_1(x',t_j)) / \sum_{j=1}^k \left( z(x,t_j) - z(x',t_j) \right)^2}{\sum_{j=1}^k \left( z(x,t_j) - z(x',t_j) \right)^2}
\end{equation}
is the least squares estimator of "a" based on \( z(x,t_j) - z(x',t_j) \), \( j=1,\ldots,n \) (see II.4.3). Then, as \( k \to \infty \), we have
\( \hat{\alpha}_k + a_{xx'} \) in probability, where

\[
a_{xx'} = [2(C_{zz}(0) - C_{zz}(x-x'))]^{-1}[2C_{f_iZ}(0) - C_{f_iZ}(x-x') - C_{f_iZ}(x'-x)].
\]

(56)

Thus, as \( k \to \infty \),

\[
\hat{\gamma}_*(x,x') = \mathbb{E}\left[ z(x,t_1) - z(x',t_1) - a_{xx'} (f_1(x,t_1) - f_1(x',t_1)) \right]^2.
\]

(57)

In general, we will not have \( a_{xx'} = a_* \), and thus \( \hat{\gamma}_*(x,x') \) will not necessarily converge to \( \gamma_*(x,x') \). However, when the regression model is "correct," and \( a_* \) equals the true value of "a," then we will have \( a_{xx'} = a_* = a \).
CHAPTER V: Kriging with Repeated Observations in Time and an Unknown, Varying Spatial Mean Function

V.1 Description of Model

In this chapter, we will consider models in which we have several realizations of a spatial field and the mean function is not a constant. In particular, we will consider spatio-temporal phenomena which are spatially and temporally non-stationary. More specifically, suppose we observe \( z(x,t) \) at \( (x_i, t_j) \) for \( i=1,\ldots,n \), and \( j=1,\ldots,k \). The basic model for this chapter is that

\[
z(x,t) = m(x) + \mu(t) + e(x,t).
\]  

(1)

We will assume that \( \mu(t_1), \ldots, \mu(t_k) \) can be considered an unknown and arbitrary deterministic time series. This assumption may cause us to miss much of the structure in the data if the \( \mu(t_i) \)'s exhibit strong temporal patterns. In particular, this assumption may be inappropriate when there are frequent observations in time. The \( m(\cdot) \) term represents spatial variations in the \( z(\cdot, \cdot) \) field which do not change over the time range being considered. The \( e(\cdot, \cdot) \) term represents spatio-temporal variations in the \( z(\cdot, \cdot) \) field not accounted for by \( m(\cdot) \) and \( \mu(\cdot) \). We will assume

\[
E[e(x,t) = 0,
\]

and

\[
\frac{1}{2} E[e(x,t) - e(x',t)]^2 = \gamma(x-x'),
\]

\[
E[(e(x,t_i) - e(x',t_i))[e(x'',t_j) - e(x'',t_j)]] = 0, \quad i \neq j.
\]

(3)

This last condition says that differences in the \( e(\cdot, \cdot) \) field at one time at which observations are taken are uncorrelated with differences from another time. This condition is somewhat weaker than assuming
Ee(x,t_i)e(x',t_j) = 0 for i \neq j, since it does not assume Var e(x,t) is finite. However, it does say something about the lack of structure in the e(\cdot,\cdot) field from one time to another. Thus, like the assumption on the u(t_i)'s, it may be inappropriate if the t_i's are too close together.

If we are interested in doing spatial interpolation, we will need to assume that the m(\cdot) field has some spatial structure. For many phenomena, it may be reasonable to assume that the places which get consistently high readings tend to be grouped together. That is, m(\cdot) may exhibit some sort of spatial continuity. Since spatial continuity in the e(\cdot,\cdot) field is being modelled by assuming e(\cdot,t) has a stationary spatial variogram for a given t, it seems reasonable to model m(\cdot) in a similar manner. Therefore, we will assume

\[
\begin{align*}
\text{Em}(x) &= 0, \quad \text{and} \\
\frac{1}{2}E\{(m(x) - m(x'))^2 & = \eta(x - x'). \} \\
\end{align*}
\]

Throughout this chapter, we will assume the m(\cdot) and e(\cdot,\cdot) fields are independent. By using this approach, we will be able to obtain estimates of the variance of the interpolation error in linear unbiased interpolators.

Using this model, several problems will be investigated. In V.2, assuming the semi-variograms \( \gamma(\cdot) \) and \( \eta(\cdot) \) have been specified, linear unbiased interpolators for \( z(x_0,t_\beta) \) will be considered. In V.3, assuming \( \gamma(\cdot) \) is known, estimators for time trends in \( z(\cdot,\cdot) \) will be discussed. Finally, in V.4, there is a short discussion on the estimation of the semi-variograms, \( \gamma(\cdot) \) and \( \eta(\cdot) \).
V.2 Interpolation

V.2.1 Mean changing in time

Using the model defined in (1)-(4), we consider estimators of $z(x_0, t_\beta)$ of the form

$$\hat{z}(x_0, t_\beta) = \sum_{i,j} \lambda_{ij} z(x_i, t_j),$$

(5)

where $t_\beta$ is one of the times at which we have taken observations. Such an estimator will be unbiased if and only if

$$\sum_i \lambda_{i\beta} = 1, \text{ and}$$

$$\sum_i \lambda_{ij} = 0, \quad j \neq \beta.$$ 

(6)

For any set of weights satisfying (6), we can show, as in I.1.2,

$$\text{Var}\left( \sum_{i,j} \lambda_{ij} z(x_i, t_j) - z(x_0, t_\beta) \right) = 2 \sum_{ij} \lambda_{ij} \eta(x_i - x_0) + 2 \sum_i \lambda_{i\beta} \gamma(x_i - x_0)$$

$$- \sum_{ijrs} \lambda_{ij} \lambda_{rs} \eta(x_i - x_r) - \sum_{irj} \lambda_{ij} \lambda_{rj} \gamma(x_i - x_r).$$

(7)

Thus, conditions (2) and (3) sufficiently specify the second order structure of the process to evaluate the variance of the error of linear unbiased interpolators.

By symmetry considerations, it is clear that the weights which minimize the variance in (7) must satisfy

$$\lambda_{ij} = \lambda_{ij}', \text{ if } j, j' \neq \beta.$$ 

(8)

Let us consider interpolators which satisfy (8), and define

$$\lambda_{ij} = \omega_i, \quad j \neq \beta, \quad i=1, \ldots, n, \text{ and}$$

$$\lambda_{i\beta} - \omega_i = \nu_i, \quad i=1, \ldots, n.$$ 

(9)

For estimators satisfying (6) and (8), we then have
\[ \text{var}(\hat{z}(x_0, t_\beta) - z(x_0, t_\beta)) = 2 \sum_{i} (kw_i + v_i)\eta(x_i - x_0) + 2 \sum_{i} (w_i + v_i)\gamma(x_i - x_0) \]
\[ - \sum_{i} (k^2 w_i w_r + 2k w_i v_r + v_i v_r)\eta(x_i - x_r) - \sum_{i} (k w_i w_r - 2w_i v_r + v_i v_r)\gamma(x_i - x_r). \]

The unbiasedness conditions can be written as
\[
\begin{align*}
\sum_{i} w_i &= 0, \text{ and } \\
\sum_{i} v_i &= 1.
\end{align*}
\]

To obtain the \( v_i \)'s and \( w_i \)'s which minimize (10) subject to (11), we introduce Lagrange multipliers \( v_1 \) and \( v_2 \) and obtain that the minimizing weights satisfy (11) and
\[
\begin{aligned}
k\eta(x_\ell - x_0) + (x_\ell - x_0) - \sum_{r} (k^2 w_r + kv_r)\eta(x_\ell - x_r) \\
- \sum_{r} (kw_r + v_r)\gamma(x_\ell - x_r) - \frac{1}{2} v_1 &= 0, \; \ell = 1, \ldots, n,
\end{aligned}
\]
and
\[
\begin{aligned}
\eta(x_\ell - x_0) + \gamma(x_\ell - x_0) - \sum_{r} (kw_r + v_r)\eta(x_\ell - x_r) \\
- \sum_{r} (w_r + v_r)\gamma(x_\ell - x_r) - \frac{1}{2} v_2 &= 0, \; \ell = 1, \ldots, n.
\end{aligned}
\]

Define \( N = (\eta(x_i - x_j)) \), \( N_0 = (\eta(x_i - x_0)) \), \( \Gamma = (\gamma(x_i - x_j)) \), \( \Gamma_0 = (\gamma(x_i - x_0)) \), \( y = (v_1 \ldots v_n)' \), and \( w = (w_1, \ldots, w_n)' \). Then (11) and (12) can be written as
\[
\begin{pmatrix}
k^2 N + k \Gamma & k N + \Gamma & e & 0 \\
k N + \Gamma & N + \Gamma & 0 & e \\
e' & 0' & 0 & 0 \\
0' & e' & 0 & 0
\end{pmatrix}
\begin{pmatrix}
w \\
y \\
\frac{1}{2} v_1 \\
\frac{1}{2} v_2
\end{pmatrix}
= \begin{pmatrix}
KN_0 + \Gamma_0 \\
N_0 + \Gamma_0 \\
0 \\
1
\end{pmatrix}
\]

Define \( z(t_j) = (z(x_1, t_j) \ldots z(x_n, t_j))' \), and \( \bar{x} = \frac{1}{k} \sum_{j=1}^{k} z(t_j) \). Then the BLUE of \( z(x_0, t_\beta) \) is
\[ \hat{z}(x_0, t_\beta) = k_{y'} \hat{z} + y' \hat{z}(t_\beta), \] (14)

where \( y \) and \( y' \) are defined by (13). Straightforward calculations yield

\[ v = \Gamma^{-1} \Gamma_0 + \Gamma^{-1} \varepsilon \left\{ \frac{1 - \varepsilon'(kN + \Gamma)^{-1}}{\varepsilon' \Gamma^{-1} \varepsilon} \right\}, \] and

\[ k_{y'} + y = (kN + \Gamma)^{-1} (kN_0 + \Gamma_0) + (kN + \Gamma)^{-1} \varepsilon \left\{ \frac{1 - \varepsilon'(kN + \Gamma)^{-1}}{\varepsilon' (kN + \Gamma)^{-1} \varepsilon} \right\}. \] (15)

If \( m(x) \equiv 0 \), then \( N = 0 \) and \( N_0 = 0 \), which imply \( y = 0 \). In this case, \( \hat{z}(x_0, t_\beta) \) is just the ordinary kriging interpolator based on \( z(t_\beta) \), so that observations from other times do not improve the BLUE if there are no persistent spatial variations. We should note that this and other interpolators in this chapter can also be obtained from the principles of co-kriging (see Chapter IV) by considering observations from other times as covariates.

Also, as the number of observation times \( k \rightarrow \infty \), we have

\[ y + k_{y'} = N^{-1} N_0 + N^{-1} \varepsilon \left\{ \frac{1 - \varepsilon' N^{-1} N_0}{\varepsilon' N^{-1} \varepsilon} \right\}. \] (16)

Thus, for \( k \) large, we have roughly

\[ \hat{z}(x_0, t_\beta) = \left\{ \Gamma^{-1} \Gamma_0 + \Gamma^{-1} \varepsilon \left\{ \frac{1 - \varepsilon' \Gamma^{-1} \Gamma_0}{\varepsilon' \Gamma^{-1} \varepsilon} \right\} \right\}' (\hat{z}(t_\beta) - \hat{z}) \]

\[ + \left\{ N^{-1} N_0 + N^{-1} \varepsilon \left\{ \frac{1 - \varepsilon' N^{-1} N_0}{\varepsilon' N^{-1} \varepsilon} \right\} \right\}' \hat{z}. \] (17)

This result is easily interpretable. The first term essentially interpolates the \( e(\cdot, \cdot) \) field at time \( t_\beta \), and the second term interpolates the \( m(\cdot) \) field.
If only observations from time $t_B$ are used to estimate $z(x_0, t_B)$, then the best estimator based on observations from that time is

$$\left( (N+\Gamma)^{-1} (N_0 + \Gamma 0') + (N+\Gamma)^{-1} \frac{1 - \varepsilon(N+\Gamma)^{-1}(N_0 + \Gamma 0')}{\varepsilon'(N+\Gamma)^{-1} \varepsilon} \right)' \hat{z}(t_B). \quad (18)$$

By restricting ourselves to observations from time $t_B$, we lose information about the part of the spatial structure of the data which does not vary in time. If a substantial amount of the spatial structure is independent of time, then the estimator given in (14) may be significantly better than the one in (18).

In many spatio-temporal phenomena, there may be a component of the process which, while not unchanging in time, does vary much more slowly than other components. For example, meteorological phenomena show day to day fluctuations and also seasonal fluctuations. However, over a small enough number of days, the seasonal effect may be inconsequential. Thus, the seasonal effect on the spatial structure of meteorology, while having a profound effect over long periods of time, may be negligible over shorter periods. To be concrete, let us consider two sites, $x$ and $x'$, with $x$ nearer to the ocean than $x'$. As a result, $x$ tends to be cooler than $x'$ in the summer, but warmer in the winter. Therefore, the model defined by (1)-(3) would be inappropriate over the entire year, since the seasonal component of temperature variation could not be put into any of the terms in (1). The $m(\cdot)$ term cannot be used, since it cannot change in time. The $\mu(\cdot)$ term cannot be used, since it cannot depend on the place, and the seasonal variation is larger at $x'$ than at $x$. Finally, the $e(\cdot, \cdot)$ term cannot be used because of the uncorrelatedness condition between times in (3). However, if our analysis focused on a
single month, say, than there would be a consistent difference in temperature between \( x \) and \( x' \). Thus, over a single month, our model may be approximately correct, and can use the consistent difference in temperature between seashore and inland sites to produce better temperature interpolators than a method which only looked at a single time.

We see that the way this model treats temporal variations is rather simplistic. The \( m(\cdot) \) term represents that part of the process which does not vary in time, or the zero frequency variations, we could say. The \( e(\cdot,\cdot) \) term represents the part of the process which shows no temporal structure from one observation time to the next, or very high frequency variations. The model leaves no room for variations at other frequencies. Allowing temporal correlations in the \( e(\cdot,\cdot) \) field would enable us to include variations at other frequencies. Such models will not be discussed in this work.

V.2.2 Mean fixed in time

In this section, we consider the effect of making \( \mu(t) \equiv \mu \) on the BLUE of \( z(x_0, t_\beta) \). That is, consider the model

\[
z(x, t) = \mu + m(x) + e(x, t),
\]

where \( \mu \) no longer depends on \( t \), but \( e(\cdot, \cdot) \) and \( m(\cdot) \) still obey (2)-(4). Again, we wish to estimate \( z(x_0, t_\beta) \) based on \( z(x_i, t_j) \) for \( i=1, \ldots, n \), and \( j=1, \ldots, k \), using a linear estimator as given in (5).

Now, to obtain unbiasedness, we only need

\[
\sum_{ij} \lambda_{ij} = 1.
\]

However, equations (2) and (3) do not specify the correlation structure of the \( e(\cdot, \cdot) \) field sufficiently to determine the variance of all
unbiased estimators. Therefore, we make the additional assumption

$$\frac{1}{2} E(e(x,t_i^j) - e(x',t_j^i))^2 = \gamma^*, \ i \neq j. \quad (21)$$

We should note that $\gamma^*$ cannot take on any value. If it is too small, variances of some linear combinations of the $e(\cdot,\cdot)$ field will be negative. In Appendix F, we show that $\gamma^*$ must be greater than or equal to the sill value of $\gamma(\cdot)$ which we denote by $\gamma(\infty)$. By allowing $\gamma^*$ to be greater than $\gamma(\infty)$, we can include a "white noise" time component in our model. That is, we can write

$$\mu + e(x,t_j^j) = \mu(t_j^j) + e^*(x,t_j^j),$$

where $E(\mu(t_j^j)) = \mu$,

$$\frac{1}{2} E(\mu(t_i^j) - \mu(t_j^j))^2 = \gamma^* - \gamma(\infty), \ i \neq j, \text{ and}$$

$$\frac{1}{2} E(e^*(x,t_i^j) - e^*(x',t_j^i))^2 = \begin{cases} \gamma(x-x'), & i=j, \\ \gamma(\infty), & i \neq j. \end{cases}$$

If $\gamma^* < \infty$, then (21) and (2) imply both (3) and the stronger condition $E(e(x,t_i^i)e(x',t_j^j)) = 0$, for $i \neq j$. If $\gamma^* = \infty$, then condition (3) is not redundant. When $\gamma^* = \infty$, we also have that (20) is not sufficient to obtain an estimator whose error has finite variance. That is, unless Equation (6) holds, $\text{Var}\left[\sum_{i,j} \lambda_{ij} z(x_i^i,t_j^j) - z(x_0^i,t_B^j)\right]$ will be infinite. When (6) is satisfied, Equation (7) gives the correct value of the variance of the interpolation error. Therefore, the $\lambda_{ij}$'s which minimize the variance will be the same as before, so the BLUE of $z(x_0^i,t_B^j)$ is again given by (14), with $\gamma$ and $\gamma^*$ defined by (15), when $\gamma^* = \infty$ and $\mu(t) \equiv \mu$. 


If $\gamma^* < \infty$, then, for $\lambda_{ij}$'s satisfying (20), we can show

$$\text{Var}(\sum_{ij} \lambda_{ij} z(x_i, t_j) - z(x_0, t_\beta)) = 2 \sum_{ij} \lambda_{ij} \gamma(x_i - x_0) + 2 \sum_1 \lambda_{i\beta} \gamma(x_i - x_0)$$

$$- \sum_{i j r s} \lambda_{ij} \lambda_{rs} \eta(x_i - x_r) - \sum_{i j r} \lambda_{ij} \lambda_{rj} \gamma(x_i - x_r) + \gamma^*(1 - 2 \sum_1 \lambda_{i\beta} + \sum_{i j r} \lambda_{ij} \lambda_{rj}).$$ (22)

As in the previous section, by symmetry considerations, the BLUE must satisfy (8). Define $w_i$ and $v_i$ as in (9). We can write the unbiasedness constraint as

$$\sum_i (kw_i + v_i) = 1. \tag{23}$$

Repeating the argument in the preceding section, we can show

$$\begin{pmatrix}
    k^2 N + k \Gamma - \gamma^* e' e' \\
    kN + \Gamma - \gamma^* e e' \\
    kN + \Gamma - \gamma^* e e' \\
    k e'
\end{pmatrix}
\begin{pmatrix}
    y \\
    y \\
    y \\
    k
\end{pmatrix}
= \begin{pmatrix}
    kN_0 + \Gamma_0 - \gamma^* e \\
    N_0 + \Gamma_0 - \gamma^* e \\
    0 \\
    \frac{k}{e'}
\end{pmatrix} \begin{pmatrix}
    w_i \\
    v_i \\
    1
\end{pmatrix} \tag{24}
$$

where $\nu$ is a Lagrange multiplier. By a straightforward but tedious argument, we can obtain

$$y = \Gamma^{-1} \Gamma_0 + \gamma^* \left(1 - \frac{e' \Gamma^{-1} \Gamma_0}{\gamma^* e' \Gamma^{-1} e - 1}\right) \Gamma^{-1} e, \text{ and}$$

$$kw + y = (kN + \Gamma)^{-1} (kN_0 + \Gamma_0) + (kN + \Gamma)^{-1} e \left[1 - \frac{e' (kN + \Gamma)^{-1} (kN_0 + \Gamma_0)}{e' (kN + \Gamma)^{-1} e}\right];$$

and the BLUE is

$$\hat{z}(x_0, t_\beta) = (kw + y)' \tilde{z} + y' (z(t_\beta) - \tilde{z}). \tag{26}$$

Roughly speaking, for $k$ large, we have
\[
\hat{z}(x_0, t_B) = \left[ \tau^{-1} \Gamma_0 + \tau^{-1} \gamma^* (1 - e^{-\gamma^{-1} \Gamma_0}) \right] \left( \tilde{z}(t_B) - \bar{z} \right) + N^{-1} N_0 + N^{-1} e \left( \frac{1 - e^{-\gamma^{-1} N N_0}}{e^{-\gamma^{-1} e}} \right) \tilde{z}.
\]

(27)

Since \( \gamma^* \) is finite, covariances of the \( e(\cdot, \cdot) \) field are finite. Thus, we can write

\[
\sigma(x-x') = E e(x,t) e(x',t) = \gamma^* - \gamma(x-x').
\]

(28)

If we define \( \Sigma = (\sigma(x_i-x_j)) \), and \( \Sigma_0 = (\sigma(x_i-x_0)) \), then \( \gamma = \Sigma^{-1} \Sigma_0 \). Let us compare (17) and (28), the asymptotic results in the cases \( \mu \) varying in time and \( \mu \) fixed, respectively. When \( \mu \) is fixed, for \( k \) large, \( z(t_B) - \bar{z} \) is almost equal to \( \tilde{e}(t_B) = (e(x_1, t_B) \ldots e(x_n, t_B))' \). Thus, we can essentially interpolate the \( e(\cdot, \cdot) \) field, which we know has mean zero. Interpolation of a field with known mean is called "simple kriging" (see Journel and Huijbregts). Since \( \gamma = \Sigma^{-1} \Sigma_0 \), the first term in (28) does this interpolation. When \( \mu \) varies in time, \( z(t_B) - \bar{z} \) will be approximately equal to \( \tilde{e}(t_B) + c \epsilon \), where \( c \) is unknown. Thus, we must interpolate a random field with an unknown constant mean, which the first term in (17) does. In each case, the interpolation of the \( m(\cdot) \) field is the same.

One last point worth mentioning is in what sense the interpolators given here are unbiased. For example, at any one place \( x_0 \), we may consistently underpredict \( z(x_0, t) \) over time if \( m(x_0) \) is unusually large. Thus, these interpolators are only unbiased at a single point because \( m(x) \) has been modelled stochastically with a constant mean. In general, these estimators are not unbiased if we condition on the value of the \( m(\cdot) \) field.
V.3 Estimation of Time Trends

In pollution control, one critical statistical problem is to estimate time trends in the average level of a pollutant over some region. For example, we may be interested in estimating how much the total acid content in rainfall in the Northeast is changing from one year to the next, and in knowing the approximate variance of this estimate. In doing this analysis, we would like to make use of the fact that, year after year, certain places get consistently more acid deposition than others.

Let us assume that Equations (1)-(4) give an appropriate model for the data. Before deciding how to estimate the time change, we must decide exactly what it is we want to estimate. The form of Equation (1) suggests that we might be interested in $\mu(t_i) - \mu(t_j)$, where $t_i$ and $t_j$ are the two times of interest. However, $\mu(t_i) - \mu(t_j)$ is not an observable quantity, but is a creation of our model. Further, the variance of a linear unbiased estimator of $\mu(t_i) - \mu(t_j)$ will not be determined by the semi-variograms $\gamma(\cdot)$ and $\eta(\cdot)$, but will also depend on $\gamma^*$. As a result, if $\gamma^*$ is infinite, the variance of a linear unbiased estimator will also be infinite. More importantly, when $\mu(t_i)$ is modelled as an unknown, deterministic time series, $\gamma^*$ cannot be estimated, since there is no way to separate the "random" fluctuations in the time trend from the "deterministic" fluctuations. Thus, we see that there are considerable problems in interpreting what $\mu(t_i) - \mu(t_j)$ really is, and in estimating the variance of linear estimators of it.

Instead, denote by $R$ the region of interest, and let $A(R)$ be the area of the region. Then, if we try to estimate the average change in $z(\cdot, \cdot)$ over $R$,

$$A(R)^{-1} \int_R z(x, t_i) dx - A(R)^{-1} \int_R z(x, t_j) dx, \quad (30)$$
the above problems are avoided. This quantity is real and observable, so its true value does not depend on the model. Also, the variance of any linear unbiased estimator of (30) will be a function of $\gamma(\cdot)$ and $\eta(\cdot)$, and will be finite even if $\gamma^*$ is infinite. By estimating (30), no attempt is made to determine whether a time change is due to "chance" fluctuations in the particular region chosen, or due to more spatially global trends over time. Thus, we shall keep in mind that our estimator is a measure of what actually happened in $R$, and does not necessarily tell us anything about what happened outside of $R$ or what could have happened in $R$ if "random" factors in $R$ had been different.

Let us now consider the problem of how to estimate the quantity in (30). Consider the case of a random field observed at the same set of sites at only two times. The obvious way to deal with consistent trends in the field over time is to take differences at each site and to estimate the average difference in the field over $R$ based on these pairwise differences. That is, suppose we observe

$$z(x_i, t_j), \ i=1, \ldots, n, \text{ and } j=1, 2.$$  

Define $\Delta z(x) = z(x, t_1) - z(x, t_2)$. Then estimate

$$A(R)^{-1} \int_R (z(x, t_1) - z(x, t_2)) \, dx = A(R)^{-1} \int_R \Delta z(x) \, dx$$

based on the $n$ $\Delta z(x_i)$'s. By looking at differences, the problem of estimating $m(\cdot)$ or its semi-variogram, $\eta(\cdot)$, is avoided. In fact, it is not necessary to make any stochastic assumptions about $m(\cdot)$ for this procedure to make sense.

Let us compare this approach to the one in which $A(R)^{-1} \int_R z(x, t_1) \, dx$ is separately estimated for $t_1$ and $t_2$, and then (30) is estimated using the difference of these estimates, which are assumed to be
uncorrelated. While these two procedures will give approximately the same estimators, the estimated variance of these estimators may be radically different. If a substantial part of the variation in \( z(\cdot, \cdot) \) persists from \( t_1 \) to \( t_2 \), then this second procedure will produce much higher estimates of the variance of our estimator of (30). Thus, by using the second procedure, we may conclude that our estimate of (30) is not statistically significantly different from zero, whereas using the first procedure will lead to the conclusion that the change over time is statistically significant. The problem is analogous to using a paired t-test rather than a two-sample t-test. That is, if the observations within a pair (from the same site) tend to be similar, then we can lower the variance of our estimator of the average pairwise difference (Equation (30)) by using the paired t-test (first procedure) instead of a two-sample t-test (second procedure). If in fact \( m(x) = 0 \), then what is lost by using the first procedure is half of the information about the semi-variogram \( \gamma(\cdot) \), which is analogous to the lost degrees of freedom in estimating \( \sigma^2 \) when the paired t-test is used. If the number of observations is not too small, then losing information about \( \gamma(\cdot) \) is not of great consequence, so we should greatly prefer to use the differencing procedure if there is any reason to believe spatial variations may be persistent in time.

Now let us consider the case where we observe the \( z(\cdot, \cdot) \) field at more than two times, but are still interested in estimating the average difference in the field over \( R \) between two times. That is, using the model defined in (1)-(4), suppose we observe \( z(x_i, t_j) \), \( i = 1, \ldots, n \), and \( j = 1, \ldots, k \), and we wish to estimate \( A(R)^{-1} \int_R (z(x, t_0) - z(x, t_a)) \, dx \).
First, using (14) and (15), given a place \( x_0 \), we obtain that the optimal estimator of \( z(x_0, t_\beta) - z(x_0, t_\alpha) \) is

\[
\hat{z}(x_0, t_\beta) - \hat{z}(x_0, t_\alpha) = \left( \Gamma^{-1} \Gamma_0 + \varepsilon \Gamma^{-1} \left[ \frac{1 - \varepsilon \Gamma^{-1} \Gamma_0}{\varepsilon \Gamma^{-1} \varepsilon} \right]^{'} \right) (z(t_\beta) - z(t_\alpha)), \tag{31}
\]

where \( \Gamma, \Gamma_0 \), and \( z(t_\alpha) \) are defined in V.2.1. As noted by Journel and Huijbregts (p. 321), the optimal estimator of a regional average is a linear functional of the optimal point estimators. Thus, the best estimator of (30) is of the form

\[
\nu_R^{'} (z(t_\beta) - z(t_\alpha)),
\]

where \( \nu_R \) is a vector depending on \( R \) and \( (\cdot) \). This estimator is just a linear combination of the differences of the observations between \( t_\alpha \) and \( t_\beta \), and does not depend on observations from other times. Therefore, the only gain of having observations at times other than \( t_\alpha \) and \( t_\beta \) is that they allow us to estimate \( \gamma(\cdot) \) better.

V.4 Estimation of the Variograms

In the preceding sections, we have assumed that the semi-varyograms, \( \gamma(\cdot) \) and \( \eta(\cdot) \), have been specified. In practice, these functions must be estimated based on the data. In this section, we briefly consider unbiased non-parametric estimators of \( \gamma(\cdot) \) and \( \eta(\cdot) \). We will use the model defined by (1)-(4), although the properties of the estimators given will not depend on how \( \mu(t) \) is modelled.

Condition (3), which says that differences from one time are uncorrelated with those from another, should be exploited when estimating \( \gamma(\cdot) \). Define

\[
\bar{z}(x) = \frac{1}{k} \sum_{j=1}^{k} z(x, t_j).
\]
As noted in II.4, if \( x \) and \( x' \) are two sites at which \( z(\cdot, \cdot) \) is observed, then we can estimate \( \gamma(x, x') \) by

\[
\hat{\gamma}(x, x') = \frac{1}{2(k-1)} \sum_{j=1}^{k} \left( \left( z(x, t_j) - \bar{z}(x) \right) - \left( z(x', t_j) - \bar{z}(x') \right) \right)^2,
\]

which yields an unbiased estimator of \( \gamma(x, x') \). If \( e(\cdot, \cdot) \) is Gaussian, then

\[
\hat{\gamma}(x, x') \sim \frac{\gamma(x, x')}{k-1} \chi^2_{k-1}.
\]

These results do not depend on how either \( m(x) \) or \( \mu(t) \) are modelled.

If \( \gamma(\cdot) \) is isotropic, we can estimate \( \gamma(d) \) using (17) in Chapter II.

If we are only interested in estimating differences over time in the \( z(\cdot, \cdot) \) field, then \( \gamma(\cdot) \) is all we need to estimate. However, if we are interested in doing interpolation at a specific time, then we also need to estimate \( \eta(\cdot) \), the semi-variogram of the \( m(\cdot) \) field. An unbiased estimator of \( \eta(x, x') \) is given by

\[
\hat{\eta}(x, x') = \frac{k}{2(k-1)} \left( \bar{z}(x) - \bar{z}(x') \right)^2 - \frac{1}{2(k-1)} \sum_{j=1}^{k} \left( z(x, t_j) - z(x', t_j) \right)^2.
\]

If \( \eta(\cdot) \) is isotropic, we could estimate \( \eta(d) \) by

\[
\hat{\eta}(d) = \frac{1}{m} \sum_{|x_i - x_j| = d} \hat{\eta}(x_i, x_j),
\]

where the sum is over all pairs of observation sites which satisfy \(|x - x'| = d\), and \( m \) is the number of such pairs. We should note that \( \hat{\eta}(d) \) may be negative, although if we demand unbiasedness, this problem seems unavoidable. A similar problem occurs when estimating the variance of a treatment effect in a one-way random effects ANOVA model (see Scheffe (1959), for example). However, if \( m \) is large and \( \eta(d) \) is of the same magnitude as \( \gamma(d) \), then the probability of \( \hat{\eta}(d) \) being negative will be
small. In general, the number of observations in any one time, \( n \), must be large if \( \eta(\cdot) \) is to be estimated reasonably well. Large numbers of observations in time do not help much in estimating \( \eta(\cdot) \); after about ten replications in time, further replications will provide only negligible additional information about \( \eta(\cdot) \). On the other hand, estimates of \( \gamma(\cdot) \) do continue to improve as the number of replications in time increase. Thus, if we are only interested in estimating differences over time, we may be able to estimate the variances of these estimators, which depend only on \( \gamma(\cdot) \), with a relatively small number of sites, if there are sufficient replications in time. However, if we wish to interpolate the \( z(\cdot,\cdot) \) field at a particular time, we need a large number of sites if the variances of the interpolation errors are to be estimated well.
CHAPTER VI: The Effect of Sequential Siting on Estimation

VI.1 Description of the Problem

In some studies of spatial or spatio-temporal processes, not all of the observation sites are picked before any observations are made. An initial set of observations may be taken, and then future sites are chosen depending on the outcome of the previous observations, a technique we will call sequential siting. One likely instance in which such a sampling scheme can occur is when it is important to estimate a random field well in regions where it takes on relatively high levels. In such a case, we might take some initial observations, and then concentrate future observations in regions where the previous observations were relatively large. In general, the distribution of any random quantity which depends on these latter observations will be affected by the fact that the locations of the observations are functions of other random observations. However, if we observe a Gaussian process with known covariance function, then the distribution of a best linear unbiased interpolator will be unaffected by the siting scheme, as long as all observation sites are functions of the contrasts of previous observations.

VI.2 A Theorem on the Effect of Sequential Siting

Consider a random field \( z(\cdot) \) defined by

\[
z(x) = \mathbf{a}' \mathbf{f}(x) + \mathbf{e}(x), \tag{1}
\]

where \( \mathbf{f}(\cdot) \) is a known vector-valued function with \( p \) components, \( \mathbf{a} \) is a vector of unknown regression coefficients, and \( \mathbf{e}(\cdot) \) is a disturbance term whose distribution is known. Suppose there are \( m \) initial observations taken, and each of \( n \) subsequent sites is chosen
one at a time, so that there is a total of \( n+m \) sites. Each of the \( n \)
sequentially chosen sites is allowed to depend on all previous observa-
tions, including the \( m \) initial observations, as long as the site is a
function of the contrast space of the previous observations. That is,
for any set of sites \( y_1, \ldots, y_k \), define \( F(y_1, \ldots, y_k) = (f(y_1) \ldots f(y_k)) \).
Choose \( C(y_k) \) to be any element of \( \mathcal{M}_{F'(y_k)} \), where \( \mathcal{M}(\cdot) \) is defined
in I.4, and \( \omega_q \) will be taken to mean \((\omega_1, \ldots, \omega_q)\) for any vector with
a subscript throughout this chapter. Call the \( m \) initial sites
\( x_m = (x_1, \ldots, x_m) \), and assume \( F_m(x_m) \) is of full rank, so that we must
have \( m > p \) for \( C(x_m) \) to be non-trivial. For \( 1 \leq k \leq n \), let \( Y_k \) be the
random value of the \( k \)'th site chosen after the initial \( m \) sites, and let
\( y_k \) be the observed value of \( Y_k \). Also define \( Z(\omega_q) = (z(\omega_1) \ldots z(\omega_q)) \),
and let \( Z^*(\omega_q) \) be the observed value of \( Z(\omega_q) \). Then \( C(\omega_q) Z(\omega_q) \) is a
basis for all contrasts of \( Z(\omega_q) \). We assume that \( y_k \) is of the form
\[
y_k = f_{k-1}[C(x_m, y_{k-1}) Z^*(x_m, y_{k-1})'; x_m, y_{k-1}],
\]
for some measurable function \( f_{k-1}(\cdot) \).

Now suppose we are trying to estimate \( G \), which may be vector-
valued. Likely candidates for \( G \) include \( a \), the vector of regression
coefficients, \( z(x_0) \), the value of the random field at a place it was
not observed, or an average of the random field over some region. For
any fixed vector of sites, \( \omega_q \), suppose \( \hat{G}(Z(\omega_q); \omega_q) \) is an estimator of
\( G \) which has the property
\[
\hat{G}(Z(\omega_q); \omega_q) - G \text{ and } C(\omega_q) Z(\omega_q)' \text{ are independent},
\]
where if \( C(\omega_q) \) has zero rows (there are no contrasts of \( Z(\omega_q) \)), then
(3) is assumed to be trivially satisfied. Then we have the following:
Theorem VI.1: Under conditions (2) and (3),

\[
\mathcal{L}\left[ \hat{G}(\tilde{Z}_{x_m}, \tilde{Z}_{x_n}; \tilde{x_m}, \tilde{x_n}) - G \mid \tilde{y_n} = y_n \right] = \mathcal{L}\left[ \hat{G}(\tilde{Z}_{x_m}, \tilde{Z}_{x_n}; \tilde{x_m}, \tilde{x_n}) - G \right],
\]

(4)

where \( \mathcal{L}(\cdot) \) means "the distribution of."

Before proving this result, let us consider two applications of it.

Example 1: Suppose \( e(\cdot) \) is a Gaussian field with mean zero and covariance structure specified up to a multiplicative constant, and let

\( G = z(x_0) \). For any set of fixed observations \( \tilde{\omega}_q \), let \( \hat{G} = \tilde{z}(x_0; \tilde{\omega}_q) \), the BLUE (and UMVUE) of \( z(x_0) \) based on \( \tilde{Z}(\tilde{\omega}_q) \), which is given in III.2.2.

Then \( \hat{z}(x_0; \tilde{\omega}_q) - z(x_0) \) is independent of \( \tilde{C}(\tilde{\omega}_q) \tilde{Z}(\tilde{\omega}_q)' \) by the projection property of BLUE's (see III.2.2) and the Gaussian assumption, so that (3) is satisfied. Then if \( \tilde{y}_n \) is picked in a method which satisfies (2), the theorem applies, and we have

\[
\mathcal{L}[\hat{z}(x_0; \tilde{x_m}, \tilde{y}_n) - z(x_0) \mid \tilde{y}_n = y_n] = \mathcal{L}[\hat{z}(x_0; \tilde{x_m}, \tilde{y}_n) - z(x_0)].
\]

(5)

That is, the distribution of the error of our interpolator is unaffected by the sequential situating.

In many cases, the mean will just be a constant; i.e., we will have \( \tilde{z}(x) \equiv \mu \). In this situation, the restriction on choosing \( \tilde{y}_n \), given by (2), corresponds to choosing the sites in a way which is invariant with respect to a location shift in the \( z(\cdot) \) field. That is, if instead of \( z(x) \), we observe \( \tilde{z}(x) = z(x) + C \), we will get the same value for \( \tilde{y}_n \). Thus, in this case, equation (2) is a weak restriction on the rule for picking \( \tilde{y}_n \).

Example 2: Making the same assumptions about \( e(\cdot) \) as in the previous example, consider the problem of estimating \( \tilde{\varphi} \). For a fixed set of
sities $\omega_q$, let $\hat{a}(\omega_q)$ be the BLUE of $a$ based on $Z(\omega_q)$. As in the previous example, if $Y_n$ is picked in accordance with condition (2), we will have

$$\mathcal{L}(\hat{a}(x_m, y_n) - a | y_n = y_n) = \mathcal{L}(\hat{a}(x_m, y_n) - a).$$  \hspace{1cm} (6)

We now prove the theorem.

**Proof:** From (2), we have

$$Y_k = f_{k-1} \left\{ C(x_m, y_{k-1}) \left( Z(x_m) \tilde{Z}(y_{k-1}) \right)'; x_m, y_{k-1} \right\}. \hspace{1cm} (7)$$

Now $y_{k-1}$ is a function of

$$\{ C(x_m, y_{k-2}) \left( Z(x_m) \tilde{Z}(y_{k-2}) \right)'; x_m, y_{k-2} \}. \hspace{1cm}$$

Therefore, $Y_k$ can be written as a function of

$$\{ C(x_m, y_{k-1}) \left( Z(x_m) \tilde{Z}(y_{k-1}) \right)'; x_m, y_{k-2} \};$$

i.e., $Y_k$ is expressible as a function which does not include $y_{k-1}$ as an explicit argument. Repeating this argument, we can conclude that there exists a measurable function $f_{k-1}^*$ such that

$$Y_k = f_{k-1}^* \left\{ C(x_m, y_{k-1}) \left( Z(x_m) \tilde{Z}(y_{k-1}) \right)'; x_m \right\}. \hspace{1cm} (8)$$

Now let us consider the event $x_n = y_n$. We will show that its occurrence depends only on the value of $C(x_m, y_n) \{ Z(x_m) \tilde{Z}(y_n) \}'. First, $Y_1 = y_1$ is equivalent to

$$\{ C(x_m) \tilde{Z}(x_m)' ; x_m \} \in f_{k-1}^{*-1}(y_1), \hspace{1cm} (9)$$

where $f_{k-1}^{*-1}(\cdot)$ gives the inverse image of $f_{k-1}^*$. And, using (8), $Y_2 = y_2$ is equivalent to
\( \{ C(x_m, Y_1) [Z(x_m) z(Y_1)]'; x_m \} \in f_1^*(y_2). \)  
\[ (10) \]

Thus, \( (Y_1, Y_2) = (y_1, y_2) \) is equivalent to

\[
\begin{align*}
\{ C(x_m)[Z(x_m)]'; x_m \} \in f_0^{* -1}(y_1), & \quad \text{and} \\
\{ C(x_m, y_1)[Z(x_m) z(y_1)]'; x_m \} \in f_1^{* -1}(y_2),
\end{align*}
\]

\[ (11) \]

since \((9)\) implies that \( Y_1 \) can be replaced by \( y_1 \) in \((10)\). Repeating this argument, we can conclude that \( Y_n = y_n \) is equivalent to

\[
\begin{align*}
\{ C(x_m)[Z(x_m)]'; x_m \} \in f_0^{* -1}(y_1), & \\
\vdots & \\
\{ C(x_m, y_{n-1})[Z(x_m) z(y_{n-1})]' x_m \} \in f_{n-1}^{* -1}(y_n)
\end{align*}
\]

\[ (12) \]

Since \( x_m \) is not random, we see that \( Y_n = y_n \) is measurable with respect to

\( C(x_m, y_n)[Z(x_m) z(y_n)]' \). That is, we can say

\[
Y_n = y_n \iff C(x_m, y_n)[Z(x_m) z(y_n)]' \in T(y_n),
\]

\[ (13) \]

where \( T(y_n) \) is a measurable set.

Then,

\[
\mathbb{P}\left( \hat{G}(Z(x_m), Z(Y_n); x_m, Y_n) - G \mid Y_n = y_n \right) = \mathbb{P}\left( \hat{G}(Z(x_m), Z(Y_n); x_m, Y_n) - G \mid Y_n = y_n \right)
\]

\[
= \mathbb{P}\left( \hat{G}(Z(x_m), Z(y_n); x_m, y_n) - G \mid C(x_m, y_n)[Z(x_m) z(y_n)]' \in T(y_n) \right)
\]

\[ (by \ (13)) \]

\[
\mathbb{P}\left( \hat{G}(Z(x_m), Z(y_n); x_m, y_n) - G \right), \ (by \ (3)),
\]

which is the required result.

For the sake of generality, this result has been stated in a rather arcane manner. Hopefully, the two examples illustrated the point of the theorem. We could further generalize the result by allowing \( n \) to be a
finite random stopping time $N$. If the event \{N=n\} is measurable with respect to $C(x_m, y_n)(Z(x_m) Z(y_n))'$, and (3) is true for all possible values of $n$, then the theorem will still be true with $n$ replaced by $N$.

An important practical limitation of the two applications of the theorem is the assumption that the covariance structure is specified up to a multiplicative constant. In practice, the covariance structure cannot be a priori specified up to a multiplicative constant, so it will have to be estimated from the data. Unfortunately, the distribution of an estimator of the covariance structure which uses $Z(y_n)$ will, in general, depend on the way in which $y_n$ was chosen. If $m$, the initial number of observations, is large, we may be able to estimate the covariance structure adequately using only $Z(x_m)$. However, if $m$ is not sufficiently large, we will also want to use the other $n$ observations. The simplest strategy would be to pretend that all of the sites were chosen ahead of time and use the methods described in Chapters II and III on all $m+n$ observations. However, such a strategy could conceivably lead to severely biased estimators of the covariance structure, although the exact effect is unclear, as no work has been done on this problem. In any case, the two applications are only valid when the true covariance, not the estimated covariance, is used to generate the estimators (see I.1.2). Thus, in practice, the conditions of the theorem will not be met exactly. However, if the covariance structure can be estimated well, then we can hope that the properties of the estimators in the examples will be only slightly affected by the sequential siting.
APPENDIX F: Some Notes on Spatio-Temporal Variograms

In this appendix, we prove some basic results about the quantity \( \gamma^* \) defined in Equation (21) of Chapter V. Recall that the semi-variogram of the disturbance term is given by

\[
\frac{1}{2} E(e(x_1, t_j) - e(x_k, t_\ell)^2 = \begin{cases} 
\gamma(x_1 - x_k), & j = \ell, \\
\gamma^*, & j \neq \ell,
\end{cases}
\]

which is a sort of temporal uncorrelatedness. It is clear that there must be some restrictions on the value of \( \gamma^* \) if we are to have a permissible semi-variogram. We now prove some results to this effect.

One consequence of these theorems is that if \( \gamma(\cdot) \) has a well-defined, finite sill, then the semi-variogram in (1) will be valid if and only if \( \gamma^* \) is greater than the sill value (see I.1.3).

**Theorem F.1:** Define for the vector \( r \)

\[
K(r) = \lim_{b \to \infty} \inf \gamma(br). \tag{2}
\]

Let

\[
K = \sup_r K(r). \tag{3}
\]

Then if \( \gamma^* < K \), the spatio-temporal semi-variogram defined in (1) is not valid.

**Proof:** Assume \( \gamma^* < K \). By (3) and the definition of \( K(r) \), there exists \( r \) and \( \epsilon > 0 \) which satisfy

\[
\gamma(nr) > \gamma^* + \epsilon \quad \text{for} \quad n=1,2,3,\ldots. \tag{4}
\]

Let us consider

\[
W_n = \begin{pmatrix}
e(r, t_2) - e(0, t_1) \\
e(2r, t_2) - e(r, t_1) \\
\vdots \\
e(nr, t_2) - e((n-1)r, t_1)
\end{pmatrix}. \tag{5}
\]
Now, \( \psi_n \) is a vector of contrasts, so the semi-variogram defined in (1) must yield that \( \psi_n = \text{Cov}(\psi'_n, \psi'_n) \) is p.s.d. Now, the diagonal elements of \( \psi_n \) all equal \( 2\gamma^* \). And, the off-diagonal elements satisfy

\[
\text{Cov}\{e(jr, t_2) - e((j-1)r, t_1), e(kr, t_2) - e((k-1)r, t_1)\} = 2\gamma^* - 2\gamma((j-k)r) \leq -2\varepsilon,
\]

for \( j \neq k \), by (4). Thus

\[
e' \psi_n e \leq 2n\gamma^* - 2n(n^2 - n)\varepsilon < 0,
\]

for \( n \) sufficiently large, which contradicts the requirement that \( \psi_n \) be p.d. Therefore, we cannot have \( \gamma^* < K \). Note that this proof is valid even if \( K = \infty \), so we are done. Also, in most situations, \( \gamma(\cdot) \) will have a sill value (possibly infinite), in which case, \( K \) will equal the sill value.

We now prove a converse of sorts.

**Theorem F.2:** Suppose \( L - \gamma(\cdot) \) is a valid spatial covariance function.

Then if \( \gamma^* \gg L \), the spatio-temporal semi-variogram defined in (1) is valid.

**Proof:** Consider a set of space-time coordinates,

\((x_1^1, t_1), \ldots, (x_{n_1}^1, t_1), (x_1^2, t_2), \ldots, (x_{n_2}^2, t_2), \ldots, (x_1^k, t_k), \ldots, (x_{n_k}^k, t_k)\).

We need to show, for any set of weights \( \lambda_{11}, \ldots, \lambda_{n_k} \) satisfying

\[
\sum_{j=1}^{k} \sum_{i=1}^{n} \lambda_{ij} = 0, \quad \text{that}
\]

\[
- \sum_{i,j,k,l} \lambda_{ij} \lambda_{kl} \frac{1}{4} \text{E}(e(x_i^j, t_j) - e(x_k^l, t_l))^2 \geq 0.
\]
Now
\[- \sum_{ijkl} \lambda_{ij} \lambda_{kl} \frac{1}{2} E(e(x_i^j, t_j) - e(x_k^l, t_l))^2 = \]
\[- \sum_{ijkl} \lambda_{ij} \lambda_{kl} \left[ \gamma(x_i^j - x_k^l) I_{\{j = l\}} + \gamma^* I_{\{j \neq l\}} \right] = \]
\[\sum_{ijk} \lambda_{ij} \lambda_{kj} (\gamma^* - \gamma(x_i^j - x_k^j)) \]
(using Equation (8))
\[= (\gamma^* - L) \sum_{ijk} \lambda_{ij} \lambda_{kj} + \sum_{ijk} \lambda_{ij} \lambda_{kj} (L - \gamma(x_i^j - x_k^j)) \]
\[= (\gamma^* - L) \sum_{j} (\sum_{i} \lambda_{ij})^2 + \sum_{j} \left[ \sum_{ik} \lambda_{ij} \lambda_{kj} (L - \gamma(x_i^j - x_k^j)) \right].\]
The first term is positive since $\gamma^* > L$, and
\[\sum_{ik} \lambda_{ij} \lambda_{kj} (L - \gamma(x_i^j - x_k^j)) > 0\]
for each $j$ by the assumption that $L - \gamma(\cdot)$ is a valid covariance function, so we are done. This proof is valid if $\gamma^* = \infty$, in which case, the variance of a contrast will be infinite unless $\sum_{i} \lambda_{ij} = 0$ for all $j$.

Finally, we show that if $\gamma(\cdot)$ has a well-defined, finite sill value, then $L$ in the previous theorem can be taken to be the sill value. We will say that $L$ is the sill of $\gamma(\cdot)$ if given any $\varepsilon > 0$, we can find a distance $d$ such that
\[|\gamma(r) - L| < \varepsilon \text{ for all } r \text{ satisfying } |r| > d. \quad (10)\]

**Theorem F.3:** If $\gamma(\cdot)$ is a valid spatial semi-variogram with a sill value $L < \infty$, then $L - \gamma(\cdot)$ is a valid spatial covariance function.

**Proof:** For any given set of sites $x_1, \ldots, x_n$ and weights $\lambda_1, \ldots, \lambda_n$, we need to show
\[\sum_{i,j=1}^{n} \lambda_i \lambda_j (L - \gamma(x_i^j - x_j^j)) > 0. \quad (11)\]
Now, by the definition of \( L \), and using \( L < \infty \), given \( \varepsilon > 0 \), we can find a vector \( r \) such that

\[
|\gamma(x_i + r - x_j) - L| < \varepsilon \quad \text{for } i, j = 1, \ldots, n. \tag{12}
\]

Define \( x_{n+1} = r + x_i \) for \( i = 1, \ldots, n \), and \( \lambda_{n+1} = -\lambda_i \) for \( i = 1, \ldots, n \).

Then, since \( \gamma(\cdot) \) is a valid semi-variogram, and \( \sum_{i=1}^{2n} \lambda_i = 0 \), we have

\[
0 \leq -\sum_{i,j=1}^{2n} \lambda_i \lambda_j \gamma(x_i - x_j) - \sum_{i,j=n+1}^{2n} \lambda_i \lambda_j \gamma(x_i - x_j)
\]

\[
- \sum_{i=1}^{n} \sum_{j=n+1}^{2n} \lambda_i \lambda_j \gamma(x_i - x_j) - \sum_{i=n+1}^{2n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j) =
\]

\[
0 < -2 \sum_{i,j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j) + \sum_{i,j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j - r) + \sum_{i,j=1}^{n} \lambda_i \lambda_j \gamma(x_i + r - x_j). \tag{13}
\]

Now,

\[
\left| \sum_{i,j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j - r) - L \sum_{i,j=1}^{n} \lambda_i \lambda_j \right| =
\]

\[
\left| \sum_{i,j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j - r) - L \right| \leq
\]

\[
\sum_{i,j=1}^{n} |\lambda_i \lambda_j| \left| \gamma(x_i - x_j - r) - L \right| \leq \varepsilon \sum_{i,j=1}^{n} |\lambda_i \lambda_j|, \text{ by (12)}. \]

By taking \( r \) sufficiently large, we can make \( \varepsilon \) arbitrarily small. We can obtain a similar expression for the last term in (13). Thus, we can conclude,

\[
0 \leq -2 \sum_{i,j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j) + 2L \sum_{i,j=1}^{n} \lambda_i \lambda_j = 2 \sum_{i,j=1}^{n} \left( L - \gamma(x_i, x_j) \right) \lambda_i \lambda_j, \tag{14}
\]
as required. This last theorem illustrates a basic relationship between variograms and covariance functions which has been generally assumed to be true, although there does not appear to be a published proof of this result.
References


