

**A Comparative Study of Spatial Analysis Methods  
for Forestry Nelder Experiments**

by

David L.R. Affleck

B.Sc. (Forest Science), University of British Columbia, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

**Master of Science**  
in  
THE FACULTY OF GRADUATE STUDIES

THE FACULTY OF FORESTRY

Department of Forest Resources Management

We accept this thesis as conforming  
to the required standard

**The University of British Columbia**

July 2001

© David L.R. Affleck, 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Forest Resources Management

The University of British Columbia  
Vancouver, Canada

Date July 27, 2001

# Abstract

The Nelder (1962) series of systematic spacing wheel designs define compact and spatially explicit layouts for forestry spacing experiments. One difficulty in the analysis of Nelder experiments is that the compact arrangements of trees may result in significant correlations among neighbouring variable values when these are geographically ordered. When this is the case, classical analysis methods, based on random sampling models and ordinary least squares estimators, provide inefficient estimates of treatment effects and biased estimates of variance parameters. However, the formal geometric structure of Nelder experiments presents an opportunity to evaluate the properties of alternative analysis methods based on spatial correlated error models.

The statistical and practical properties of four spatial analysis methods were evaluated in the context of the interpretation of data from forestry Nelder (1962) experiments. The spatial analysis methods considered were based on either regionalized variables or nearest neighbour models and restricted maximum likelihood estimators. The validity and relative efficiency of the spatial analysis methods were assessed under different magnitudes of spatial autocorrelation through a simulation study. The same spatial analysis methods were also applied to data from a Nelder experiment in the University of British Columbia Malcolm Knapp Research Forest in Haney, British Columbia.

The spatial analysis methods based on regionalized variables models were valid and more efficient than the classical analysis method when data were strongly spatially correlated. The spatial analysis methods based on nearest neighbour models were not valid and were less efficient than the classical analysis method. Application of the spatial analysis methods required more intensive estimation procedures.

# Contents

|   |            |
|---|------------|
| <b>Abstract</b>                                       | <b>ii</b>  |
| <b>Contents</b>                                       | <b>iii</b> |
| <b>List of Tables</b>                                 | <b>v</b>   |
| <b>List of Figures</b>                                | <b>vi</b>  |
| <b>Acknowledgements</b>                               | <b>vii</b> |
| <b>1 Introduction</b>                                 | <b>1</b>   |
| <b>2 Nelder Experiments</b>                           | <b>3</b>   |
| 2.1 Spacing Experiments . . . . .                     | 3          |
| 2.2 Nelder Spacing Wheel Designs . . . . .            | 6          |
| 2.3 The MKN Experiment . . . . .                      | 9          |
| 2.4 Randomization and Statistical Inference . . . . . | 11         |
| <b>3 Analysis Methods for Nelder Experiments</b>      | <b>15</b>  |
| 3.1 Analysis Methods . . . . .                        | 15         |
| 3.1.1 The General Linear Model . . . . .              | 15         |
| 3.1.2 Statistical Properties . . . . .                | 17         |
| 3.2 Classical Analysis Methods . . . . .              | 19         |
| 3.2.1 The ANOVA Model . . . . .                       | 19         |
| 3.2.2 OLS Estimation . . . . .                        | 20         |
| 3.2.3 Robustness . . . . .                            | 21         |
| 3.3 Spatial Analysis Methods . . . . .                | 22         |
| 3.3.1 Spatial Correlated Error Models . . . . .       | 22         |

|          |   |           |
|----------|---|-----------|
| 3.3.2    | REML Estimation . . . . .                           | 29        |
| 3.4      | Model Selection . . . . .                           | 32        |
| <b>4</b> | <b>Methods</b>                                      | <b>34</b> |
| 4.1      | Simulation Study . . . . .                          | 34        |
| 4.2      | Analysis of the MKN Experiment . . . . .            | 38        |
| <b>5</b> | <b>Results and Discussion</b>                       | <b>40</b> |
| 5.1      | Results . . . . .                                   | 40        |
| 5.1.1    | Simulation Study . . . . .                          | 40        |
| 5.1.2    | MKN Analysis . . . . .                              | 45        |
| 5.2      | Discussion . . . . .                                | 47        |
| 5.2.1    | Simulation Study . . . . .                          | 48        |
| 5.2.2    | MKN Analysis . . . . .                              | 53        |
| <b>6</b> | <b>Conclusion</b>                                   | <b>55</b> |
|          | <b>Literature Cited</b>                             | <b>57</b> |
|          | <b>Appendix A REML Estimation</b>                   | <b>63</b> |
|          | <b>Appendix B On the Equivalence of OLS and GLS</b> | <b>67</b> |

# List of Tables

|     |   |    |
|-----|---|----|
| 2.1 | Tree density levels in the MKN experiment, by sector. . . . .   | 11 |
| 4.1 | Simulation models, analysis methods, and parameter values used in the simulation study. . . . .   | 35 |
| 4.2 | Simulated treatments effects by tree density class. . . . .   | 35 |
| 4.3 | Analysis of variance table for the simulated Nelder experiments. . . . .  | 38 |
| 4.4 | Analysis of variance table for the MKN experiment DBH15 data set. . . . .   | 38 |
| 5.1 | Mean and standard deviations of OLS and REML variance parameter estimators at each level of spatial autocorrelation, by simulation trial. . . . . | 42 |
| 5.2 | ANOVA/OLS analysis of variance table for the MKN experiment DBH15 data set. . .   | 45 |
| 5.3 | Estimates of variance parameters, AIC, and restricted LR statistics for the MKN DBH15 data set, by analysis method. . . . .                       | 46 |
| 5.4 | Rankings of estimated tree density effects for the MKN DBH15 dataset, by analysis method. . . . .   | 47 |

# List of Figures

|     |  |    |
|-----|--|----|
| 2.1 | Nelder (1962) systematic spacing designs. . . . .  | 6  |
| 2.2 | A Nelder (1962) type Ia spacing wheel. . . . .   | 7  |
| 2.3 | Schematic of the Nelder (1962) type Ia spacing wheel. . . . .  | 7  |
| 2.4 | Layout of the MKN experiment. . . . .  | 10 |
| 2.5 | Aerial photograph of the MKN experiment taken in 1987. . . . .   | 12 |
| 3.1 | Rate of decay of the pairwise covariance between observations as a function of their separation distance under the exponential and Gaussian covariance functions. . . . .  | 28 |
| 5.1 | Relative variance bias and relative empirical variance of analysis methods by simulation trial. . . . .  | 41 |
| 5.2 | Empirical rejection rates of among- and within-class treatment $F$ -tests by analysis method, level of spatial autocorrelation, and simulation trial. . . . .  | 43 |
| 5.3 | Empirical rejection rates of the restricted LR test, and the rates at which the AIC for ANOVA/OLS exceeded the AIC for the spatial analysis method, in each simulation trial as a function of the spatial autocorrelation. . . . . | 44 |
| 5.4 | Frequency distribution and residual plot for the standardized ANOVA/OLS MKN DBH15 residuals. . . . .   | 46 |

# Acknowledgements

I am grateful to Dr. Valerie LeMay, my supervisor, for her assistance and patience, and for many helpful comments concerning the ideas embodied in the present work and in previous drafts. I am also grateful to the other members of my committee, Dr. Antal Kozak, Dr. Peter Marshall, and Dr. David Tait, for their advice and support. Albert Nussbaum of the B.C. Ministry of Forests, Research Branch, generously provided the data for the Malcolm Knapp Research Forest Nelder experiments.

Funding for this research was provided by a series of University Graduate Fellowships administered by the University of British Columbia and by fellowships endowed by the Sopron Alumni and the Li Tze Fong family. A Natural Sciences and Engineering Research Council of Canada research assistantship within the Department of Forest Resources Management also provided financial support. Guillaume Thérien and J.S. Thrower and Associates Ltd. provided me the time and support I needed during the final months of thesis preparation.

Finally, I would like to thank Willow, for her support and encouragement throughout.

DAVID L.R. AFFLECK

*The University of British Columbia  
July 2001*

# Chapter 1

## Introduction

Field experiments are used in forest research to evaluate the effects of different treatments or stand management regimes. Many forestry field experiments consist of large experimental units and span large areas. Spacing experiments based on the Nelder (1962) series of spacing wheel designs are notable exceptions. Individual trees are the basic units of study in these experiments, and are distributed in carefully controlled and relatively compact arrangements. The spatial dimensions of Nelder spacing wheel designs make them suitable candidates for evaluating the statistical and practical properties of spatial analysis methods.

Classical analysis methods are based on the principles of random sampling and mutual independence. Data from forestry Nelder experiments may not satisfy the condition of independence because some trees grow in more similar environments than others. Given the relatively compact arrangement of trees in Nelder designs, this may result in appreciable correlations in the growth habits of neighbouring trees.

Classical analysis methods cannot account for spatial autocorrelation among observations directly. When data in field experiments are spatially autocorrelated, treatment effect estimators for classical analysis methods are not efficient and variance estimators are biased (Cressie 1993). The device of experimental randomization can be relied upon to neutralize the effects of spatial autocorrelation over repeated trials (Harville 1975). However, it is generally not feasible to repeat experiments. Also, in Nelder spacing wheel designs, spatial randomization of treatments is not possible.

In this thesis, spatial analysis methods based on correlated error (CE) models are examined. Spatial CE models allow for spatially explicit autocorrelation among geo-referenced data. The potential advantages of spatial analysis methods are increased precision in treatment effect estimation and unbiased variance estimation (Cressie 1993). Estimates of spatial autocorrelation

parameters may also provide useful insights into the form and pattern of uncontrolled sources of variation. However, spatial analysis methods require more involved estimation and computation techniques. The specific objectives of this thesis were:

1. to review the effects of spatial autocorrelation among data in field experiments on the performance of classical analysis methods consisting of analysis of variance (ANOVA) models and ordinary least squares (OLS) estimators;
2. to examine the efficiency and validity of four spatial analysis methods consisting of spatial CE models and restricted maximum likelihood (REML) estimators, for the interpretation of forestry Nelder experiments; and
3. to evaluate the practicality of spatial analysis methods for forestry Nelder experiments.

A description of the Nelder (1962) series of systematic spacing designs and a review of different perspectives on experimental inference are provided in Chapter 2. In Chapter 3, the effects of spatial autocorrelation on the OLS estimators of treatment effects and variance parameters are examined. Spatial CE models and REML estimators for spatial analysis methods are also introduced in that chapter. The second and third objectives of this thesis were addressed by means of the application of classical and spatial analysis methods to simulated and actual Nelder experiments, as described in Chapter 4. The results of these two applications are presented and discussed in Chapter 5. Chapter 6 is included as a summary.

## Chapter 2

# Nelder Experiments

### 2.1 Spacing Experiments

In 1943, Toumey and Korstian (p.18) wrote that one of the chief aims of silviculture was to bring about a stand of "crop trees of best form and most rapid growth consistent with fullness of stand and quality of product." Although somewhat dated, the preceding statement is of particular value because of its explicit recognition of the relationships among tree form, growth rate, and density (*i.e.*, stems per hectare). Silviculturalists may strive to improve site productivity and to ensure the establishment of desired crop species, but managing the distribution of productivity among and within individual crop trees is frequently of equal importance. Often this latter objective is achieved through the regulation of tree density.

Silviculturalists may regulate tree density at any age through thinnings. However, because controlling tree characteristics such as lateral branching early in stand development can greatly influence the future economic value of a crop (Nyland 1996), attending to seedling or sapling density can be a better investment. Other economic incentives and legal responsibilities may also render rapid initial tree growth and appropriate re-stocking levels expedient objectives. A detailed knowledge of the effects of initial density on the growth of forest tree crops can therefore be of considerable value.

Evaluating the effects of tree density on tree growth has been the focus of many spacing experiments in forestry. Smith (1959) discussed some of the early approaches to studying the problem and Evert (1971, 1984) provided thorough reviews and bibliographies of the subject. The archetypal forestry spacing experiment consists of a number of trees that are grown at different densities by arranging them into a collection of rectangular experimental units. After a period of time, some attribute of the trees is measured (*e.g.*, diameter or survivorship) and trees grown at

different densities are compared.

These experiments are largely based on the same design principles as agricultural spacing experiments, but differ in two fundamental ways. First, agricultural crops often take only a single season or year to fully mature, making agricultural spacing experiments relatively short-lived. Forestry spacing experiments that persist for years or decades may become increasingly complicated by data compilation issues and tree mortality. Second, agricultural spacing experiments are typically designed to study the effects of both the size and rectangularity<sup>1</sup> of the growing space on plant growth. Few spacing experiments in forestry have examined the effects of the spatial arrangement of trees independently of the effects of tree density; the study by Smith (1978) is a notable exception. It has generally been held that tree growth will not be unduly influenced by rectangularity, except at extremes, since tree crops have more time to adjust to the dimensions of their growing space (Namkoong 1966). At the same time, the degree of mechanization and control applied in the establishment of agricultural crops is generally not feasible for forest tree crops under operational conditions.

The design of forestry spacing experiments is greatly complicated by the nature of the treatment employed, namely, tree density. The design of field experiments typically proceeds from the identification of discrete plots of ground as the experimental units. Appreciable systematic heterogeneity can be found across and within experimental units, so replication and interspersions of treatments are desired. In many field experiments, it is possible to allocate treatments to experimental units subsequent to their demarcation. Then experimental randomization (*i.e.*, random allocation of treatments to experimental units) can easily be accomplished. However, in forestry spacing experiments, the factor of interest is inter-tree competition, represented as a set of treatment levels by a range of tree densities. As such, the delineation of experimental units and the allocation of treatments must be done simultaneously.

Under these circumstances, two options exist for allocating different levels of tree density to experimental units to permit random allocation or interspersions of treatments. One option is to fix the dimensions of the experimental units and plant (or space) each to the desired density. This option permits the experimenter to employ experimental randomization. However, the use of experimental units of a fixed size precludes the use of a single tree per experimental unit for all but the lowest density treatments. At higher densities, the experimental units must contain more trees. Therefore, this strategy entails a less efficient use of materials. Single-tree experimental units are economical in the use of both trees and space, and can lead to more efficient designs than

---

<sup>1</sup>The ratio of the dimensions of the growing space (*i.e.*, length : width).

multiple-tree experimental units (Loo-Dinkins and Tauer 1987). Multiple-tree experimental units do allow for estimation of within-experimental-unit variability, but different numbers of trees per experimental unit results in differential precision of the estimates. Also, at high densities there can be far more trees than necessary for this purpose (Namkoong 1966, Smith 1978). This results in the collection of redundant information and a potentially wasteful use of trees.

The alternative option for treatment arrangement in spacing experiments is to use a constant number of trees in each experimental unit. Experimental units with different tree densities will then have different areas. Unfortunately, this can result in considerable variability in the size of the experimental units, especially when multiple-tree experimental units are used or a broad range of tree densities is studied. In principle, variable-sized experimental units could be randomly allocated over the experimental field (or within blocks). In practice, this is impossible; it would not allow for close packing of plots and would result in a very inefficient use of space. Instead, the experimental field must be partitioned (systematically) into variable-sized experimental units. Often this results in the use of a single experimental unit for each density class, with sub-sampling supplying an estimate of experimental error (*e.g.*, the study described by Reukema 1979). However, replication and interspersed tree density classes is possible; Smith (1978) described a spacing experiment using multiple-tree, variable-sized experimental units where replicates were interspersed over the field.

Ultimately, the most detracting element of spacing experiments employing experimental units with either a constant area or a constant number of trees, is the need for a large number of buffer trees. Maintaining specific tree densities within each experimental unit, while allowing for some degree of randomization (or at least interspersed) of treatments, necessarily prevents the use of spatially contiguous plots. Buffer trees are required around the perimeter of all experimental units, but do not contribute any information for the estimation of treatment effects. Moreover, the proportion of the experimental area (and trees) lost to buffer elements increases as the size of the experimental units is decreased. Thus, in one sense, any efforts to trade off experimental unit size for increased replication results in even less economical use of space and trees.

In summary, either constant area or constant number of tree experimental units can be used in spacing experiments in order to adhere to some of the conventions of classic field experiment methodology, such as experimental randomization. However, neither option makes very economical use of trees or space. As a result, researchers have had to consider using alternative designs for forestry spacing experiments. Among such alternatives are Nelder (1962) systematic spacing designs.

## 2.2 Nelder Spacing Wheel Designs

In 1962, Nelder published a description of a series of systematic experimental designs specifically for spacing trials. Two essential features of this series of designs are that the experimental units consist of single trees and that tree density, rectangularity, or both, are varied in a continuous and systematic manner across the field. Nelder described five designs: type Ia, in which only tree density is varied; type Ib, in which only rectangularity is varied; and types Ic, Id, and II, in which these two factors are varied together but in different directions (Figure 2.1). The first two designs (types Ia and Ib) can be formulated so as to run through any angle, including  $360^\circ$ . In this case

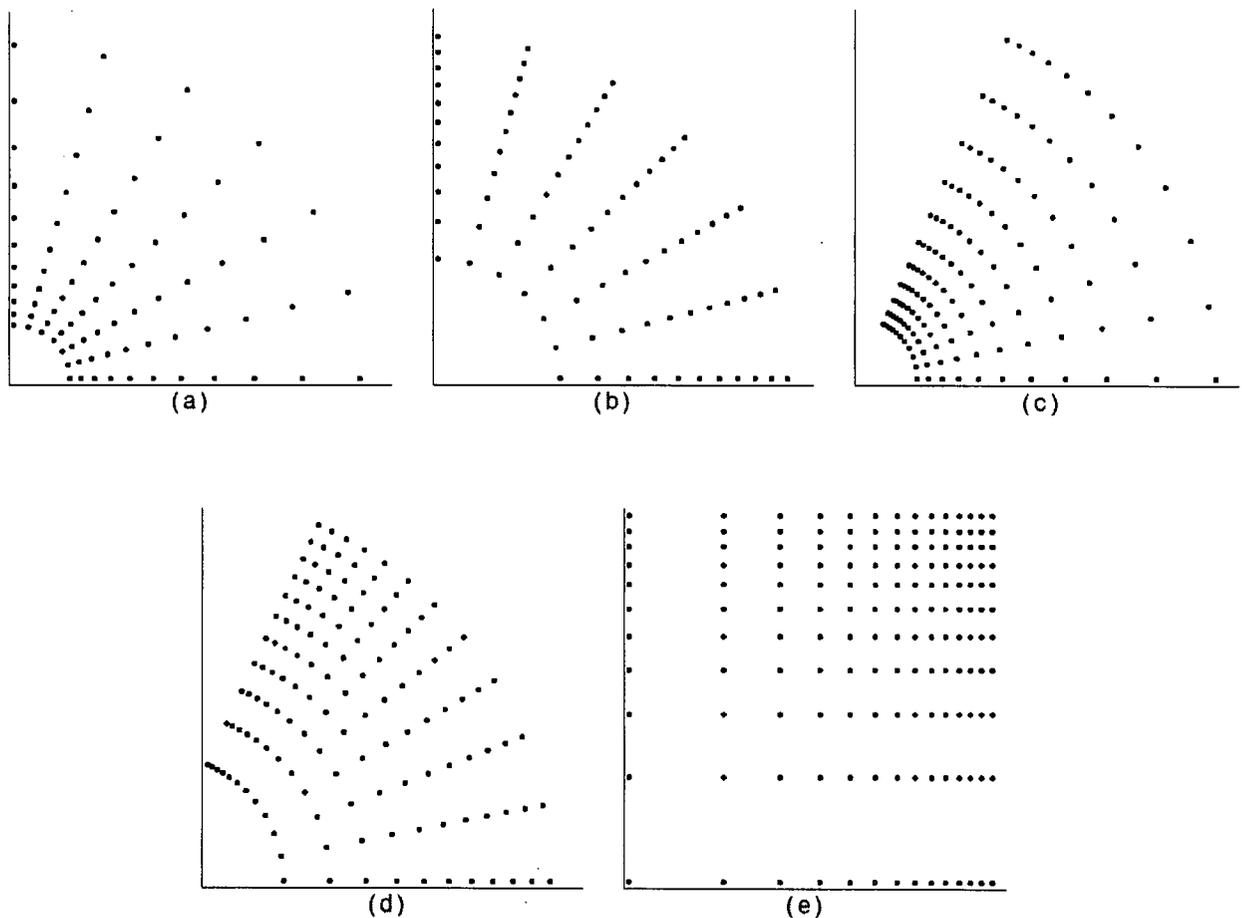


Figure 2.1: Five systematic designs for spacing experiments: (a) type Ia: density decreases with increasing distance from the origin (rectangularity fixed); (b) type Ib: rectangularity decreases with distance from the origin (density fixed); (c) type Ic: density changes along the horizontal while rectangularity changes on the vertical; (d) type Id: density changes on the vertical while rectangularity changes on the horizontal; (e) type II: density and rectangularity change along both axes but in different directions. Adapted from Nelder (1962).

they take on the appearance of a spoked wheel (Figure 2.2). As a result, these designs are often referred to as spacing wheels. In the following text, the term “Nelder experiment” will refer to spacing experiments consisting exclusively of one or more of these spacing wheels.

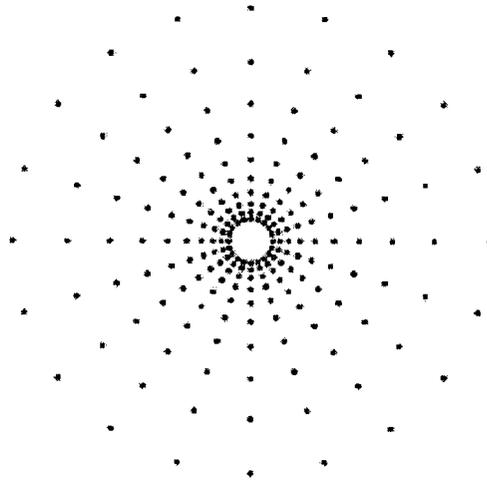


Figure 2.2: A Nelder (1962) type Ia spacing wheel spanning 360°.

Of the five designs proposed by Nelder (1962), the type Ia has been most widely applied in forestry research (*e.g.*, Freeman 1964; Zedaker 1982; Cole and Newton 1987; Galinski *et al.* 1994). In the type Ia spacing wheel, trees are positioned along spokes radiating from the centre at a constant angle ( $\varpi$ ) and concentric arcs evolve at increments ( $r_i$ ) along these spokes (Figure 2.3). Tree density decreases (*i.e.*, growing space per tree increases) along the spokes while the arcs form

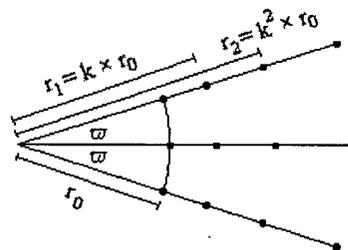


Figure 2.3: Schematic of the Nelder (1962) type Ia spacing wheel showing the angle ( $\varpi$ ) between spokes and the arc radii ( $r_i$ ).

iso-density contours. Rectangularity is constant throughout. Because trees are located at the intersections of arcs and spokes, the design is most easily described in terms of polar co-ordinates.

The radii (in metres) of the concentric arcs form a geometric progression given by

$$r_i = r_{i-1} \times k = r_0 \times k^i \quad (2.1)$$

where  $r_i$  is the radius of the  $i$ th arc ( $i = 1, 2, \dots, t, (t+1)$ ),  $r_0$  is the radius of the innermost arc,  $k$  is a rate constant, and  $(t+1)$  is the index of the outermost arc. The innermost and outermost arcs of a spacing wheel act as buffers so there are only  $t$  experimental levels of tree density. If the design spans less than  $360^\circ$ , then trees on two edge spokes must also stand as guards.

The area per tree ( $A_i$ ; in  $m^2$ ) and the tree density ( $D_i$ ; in stems/ha) anywhere on the  $i$ th arc in a type Ia spacing wheel are given by

$$A_i = r_i^2 \varpi (k - k^{-1})/2 = r_0^2 k^{2i} \varpi (k - k^{-1})/2 \quad (2.2)$$

and

$$D_i = 10000(A_i)^{-1} \quad (2.3)$$

for  $i = 1, 2, \dots, t$ . The rectangularity ( $\nu$ ) anywhere in a type Ia is

$$\nu = \varpi / (k^{\frac{1}{2}} - k^{-\frac{1}{2}}) \quad (2.4)$$

The type Ia spacing wheel can be completely described by specifying  $t$ ,  $r_0$ ,  $k$ , and  $\varpi$ . Alternatively, the latter three parameters can be derived given  $t$ ,  $\nu$ , and the range of tree densities of interest. Denoting the areas per tree at the endpoints of the density range by  $A_1$  and  $A_t$ , the system of equations:

$$\begin{cases} \log k = \log \left( \frac{A_1}{A_t} \right) / (2 - 2t) \\ \varpi = \nu \left( k^{\frac{1}{2}} - k^{-\frac{1}{2}} \right) \\ r_0 = \sqrt{2A_1 / \varpi (k^3 - k)} \end{cases} \quad (2.5)$$

can be solved to yield  $k$ ,  $\varpi$ , and  $r_0$ .

In his original paper, Nelder (1962) noted that as the rate constant  $k$  is increased, the deviation in plant position from the centroid of the quadrangle formed by its four nearest neighbours can become substantial. The distances to the two nearest trees on neighbouring spokes are the same, but the distances to the two nearest trees on the same spoke are not. Nelder defined the non-centrality ( $C_o$ ) of a type Ia spacing wheel as the deviation in the position of a plant from the centroid of its growing space, taken a percentage of the length of the growing space along the spoke.

This can be expressed as

$$C_o = \frac{2 - (k + k^{-1})}{2(k - k^{-1})} \times 100\% \quad (2.6)$$

Nelder advised that  $C_o$  should be maintained below 5% (*i.e.*,  $k < 1.1$ ) in agricultural spacing experiments to prevent bias in the estimation of treatment effects. However, Namkoong (1966) pointed out that, just as with rectangularity, because trees will have more time to adapt to the shape of their growing spaces, the effects of non-centrality are likely to be less significant in forestry spacing experiments. Many forestry spacing experiments employing the type Ia design have used  $k > 1.1$  (*e.g.*, Zedaker 1982; Paltzat 1984; Cole and Newton 1987; Galinski *et al.* 1994).

Several variations on the original Nelder (1962) spacing wheels have been proposed. In 1966, Namkoong described a number of designs based on the Nelder type Ia spacing wheel, in which rectangularity was varied over a relatively narrow range (approximately 1:1 to 1:4). The rationale for these modified designs was that the range of tree densities to be studied in the type Ia spacing wheel is spanned by a geometric series and most of the density levels fall towards the upper end of the range. Namkoong recognized that spacing wheels with constant density increments could be formulated if rectangularity were varied. More recently, Rustagi (1984) modified the Nelder design to allow for square or rectangular tree spacing. The continuous change in plant density, characteristic of Nelder designs, also forms the basis of several other systematic spacing designs used in intercropping research such as the replacement series design (Radosevich 1987). This latter design closely resembles the rectangular type II design (Figure 2.1e).

The type Ia spacing wheel is an extremely efficient design for studying tree density-growth relationships. When a spacing wheel spans  $360^\circ$  degrees, only the inner and outer arcs (*i.e.*, a total of  $\frac{4\pi}{\omega}$  trees) are lost as guards. Therefore, greater proportions of area and trees are allowed to contribute information to the estimation of treatment effects. Since the design is so compact, it is often possible to replicate it several times within the same area required by an experiment with multiple-tree, rectangular experimental units. The spacing wheel design is also very flexible; designs running through angles less than  $360^\circ$  can be used in various combinations to conform to the shape of the experimental field (*e.g.*, Freeman 1964; Namkoong 1966).

### 2.3 The MKN Experiment

In 1977, the staff of the University of British Columbia Malcolm Knapp Research Forest in Maple Ridge, British Columbia, initiated a Nelder experiment to investigate the effects of spacing in red alder (*Alnus rubra* Bong.) plantations (MKN Experiment). At the time, the prevalence of red alder regeneration on cutovers was sufficient for some to consider managing the species on short rotations (Paltzat 1984). However, up to that point, the species had been widely regarded

as a weed species and very little was known about the effects of initial density on its growth and development.

The MKN experiment was established at an elevation of 200 m, on a site logged the previous year (Paltzat 1984). Red alder seedlings were planted following scarification. The experiment is made up of two 0.35 ha spacing wheels, as shown in Figure 2.4. Both spacing wheels consist of

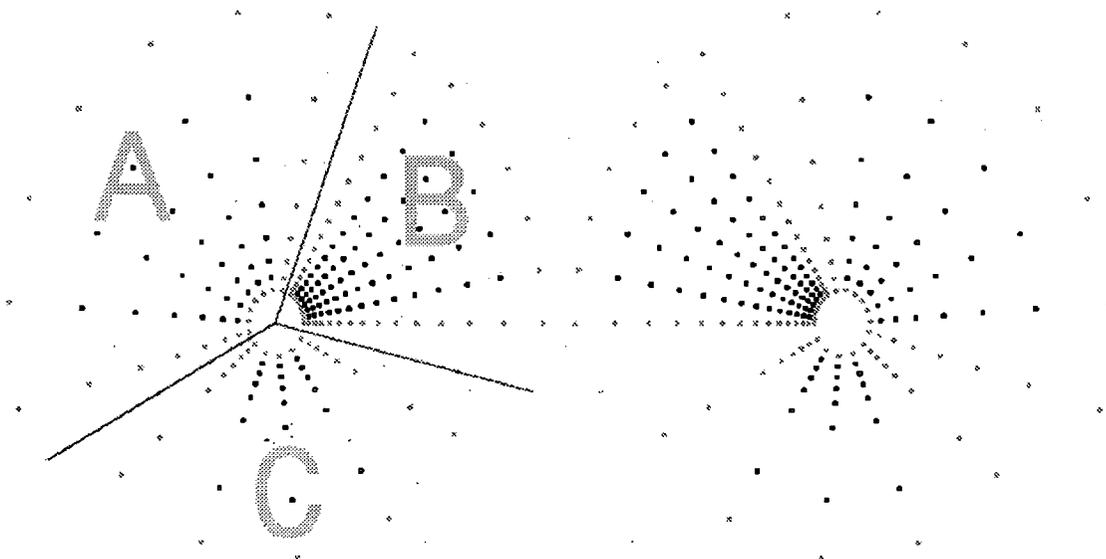


Figure 2.4: Layout of the MKN experiment. Each of the two spacing wheels consist of three sectors (A, B, and C) with different design parameters. Buffer trees are shown in grey.

three sectors, and are oriented so that each wheel is the mirror image of the other. The three sectors are defined by different design parameters and span different tree density levels (Table 2.1). Sectors A and B in Figure 2.4 are both type Ia spacing wheels (Nelder 1962). Sector A was designed with  $r_0 = 3.93$  m,  $\varpi = 19.59^\circ$ , and  $k = 1.3797$ , while sector B was laid out using  $r_0 = 3.93$  m,  $\varpi = 9.795^\circ$ , and  $k = 1.1746$  (Paltzat 1984). Sector C is a modified type Ia design based on a set of design parameters given by Namkoong (1966).

Trees on the innermost and outermost arcs, and on spokes that frame different sectors, are used to buffer the experimental trees and maintain the densities given in Table 2.1. These buffer trees are shown in red in Figure 2.4. In all there are 19 spokes per spacing wheel, 6 of which are for buffer trees, and a total of 113 experimental (*i.e.*, non-buffer) trees grown at 25 tree densities per spacing wheel.

Data were first collected from the MKN experiment in 1983 when the trees were 7 years old.

Table 2.1: Tree density levels in the MKN experiment, by sector (see Figure 2.4).

| Experimental<br>Arc | Tree density (stems/ha) |          |          |
|---------------------|-------------------------|----------|----------|
|                     | Sector A                | Sector B | Sector C |
| 1                   | 3,042                   | 17,001   | 3,089    |
| 2                   | 1,597                   | 12,321   | 2,348    |
| 3                   | 850                     | 8,931    | 1,607    |
| 4                   | 450                     | 6,475    | 875      |
| 5                   | 242                     | 4,693    | 267      |
| 6                   | 131                     | 3,401    | 134      |
| 7                   |                         | 2,464    |          |
| 8                   |                         | 1,787    |          |
| 9                   |                         | 1,295    |          |
| 10                  |                         | 949      |          |
| 11                  |                         | 690      |          |
| 12                  |                         | 504      |          |
| 13                  |                         | 368      |          |

The trees have been re-measured twice since then, first at age 15, and then at age 19. An aerial photograph taken in 1987, when the trees were 11 years old, is shown in Figure 2.5. Mortality has been an issue in this experiment. Paltzat (1984) reported that during the first year, 11% of the trees in one of the spacing wheels, and 61% of the trees in the other, died. The dead seedlings were culled and their positions were replanted in the second year.

## 2.4 Randomization and Statistical Inference

The primary objective of any spacing experiment is to obtain the most accurate estimates of the effects of tree density possible, given the resources at hand. Nelder spacing wheels are very economical designs, but have not found widespread use due, in large part, to historical biases for experimental randomization and design-based statistical inference. The systematic arrangement of treatments makes these designs appealing from a practical standpoint but also renders them invalid as a basis for statistical inference.

Since the Nelder spacing wheel designs do not allow for randomization of treatments to experimental units, these designs do not constitute a basis for assessing the bias or precision of an estimator of the effects of tree density. In any field experiment employing a systematic design, it is possible that a set of experimental units assigned the same treatment may also be exposed to a common level of another uncontrolled factor. As an extreme example, consider a spacing wheel

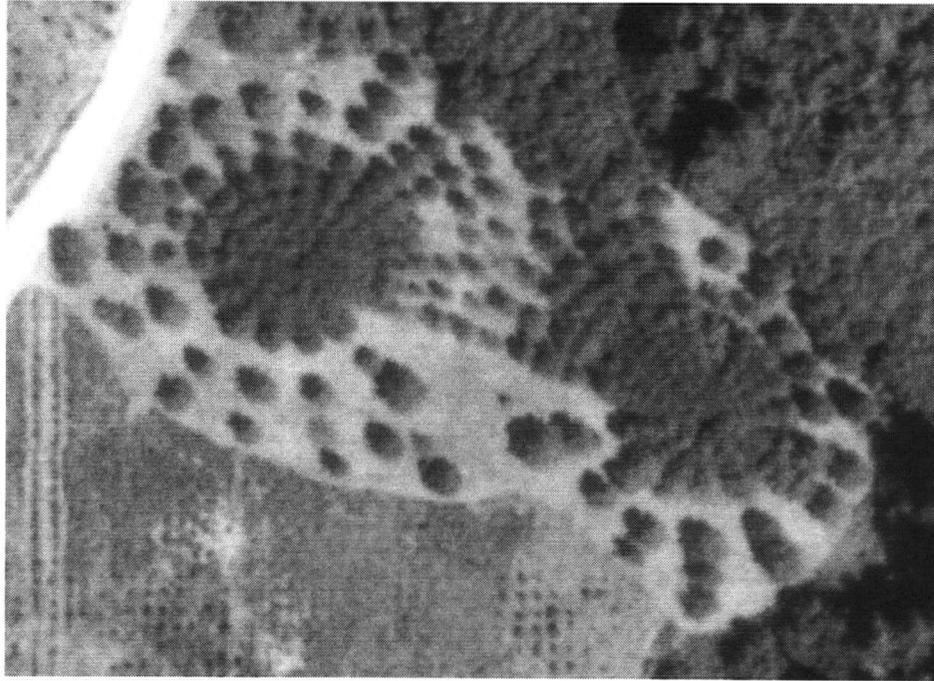


Figure 2.5: Aerial photograph of the MKN experiment taken in 1987, when the trees were 11 years old (scale is approximately 1:1,250).

laid out such that the centre of the wheel coincides with a particularly elevated microsite and the arcs of the wheel follow successive contours in elevation. In this case, the effects of tree density are completely confounded with the effects of any site factors that vary with the microtopography (*e.g.*, soil moisture). From a design-based approach to statistical inference, confounding of treatment and uncontrolled factor effects is a major flaw of systematic experimental designs. The fact that it is unlikely that the concentric patterns of tree density in a spacing wheel are going to overlap with similar expressions of site variables is not important. Nor is it important that particular layouts of randomized experimental designs may lead to a confounding of treatment and other effects; the resultant bias is zero when averaged over the set of all possible randomizations.

A second criticism of the use systematic designs in field experiments stems from the properties of the experimental material. It has often been observed (*e.g.*, Student 1914; Papadakis 1937; Milne 1959; Magnussen 1993) that data from field experiments tend to exhibit systematic spatial heterogeneity such that there can be appreciable correlation among neighbouring observations. This poses a problem for statistical inference. Spatial autocorrelation violates the conditions underlying many of the common normal-theory variance estimators and hypothesis tests. The device of experimental randomization guarantees the existence of a design-unbiased estimate of experimental error and the

validity of inferences on treatment effects. In randomized designs, randomization of treatments to experimental units neutralizes the effects of spatial dependence because it ensures that all spatial arrangements of treatments are equally likely and will occur with the same frequency over the set of all possible randomizations. From the design-based approach to statistical inference, Nelder experiments do not secure a basis for inference. Yet while there is no basis for statistical inference imparted by the design, it is still possible to draw inferences on the effects of tree density in a Nelder experiment on the basis of a statistical model.

Model-based statistical inference relies on an explicit definition of a stochastic framework under which the data arise. Inferences are made on the basis of the model and are subject to the adequacy of the model. Under a model-based approach, randomization renders inference more robust, but is not necessary for valid statistical inference (Harville 1975; Thornett 1982; Hooper 1989). Inference can be based on the statistical model alone, the assumptions of which can be examined independently. For example, Galinski *et al.* (1994) studied tree height differentiation in a Nelder experiment and investigated the tenability of their model through a series of antecedent soil and tree sampling efforts. It is also noteworthy that even in systematic designs, some sources of variation may be controlled through randomization. Chacko (1965) pointed out that variation due to differences in tree growth potential (*i.e.*, genetic differences) in spacing experiments is often of greater magnitude than variation due to differences in site factors. The effects of this source of variation can be controlled in Nelder spacing wheels through the randomization of seedlings to tree density levels.

Much of the controversy over the use of Nelder spacing wheels and other systematic designs in field experiments stems from a misunderstanding of the role of experimental randomization in statistical inference. The literature on the design and analysis of experiments extends back to the early 1900's, yet few authors have articulated, or acknowledged the implications of, their particular preference for design- or model-based inference.<sup>2</sup> In the context of design-based inference, the qualifier "over the set of all possible randomizations" has not always been adequately stressed, nor have all its implications been adequately addressed. As Harville (1975) pointed out, experimental randomization serves as a useful basis for statistical inference only when we ignore the fact that any analysis is conditional upon the particular randomization observed. We must also ignore the fact that the properties of the estimators derived from viewing the observed layout as one amongst an infinite set of layouts do not apply to the estimates based on the single layout observed. The design-based approach to statistical inference is associated more with an emphasis on significance

---

<sup>2</sup>These topics have, however, received more attention in the sampling literature (*e.g.*, Hansen *et al.* 1983).

testing, and, as Student (1937) pointed out, does not generally lead to the most accurate estimates of treatment effects. Model-based inference can be a pragmatic perspective for the analysis of field experiments, and systematic experimental designs, such as Nelder spacing wheels can provide tremendous material efficiencies. Combined with a suitable analysis method, these two approaches to experimental inference may provide more accurate estimates of treatment effects.

## Chapter 3

# Analysis Methods for Nelder Experiments

### 3.1 Analysis Methods

The interpretation of experimental data from a model-based approach to statistical inference proceeds from the identification and application of a suitable analysis method. An analysis method is here defined as consisting of a parametric stochastic model and a set of estimators for the model parameters. In this section, a general linear model for Nelder experiments is developed and the statistical properties of an analysis method are defined.

#### 3.1.1 The General Linear Model

Consider a Nelder experiment consisting of  $b$  spacing wheels,  $t$  levels of tree density, and  $m$  trees per spacing wheel per density level. The observations on the variable of interest can then be modelled as a realization of the  $n \times 1$  random vector

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

where  $n = btm$  is the total number of observations. In addition, the relative spatial positions of the trees, denoted by the  $n \times 2$  matrix  $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}^T$  of bivariate spatial co-ordinates (*i.e.*,  $\mathbf{s}_i \in \mathbb{R}^2$ ;  $i = 1, 2, \dots, n$ ), can be derived from the formal geometric structures of the Nelder spacing wheels. Thus, the vector of observations, indexed by their relative spatial positions can be given

by the realization  $\mathbf{y}_S = \{y_{s_1}, y_{s_2}, \dots, y_{s_n}\}^T$  of the random vector  $\mathbf{Y}_S$ .

The general model developed below is based on the assumption that the elements of  $\mathbf{Y}_S$  can be additively decomposed as

$$Y_{s_i} = \mu + \sum_{j=1}^b f(i, j)\alpha_j + \sum_{k=1}^t f(i, k)\tau_k + \sum_{j=1}^b \sum_{k=1}^t f(i, j)f(i, k)(\alpha \times \tau)_{jk} + \varepsilon_{s_i} \quad (3.1)$$

where  $\mu$  is the overall mean,  $\alpha_j$  is the effect of the  $j$ th spacing wheel,  $\tau_k$  is the effect of the  $k$ th level of tree density,  $(\alpha \times \tau)_{jk}$  is the interaction effect for the  $j$ th spacing wheel and the  $k$ th level of tree density,  $\varepsilon_{s_i}$  is an unobservable zero-mean random variable (indexed by its spatial position) associated with the observation taken on the  $i$ th tree, and

$$f(i, q) = \begin{cases} 1 & \text{if the } i\text{th observation is assigned to the } q\text{th factor level} \\ 0 & \text{otherwise.} \end{cases}$$

Removing the linear dependencies among the mean structure parameters, Equation 3.1 can be stated more concisely as

$$Y_{s_i} = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_{s_i} \quad (3.2)$$

where  $\mathbf{x}_i$  is a  $1 \times p$  design vector relating observations to the main effects and their interactions,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of linearly independent parameters (*i.e.*,  $p = bt$ ), and  $\varepsilon_{s_i}$  is as defined above. The tree density effects in Equation 3.1 are then obtained through a linear transformation of the elements of  $\boldsymbol{\beta}$ , that is

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_t \end{bmatrix} = \boldsymbol{\lambda}_t^T \boldsymbol{\beta} \quad (3.3)$$

where  $\boldsymbol{\lambda}_t$  is a  $p \times t$  matrix of ones and zeros that defines the appropriate transformation of  $\boldsymbol{\beta}$ .

Estimation of  $\boldsymbol{\tau}$  (treatment effects) and of the differences among the elements of  $\boldsymbol{\tau}$  (treatment contrasts) are of central importance. In Equations 3.1 and 3.2, tree density is implicitly defined as a class variable. Researchers studying the effects of tree spacing are generally interested in modelling the effects of tree density over a range of tree densities, not only at the  $t$  levels used in an experiment. However, the effects of tree density are likely to be different for different variables of interest, tree species, and site qualities. Since the emphasis in this study is on comparing different estimators of tree density effects, the discontinuous form in Equation 3.2 is used. This avoids complicating these comparisons with possible effects of mis-specifying the density-growth response. Notwithstanding, the results of this study could be extended to the estimation of more complex density-growth

response functions. Continuous and/or non-linear functions could be fitted by generalizing  $\mathbf{x}_i\boldsymbol{\beta}$  to a more flexible mean structure (e.g.,  $f(\mathbf{x}_i; \boldsymbol{\beta})$ ).

Spacing wheel effects and interactions among spacing wheels and tree densities can also be obtained from a transformation of  $\boldsymbol{\beta}$ . These factors could be modelled as either random or fixed effects. However, the focus of this study is the estimation of the effects of individual levels of tree density (or of contrasts among them) not the estimation of the expected growth response. As a result, variation due to differences among spacing wheels, and variation due to interactions among spacing wheels and tree density levels, can be treated as fixed effects without loss of generality (Casella and Berger 1990).

Throughout this thesis, it was assumed that the joint probability density function (pdf) of the random errors ( $\varepsilon_{s_i}$ ) in Equation 3.2 could be approximated by a multivariate normal (i.e., Gaussian) distribution. The Central Limit Theorem (CLT) states that the sum of a sequence of independent random variables converges in distribution to a normal random variable (Chung 1974).<sup>1</sup> The random errors in Equation 3.2 consist of the effects of many sources of variation (e.g., measurement error, differences among trees used in the experiment, variability in soil properties within spacing wheels, etc.). If the effects of these sources of variation can be taken to be additive, the CLT can be relied upon to provide a measure of assurance that their cumulative effects will be normally distributed. A multivariate normal distribution may not be suitable when the tree variable of interest is discrete (e.g., tree vitality) or has a censored domain (e.g., diameter increment cannot be negative) but may be a reasonable approximation for many continuous tree growth attributes.

At this point, Equation 3.2 can be restated in matrix notation as

$$\mathbf{Y}_S = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_S \quad (3.4)$$

where  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}^T$  is a  $n \times p$  matrix of full column rank, and  $\boldsymbol{\varepsilon}_S = \{\varepsilon_{s_1}, \varepsilon_{s_2}, \dots, \varepsilon_{s_n}\}^T$  is a  $n \times 1$  vector of random errors. In addition,  $\boldsymbol{\varepsilon}_S \sim N(\mathbf{0}, \mathbf{V})$  where  $\mathbf{0}$  is an  $n \times 1$  vector of zeros, and  $\mathbf{V}$  is an  $n \times n$  positive definite matrix. Finally, let  $\mathbf{V} = \sigma^2\boldsymbol{\Omega}$  where  $\sigma^2$  is a general scale parameter ( $\sigma^2 > 0$ ), and  $\boldsymbol{\Omega}$  is an  $n \times n$  positive definite matrix. This implies that  $\mathbf{Y}_S \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega})$ .

### 3.1.2 Statistical Properties

Each analysis method discussed in the following sections is based on the general linear model given by Equation 3.4. Evaluation of these analysis methods will be made on the basis of both practical and statistical considerations. Practical considerations including parsimony, biological

<sup>1</sup>Chung also details several formulations of the CLT in which the condition of independence is relaxed.

relevance, and feasibility will be discussed in Chapter 5. Statistical considerations will be discussed here.

The statistical properties of an analysis method considered here are bias, efficiency, and validity. An analysis method is unbiased if the estimators of treatment effects are unbiased, that is

$$E(\hat{\tau}) = E(\lambda_t^T \hat{\beta}) = \tau \quad (3.5)$$

where  $\hat{\tau} = \lambda_t^T \hat{\beta}$  is an estimator of  $\tau$ ,  $\hat{\beta}$  is an estimator of  $\beta$ ,  $\lambda_t$  is the appropriate linear transformation of  $\hat{\beta}$ , and the expectation is taken over the pdf defined by the model. An analysis method is efficient if the estimators of treatment effects are efficient, that is

$$\text{Var}(\hat{\tau}) = \text{Var}(\hat{\tau}_o) - \Delta \quad (3.6)$$

where  $\hat{\tau}_o$  is any other unbiased estimator of  $\tau$  and  $\Delta$  is a  $t \times t$  positive semidefinite matrix. In other words, the estimators of treatment effects should be as precise as possible. Finally, if an analysis method is unbiased and

$$E \left[ \sum_{k=1}^t \widehat{\text{Var}}(\hat{\tau}_k) \right] = \sum_{k=1}^t \text{Var}(\hat{\tau}_k) \quad (3.7)$$

where  $\widehat{\text{Var}}(\hat{\tau}_k)$  is an estimator of  $\text{Var}(\hat{\tau}_k)$  ( $k = 1, 2, \dots, t$ ), then it is valid (Besag and Kempton 1986).

Validity is an important property of an analysis method because, in practice, analysis methods are often selected on the basis of the estimated precision of  $\hat{\tau}$ . If these estimates are biased, the comparison is meaningless. Lack of validity is also dangerous, because it can vitiate the results of hypothesis tests. For example, a testable hypothesis in the context of the general linear model developed above is the absence of a treatment effect (*i.e.*,  $\mathcal{H}_0 : \tau = \mathbf{0}$ ). Assuming all other factors in the model (*e.g.*, spacing wheels) are considered fixed, this hypothesis can be tested using an  $F$ -statistic of the form

$$F_{(t-1), (n-p)} = \frac{MS_{TRT}}{MS_{ERR}} \quad (3.8)$$

where  $MS_{TRT}$  is the mean squared treatment effect;  $MS_{ERR}$  is the mean squared error; and  $(t-1)$  and  $(n-p)$  are the treatment and error degrees of freedom, respectively. In matrix notation, Equation 3.8 can be stated as

$$F_{(t-1), (n-p)} = \frac{\hat{\tau}^T \mathbf{L}^T \left( \mathbf{L} \lambda_t^T (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \lambda_t \mathbf{L}^T \right)^{-1} \mathbf{L} \hat{\tau}}{t-1} \quad (3.9)$$

where  $\hat{\mathbf{V}}$  is an estimator of  $\mathbf{V}$ , and  $\mathbf{L}$  is a  $(t-1) \times t$  contrast matrix

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \quad (3.10)$$

(Beckers 1997). In a randomization context, validity is defined as the equality of the expectations of the treatment and error mean squares in Equation 3.8 under  $\mathcal{H}_0$ , which implies Equation 3.7 (Azaïs *et al.* 1998). In the context of the general linear model given by Equation 3.4, the equality of the expectations of the treatment and error mean squares also implies that the  $F$ -statistic in Equation 3.9 will be distributed according to a Fisher (F) distribution with  $t-1$  and  $n-p$  degrees of freedom. If an analysis method is not valid, it is not legitimate to test  $\mathcal{H}_0$  using an  $F$ -statistic.

## 3.2 Classical Analysis Methods

The classical analysis method, ANOVA/OLS,<sup>2</sup> for field experiments is characterized by the assumption that the errors in Equation 3.4 are normal, independent, and identically distributed (i.i.d.) random variables. Estimation of model parameters proceeds using the ordinary least squares (OLS) estimators. If the assumption of normal i.i.d. random errors is justified, the OLS estimators of treatment effects are unbiased and efficient, and unbiased estimates of their variances are readily available.

### 3.2.1 The ANOVA Model

The analysis model for ANOVA/OLS is a special case of Equation 3.4. It can be stated as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.11)$$

where  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . The ANOVA model does not incorporate the information in  $\mathbf{S}$  on the spatial arrangement of observations, so the subscript is dropped from  $\mathbf{Y}$  and  $\boldsymbol{\varepsilon}$ . Also, the matrix  $\boldsymbol{\Omega}$  is reduced to  $\mathbf{I}_n$ , a  $n \times n$  identity matrix. This implies that

$$\text{Cov}(Y_i, Y_l) = \text{Cov}(\varepsilon_i, \varepsilon_l) = \begin{cases} \sigma^2 & \text{if } i = l \\ 0 & \text{otherwise} \end{cases}$$

for  $i, l = 1, 2, \dots, n$ , and that  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ .

<sup>2</sup>The naming convention for analysis methods is [*analysis model*]/[*estimation procedure*].

### 3.2.2 OLS Estimation

The simple form of the variance matrix of the random errors,  $\epsilon$ , assumed in the ANOVA model establishes a set of conditions under which the OLS estimators of  $\tau$  are unbiased and efficient. The OLS estimator of  $\tau$  is given by

$$\hat{\tau}_{ols} = \lambda_t^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (3.12)$$

Equation 3.12 defines a linear combination of  $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$ , so it follows that  $\hat{\tau}_{ols}$  also has a multivariate normal distribution. Its mean and variance matrix are given by

$$\mathbf{E}(\hat{\tau}_{ols}) = \lambda_t^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E}(\mathbf{Y}) = \lambda_t^T \beta = \tau \quad (3.13)$$

and

$$\begin{aligned} \text{Var}(\hat{\tau}_{ols}) &= \mathbf{E}(\hat{\tau}_{ols} - \tau)(\hat{\tau}_{ols} - \tau)^T \\ &= \mathbf{E} \left[ \lambda_t^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon \epsilon^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \lambda_t \right] \\ &= \sigma^2 \lambda_t^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda_t \end{aligned} \quad (3.14)$$

Equation 3.13 states that  $\hat{\tau}_{ols}$  is an unbiased estimator of  $\tau$ . It can also be shown that  $\text{Var}(\hat{\tau}_{ols})$  achieves the Cramér-Rao lower bound (Judge *et al.* 1985). Therefore,  $\hat{\tau}_{ols}$  is an efficient estimator of  $\tau$ . In addition, the mean squared error

$$\hat{\sigma}_{ols}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta}_{ols})^T (\mathbf{y} - \mathbf{X}\hat{\beta}_{ols})}{n - p} = \frac{\mathbf{e}_{ols}^T \mathbf{e}_{ols}}{n - p} \quad (3.15)$$

where  $\hat{\beta}_{ols}$  is the OLS estimator of  $\beta$  and  $\mathbf{e}_{ols}$  is the  $n \times 1$  vector of ANOVA/OLS residuals, provides an unbiased estimator of  $\sigma^2$  (Judge *et al.* 1985). Substituting  $\hat{\sigma}_{ols}^2$  for  $\sigma^2$  in Equation 3.14

$$\widehat{\text{Var}}(\hat{\tau}_{ols}) = \hat{\sigma}_{ols}^2 \lambda_t^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda_t \quad (3.16)$$

provides an unbiased estimator of  $\text{Var}(\hat{\tau}_{ols})$ .

Under the assumption of multivariate normal i.i.d. random errors, ANOVA/OLS is unbiased, efficient, and valid. Under the null hypothesis (*i.e.*,  $\tau = \mathbf{0}$ ), the treatment  $F$ -statistic

$$F_{(t-1), (n-p)} = \frac{\hat{\tau}_{ols}^T \mathbf{L}^T \left( \mathbf{L} \lambda_t^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda_t \mathbf{L}^T \right)^{-1} \mathbf{L} \hat{\tau}_{ols}}{\hat{\sigma}_{ols}^2 (t-1)} \quad (3.17)$$

is distributed according to an F distribution with  $t - 1$  and  $n - p$  degrees of freedom. These are very powerful results.

### 3.2.3 Robustness

One drawback of the classical analysis method is that the OLS estimators lose their well-defined and attractive properties when the condition of normal i.i.d. errors cannot be justified. As discussed in Section 2.4, data from field experiments can exhibit spatial dependence, such that appreciable correlation can exist among observations located close together. In such cases, the form for  $\mathbf{V}$  taken in Equation 3.11 is oversimplified (*i.e.*,  $\mathbf{V} = \sigma^2\boldsymbol{\Omega} \neq \sigma^2\mathbf{I}_n$ ).

When the random errors are not independent,  $\hat{\tau}_{ols}$  is normally distributed and unbiased but has variance

$$\begin{aligned}\text{Var}(\hat{\tau}_{ols}) &= \text{E}(\hat{\tau}_{ols} - \tau)(\hat{\tau}_{ols} - \tau)^T \\ &= \text{E}\left[\boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}_t\right] \\ &= \sigma^2 \boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}_t .\end{aligned}\quad (3.18)$$

Using Equation 3.18 it can be shown that the OLS estimators are not efficient under spatial autocorrelation. The generalised least squares (GLS) estimator of  $\tau$  given by

$$\hat{\tau}_{gls} = \boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \quad (3.19)$$

is normally distributed (under the assumption of multivariate normality for  $\boldsymbol{\varepsilon}$ ) with mean and variance

$$\text{E}(\hat{\tau}_{gls}) = \boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \text{E}(\mathbf{Y}) = \tau \quad (3.20)$$

and

$$\begin{aligned}\text{Var}(\hat{\tau}_{gls}) &= \text{E}(\hat{\tau}_{gls} - \tau)(\hat{\tau}_{gls} - \tau)^T \\ &= \text{E}\left[\boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \boldsymbol{\lambda}_t\right] \\ &= \boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \text{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \boldsymbol{\lambda}_t \\ &= \boldsymbol{\lambda}_t^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \boldsymbol{\lambda}_t \\ &= \sigma^2 \boldsymbol{\lambda}_t^T (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \boldsymbol{\lambda}_t .\end{aligned}\quad (3.21)$$

From Equation 3.20,  $\hat{\tau}_{gls}$  is an unbiased estimator of  $\tau$ . More important, however, is that  $\text{Var}(\hat{\tau}_{ols}) - \text{Var}(\hat{\tau}_{gls}) = \boldsymbol{\Delta}$  is a positive semidefnite matrix (Judge *et al.* 1985). As a result, when the error terms have a non-diagonal variance matrix, the OLS estimator of  $\tau$  is, in general, not efficient.

In addition, when  $\Omega \neq \mathbf{I}_n$  the OLS estimator of  $\sigma^2$  in Equation 3.15 is biased

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}_{ols}^2) &= \frac{1}{n-p} \mathbb{E}(\mathbf{e}_{ols}^T \mathbf{e}_{ols}) \\
&= \frac{1}{n-p} \mathbb{E} \left[ \boldsymbol{\varepsilon}^T \left( \mathbf{I}_n - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \boldsymbol{\varepsilon} \right] \\
&= \frac{1}{n-p} \mathbb{E} \left[ \text{tr} \left( \boldsymbol{\varepsilon}^T \left( \mathbf{I}_n - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \boldsymbol{\varepsilon} \right) \right] \\
&= \frac{1}{n-p} \mathbb{E} \left[ \text{tr} \left( \left( \mathbf{I}_n - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \right) \right] \\
&= \frac{1}{n-p} \text{tr} \left[ \left( \mathbf{I}_n - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{V} \right] \\
&= \frac{\sigma^2}{n-p} \text{tr} \left[ \left( \mathbf{I}_n - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \Omega \right] \tag{3.22}
\end{aligned}$$

where  $tr(\cdot)$  is the trace operator. The resulting magnitude and sign of the bias in  $\hat{\sigma}_{ols}^2$  will depend on the structure of the design and correlation matrices (Judge *et al.* 1985). But as a result, the treatment  $F$ -statistic in Equation 3.17 will also be biased.

In summary, when the multivariate normal i.i.d. condition of the ANOVA model is justified, ANOVA/OLS is unbiased, efficient, and valid. However, if the errors are autocorrelated, the OLS estimators of  $\boldsymbol{\tau}$  are not efficient and the variance estimators are biased. ANOVA/OLS will still be unbiased but, in general, will be neither efficient nor valid.

### 3.3 Spatial Analysis Methods

The gains in efficiency of treatment estimation provided by GLS estimators are of limited value since  $\mathbf{V}$  is usually unknown. In addition, because there are more elements in  $\mathbf{V}$  than there are observations, it is impossible to estimate  $\mathbf{V}$  directly. The basis of the spatial approach is to reduce the dimensions of  $\mathbf{V}$  to an estimable number of covariance parameters by making assumptions about the correlation structure (*i.e.*,  $\Omega$ ). The resultant models are generally nonlinear with respect to the covariance parameters and require likelihood-based estimation.

#### 3.3.1 Spatial Correlated Error Models

Spatial CE models are derived from the principle known as Tobler's Law (*e.g.*, Anle and Marshall 1996), that "everything is related to everything else, but near things are more related than distant things" (quoted in Gould 1970). The spatial CE models considered in this thesis are based on the parameterization  $\mathbf{V} = \sigma^2 \Omega$  where  $\Omega \equiv \Omega(\mathbf{S}; \boldsymbol{\theta})$  is a matrix-valued function of a vector of unknown parameters  $\boldsymbol{\theta}$  and the matrix of spatial coordinates  $\mathbf{S}$ . In these models, the off-

diagonal elements of  $\Omega$  (*i.e.*, the set of pairwise correlations among observations) are modelled as a function of the spatial distribution of the data. As in many time-series models, the autoregressive formulation of autocorrelation plays a central role. However, distinct autoregressive formulations can be used in a spatial setting and spatial autocorrelation can be modelled as either a continuous or discrete process.

In a temporal setting, the autoregressive model of autocorrelation can be specified in any of three equivalent ways. Denoting the observation to be taken at time  $q$  ( $q = \dots, -1, 0, 1, \dots$ ) by  $Y_q$ , the first-order autoregressive model can be specified:

1. by defining  $Y_q$  as a function of  $Y_{q-1}$

$$Y_q = \theta Y_{q-1} + \varepsilon_q \quad (3.23)$$

where  $E(Y_q) = 0$  and  $\varepsilon_q \sim N(0, \sigma^2)$  is a random error uncorrelated with  $Y_{q-a}$  and  $\varepsilon_{q-a}$  for  $a > 0$ ;

2. through the conditional distribution of  $Y_q$  given  $\mathbf{Y}_q^* = \{Y_i : i \neq q\}$

$$E(Y_q | \mathbf{Y}_q^*) = \theta y_{q-1} \quad (3.24)$$

$$\text{Var}(Y_q | \mathbf{Y}_q^*) = \sigma^2 \quad (3.25)$$

where  $y_{q-1}$  is a realization of  $Y_{q-1}$ ; or

3. through the unconditional distribution of  $Y_q$

$$E(Y_q) = 0 \quad (3.26)$$

$$\text{Cov}(Y_q, Y_{q-a}) = \frac{\sigma^2}{1 - \theta^2} \theta^{|a|} \quad (3.27)$$

(Cliff and Ord 1981). These definitions are referred to as the simultaneous, conditional, and unconditional formulations, respectively. As Cressie (1993) pointed out, the equivalence of these three formulations of the temporal autoregressive model relies on the unidirectional nature of dependence through time. In a spatial setting there is no past, present, or future - spatial dependence extends in all directions simultaneously. As a result, spatial CE models that define spatial autocorrelation according to simultaneous, conditional, or unconditional formulations are distinct.

A second distinction among spatial CE models can be made on the basis of the spatial extent of the autocorrelation. The collection of responses,  $\mathbf{y}_s$ , from a Nelder experiment can be considered as a realization of a spatial stochastic process

$$\mathbf{Y}_s \equiv \{Y_{s_i} = Y(\mathbf{s}_i) : \mathbf{s}_i \in D \subset \mathcal{R}^2, i = 1, 2, \dots, n\} \quad (3.28)$$

where  $\mathbf{s}_i$  ranges over a 2-dimensional index set  $D$ . A dichotomy between spatial CE models can be drawn on the basis of whether the index set,  $D$ , is discrete or continuous.

One class of spatial CE models that have been widely applied in the agricultural and social sciences are the nearest neighbour (NN) models. In NN models, the  $n$  rows of  $\mathbf{S}$  (the locations of the  $n$  trees) define the extent of the stochastic process  $\mathbf{Y}_{\mathbf{S}}$  (*i.e.*,  $D = \mathbf{S}$ ). Under this framework, spatial autocorrelation is defined using spatial analogues of the simultaneous and conditional formulations of temporal autoregression. Consequently, analyses based on NN models account for spatial autocorrelation through differencing or using responses from neighbouring observations as covariates.

A second class of spatial CE models applicable to the analysis of field experiments originated from geological spatial interpolation methods developed in the 1940s (Cressie 1990). Regionalized variables (RV) models have been widely used in geological and mining applications but were only introduced as potential models for the analysis of field experiments by Zimmerman and Harville in 1991. These models characterize spatial autocorrelation as a continuous process (*i.e.*,  $D \subset \mathcal{R}^2$ ) and incorporate spatial autocorrelation through explicit specification of the covariance structure. This approach is analogous to the unconditional formulation of the temporal autoregressive model.

Before examining NN and RV models in detail, it is important to note some of the assumptions involved in modelling spatial autocorrelation. In general, autocorrelation among spatial data poses a major problem for statistical inference. Unless assumptions are made about the joint pdf of the spatial process,  $\mathbf{y}_{\mathbf{S}}$  can only be viewed as a single (and possibly incomplete) realization of the stochastic process  $\mathbf{Y}_{\mathbf{S}}$ . To avoid this interpretation,  $\mathbf{Y}_{\mathbf{S}}$  is generally modelled as spatially stationary stochastic process. A stochastic process is spatially stationary if the joint pdf of  $\mathbf{Y}_{\mathbf{S}}$  depends on  $\mathbf{S}$  only through the relative spatial arrangement of the observations (Wackernagel 1998). Under the assumption of multivariate normality, this is equivalent to the conditions of covariance (or second order) stationarity

$$E(\mathbf{Y}_{\mathbf{S}}) = \mathbf{X}\boldsymbol{\beta} \quad (3.29)$$

and

$$\text{Cov}(Y_{\mathbf{s}_i}, Y_{\mathbf{s}_l}) = \sigma^2 g(\mathbf{h}_{il} | \boldsymbol{\theta}) \quad (3.30)$$

where the parameters  $\boldsymbol{\beta}$  do not depend on  $\mathbf{S}$ , and  $g(\cdot)$  is a function of  $\boldsymbol{\theta}$  and  $\mathbf{h}_{il} = (\mathbf{s}_i - \mathbf{s}_l)$ , for all  $\mathbf{s}_i$  and  $\mathbf{s}_l$  in  $D$  ( $i, l = 1, 2, \dots, n$ ). Under the condition of stationarity, the pdf of a spatial stochastic process depends only on the relative spatial distribution of the observations, not on their absolute positions. In the absence of fixed large-scale spatial trends, covariance stationarity is generally a

reasonable assumption.

A second condition that is often imposed on the pdf of spatially autocorrelated processes is isotropy. A stationary process  $\mathbf{Y}_S$  is isotropic if its joint pdf depends on  $\mathbf{S}$  only through

$$d_{il} = \|\mathbf{h}_{il}\| = \sqrt{\mathbf{h}_{il}^T \mathbf{h}_{il}} \quad (3.31)$$

for all  $\mathbf{s}_i$  and  $\mathbf{s}_l$  in  $D$ . Under covariance stationarity, this implies that

$$\text{Cov}(Y_{\mathbf{s}_i}, Y_{\mathbf{s}_l}) = \sigma^2 f(d_{il} | \boldsymbol{\theta}) \quad (3.32)$$

for all  $\mathbf{s}_i$  and  $\mathbf{s}_l$  in  $D$ . Hence the correlation between any two observations depends on their spatial positions only through their separation distance and is direction invariant. Isotropy is generally adopted as a simplifying assumption and is not a necessary condition for statistical inference. Anisotropy (*i.e.*, pairwise correlations are a function of the directional distances between observations) may be more suitable when, for example, the observations are located on a topographic gradient. In such cases, there may be much stronger pairwise correlations extending downslope than across the slope.

### Nearest Neighbour Models

In NN models, spatial autocorrelation is accounted for indirectly through simultaneous or conditional autoregressive formulations. Central to the use of NN models are the definitions of neighbourhoods and spatial weight matrices.

The simultaneous spatial autoregressive (SAR) model was introduced by Whittle (1954). The error correlation structure in the SAR model is given by

$$\begin{aligned} \boldsymbol{\varepsilon}_S &= \boldsymbol{\theta} \mathbf{W} \boldsymbol{\varepsilon}_S + \boldsymbol{\eta} \\ &= (\mathbf{I}_n - \boldsymbol{\theta} \mathbf{W})^{-1} \boldsymbol{\eta} \end{aligned} \quad (3.33)$$

where  $\boldsymbol{\theta}$  is an unknown spatial autocorrelation parameter,  $\boldsymbol{\eta} \sim N(0, \sigma^2 \mathbf{I}_n)$  is a  $n \times 1$  random vector, and  $\mathbf{W} = \{w_{il}\}$  is a  $n \times n$  spatial weight matrix with elements

$$w_{il} = \begin{cases} 1/d_{il} & \text{if } \varepsilon_{\mathbf{s}_l} \text{ is in the neighbourhood of } \varepsilon_{\mathbf{s}_i} \\ 0 & \text{otherwise.} \end{cases} \quad (3.34)$$

For a Nelder experiment, it is useful to define the neighbourhood of a tree as the set of nearest inter-spoke and intra-spoke trees. The spatial weight matrix in Equation 3.34 then reflects the

assymetrical distribution of these neighbours by weighting their influence as a diminishing function of their distance from tree  $i$ . Combining Equations 3.4 and 3.33, the SAR model can be stated as

$$\mathbf{Y}_S = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \theta\mathbf{W})^{-1}\boldsymbol{\eta} \quad (3.35)$$

The mean and variance of  $\mathbf{Y}_S$  are

$$\mathbf{E}(\mathbf{Y}_S) = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \theta\mathbf{W})^{-1}\mathbf{E}(\boldsymbol{\eta}) = \mathbf{X}\boldsymbol{\beta} \quad (3.36)$$

and

$$\begin{aligned} \text{Var}(\mathbf{Y}_S) &= \mathbf{E}(\mathbf{Y}_S - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y}_S - \mathbf{X}\boldsymbol{\beta})^T \\ &= \mathbf{E}(\mathbf{I}_n - \theta\mathbf{W})^{-1}\boldsymbol{\eta}\boldsymbol{\eta}^T(\mathbf{I}_n - \theta\mathbf{W}^T)^{-1} \\ &= \sigma^2(\mathbf{I}_n - \theta\mathbf{W})^{-1}(\mathbf{I}_n - \theta\mathbf{W}^T)^{-1} \\ &= \sigma^2(\mathbf{I}_n - 2\theta\mathbf{W} + \theta^2\mathbf{W}\mathbf{W})^{-1} \end{aligned} \quad (3.37)$$

Under the assumption of multivariate normality, the SAR model implies that

$$\mathbf{Y}_S \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n - 2\theta\mathbf{W} + \theta^2\mathbf{W}\mathbf{W})^{-1}) \quad (3.38)$$

A disadvantage of the SAR model is that the autoregressive process,  $\mathbf{Y}_S$ , is not independent of the underlying i.i.d. random process,  $\boldsymbol{\eta}$ . That is

$$\text{Cov}(\mathbf{Y}_S, \boldsymbol{\eta}) = \mathbf{E}((\mathbf{I}_n - \theta\mathbf{W})^{-1}\boldsymbol{\eta}\boldsymbol{\eta}^T) = \sigma^2(\mathbf{I}_n - \theta\mathbf{W})^{-1} \quad (3.39)$$

is not a diagonal matrix. Consequently, application of OLS or iterative least-squares methods (*e.g.*, Mora 1996) will not result in consistent estimators of  $\theta$  (Anselin 1988). In addition, this result has prompted some authors (*e.g.*, Cressie 1993) to suggest that the conditional spatial autoregressive (CAR) model (Besag 1974) is a more suitable model for spatial autocorrelation.

The CAR model can be specified through the conditional means and variances of the elements of  $\mathbf{Y}_S$

$$\mathbf{E}(Y_{s_i} | \mathbf{Y}_{s_i}^*) = \mathbf{x}_i\boldsymbol{\beta} + \theta \sum_{l=1}^n w_{il}(y_{s_l} - \mathbf{x}_l\boldsymbol{\beta}) \quad (3.40)$$

and

$$\text{Var}(Y_{s_i} | \mathbf{Y}_{s_i}^*) = \sigma^2 \quad (3.41)$$

where  $\mathbf{Y}_{s_i}^* = \{Y_{s_l} : l \neq i\}$ ,  $\sigma^2$  is the conditional variance of  $Y_{s_i}$ ,  $\mathbf{W} = \{w_{ij}\}$  is a spatial weight matrix, and  $\theta$  is an unknown autocorrelation parameter. Besag (1974) showed that under the assumption of multivariate normality the CAR model can be expressed unconditionally as

$$\mathbf{Y}_S \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n - \theta\mathbf{W})^{-1}) \quad (3.42)$$

The spatial weight matrix used for the SAR model (Equation 3.34) can also be used for the CAR model.

For Equations 3.38 and 3.42 to define valid pdfs, the variance matrices

$$\mathbf{V} = \sigma^2(\mathbf{I}_n - 2\theta\mathbf{W} + \theta^2\mathbf{W}\mathbf{W})^{-1} \quad (3.43)$$

and

$$\mathbf{V} = \sigma^2(\mathbf{I}_n - \theta\mathbf{W})^{-1} \quad (3.44)$$

must be positive definite. For the CAR model this means that  $0 \leq \theta \leq 1/\omega_{(1)}$  where  $\omega_{(1)}$  is the largest eigenvalue of  $\mathbf{W}$  (Wall 2000). For the SAR model, this requirement can be met by the less restrictive condition  $\theta \neq 1/\omega_{(i)}$  where  $\omega_{(i)}$  is the  $i$ th largest eigenvalue of  $\mathbf{W}$  ( $i = 1, 2, \dots, n$ ). However, in order to maintain the interpretation of  $\theta$  as a spatial autocorrelation parameter, the restriction  $0 \leq \theta \leq 1/\omega_{(1)}$  should also be applied in the SAR model (Wall 2000).

A drawback of the NN class of models is that spatial autocorrelation is modelled indirectly. Examination of Equations 3.38 and 3.42 reveals that the autocorrelation structure is modelled through the inverse of the covariance matrix. As Wall (2000) pointed out, this means that the actual spatial autocorrelation structure cannot be understood simply by examining  $\mathbf{W}$ . In addition, except on infinite regular lattices, SAR and CAR models are not stationary, and the diagonal elements of  $\mathbf{\Omega} = (\mathbf{I}_n - 2\theta\mathbf{W} + \theta^2\mathbf{W}\mathbf{W})^{-1}$  and  $\mathbf{\Omega} = (\mathbf{I}_n - \theta\mathbf{W})^{-1}$  are not all equal to 1. Most disturbingly, Wall showed that the actual pairwise correlations implied by modeling the inverse covariance matrices can be contrary to what is intended: pairwise correlations between non-neighbours can exceed those between neighbours.

## Regionalized Variables Models

The regionalized variables approach to spatial modelling was developed primarily for spatial interpolation and spatial prediction applications. All RV models are based on unconditional autoregressive formulations and direct modelling of the variance matrix.

An RV model requires an explicit definition of the joint pdf of  $\mathbf{Y}_S$ . The elements of  $\mathbf{\Omega}$  are modeled directly using a covariance function

$$C(\mathbf{h}_{il}; \sigma^2, \boldsymbol{\theta}) = \text{Cov}(\varepsilon_{s_i}, \varepsilon_{s_l}) \quad (3.45)$$

where  $C(\cdot)$  is a real-valued function of  $\mathbf{h}_{il} = (\mathbf{s}_i - \mathbf{s}_l)$ , the scale parameter ( $\sigma^2$ ), and a vector of covariance parameters  $\boldsymbol{\theta}$ . The covariance function should be continuous, non-negative, and

monotone decreasing. Wackernagel (1998) provided an extensive list of permissible covariance functions. Two of the simplest, and most commonly used, isotropic covariance functions are:

1. the exponential covariance function

$$C(d_{il}) = \begin{cases} \sigma^2 \theta^{d_{il}} & \text{for } d_{il} > 0 \\ 0 & \text{otherwise;} \end{cases} \quad (3.46)$$

2. the Gaussian covariance function

$$C(d_{il}) = \begin{cases} \sigma^2 \theta^{d_{il}^2} & \text{for } d_{il} > 0 \\ 0 & \text{otherwise;} \end{cases} \quad (3.47)$$

where  $d_{il} = \|\mathbf{h}_{il}\|$ ,  $\sigma^2 > 0$ , and  $\theta \in [0, 1)$ . These covariance functions specify different rates of decay in the magnitude of the correlation between two observations as a function of their separation distance (Figure 3.1).

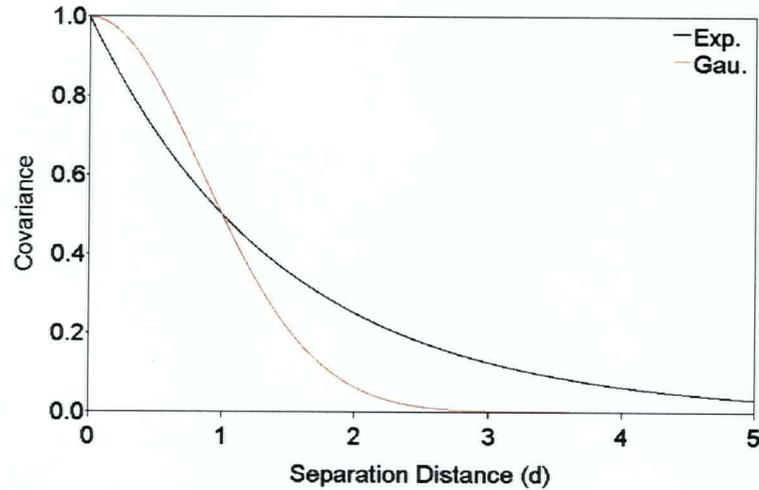


Figure 3.1: Rate of decay of the pairwise covariance between observations as a function of their separation distance under the exponential and Gaussian covariance functions ( $\sigma^2 = 1$ ;  $\theta = 0.5$ ).

An advantage of using RV models is that it is possible to interpret the form of the spatial correlation directly from the covariance function. On the other hand, the selection of an appropriate covariance function for a data set is not straightforward. One approach is to base selection on the empirical covariance function

$$\hat{C}(d) = \frac{1}{|\mathcal{N}_d|} \sum_{\mathcal{N}_d} \hat{\varepsilon}_{\mathbf{s}_i} \hat{\varepsilon}_{\mathbf{s}_j} \quad (3.48)$$

where  $\hat{\varepsilon}_{s_i}$  is an estimator of  $\varepsilon_{s_i}$ ,  $\mathcal{N}_d$  is the set of pairs of observations separated by a distance ( $d$ ), and  $|\mathcal{N}_d|$  is the number of pairs in  $\mathcal{N}_d$  (the summation is over all pairs in  $\mathcal{N}_d$ ). If  $\hat{C}(d)$  is plotted against  $d$ , it may reveal an appropriate form for the covariance function; unfortunately,  $\hat{C}(d)$  is a biased estimator of  $C(d)$  (Cressie 1993). Another approach is to base the selection of the covariance function on a cross-validation procedure. One observation at a time is excluded and subsequently predicted from the rest of the set, and the covariance function that minimizes the mean squared prediction error is taken as the model covariance function. Both approaches are complicated by heteroskedasticity and correlations among model residuals (Ripley 1981; Zimmerman and Harville 1991). In this thesis, only the exponential and Gaussian covariance functions were considered. The RV models corresponding to these two covariance functions will be referred to as the EXP and GAU models, respectively.

### 3.3.2 REML Estimation

In NN and RV models, the vector of observations is modelled as a nonlinear function of the autocorrelation parameter,  $\theta$ . Under these models, OLS estimators of  $\beta$  are not efficient and the OLS estimators of the variance parameters are not consistent. On the other hand, the restricted maximum likelihood (REML) estimators described by Patterson and Thompson (1971, 1974) have desirable asymptotic and finite sample properties in a spatial setting.

Under the assumption of multivariate normality, the log-likelihood of Equation 3.4 is

$$\begin{aligned} L(\beta, \sigma^2, \theta | \mathbf{Y}_S) &= c_0 - \frac{1}{2} \ln |\mathbf{V}| - \frac{1}{2} (\mathbf{Y}_S - \mathbf{X}\beta)^T \mathbf{V}^{-1} (\mathbf{Y}_S - \mathbf{X}\beta) \\ &= c_1 - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln |\mathbf{\Omega}| - \frac{1}{2\sigma^2} (\mathbf{Y}_S - \mathbf{X}\beta)^T \mathbf{\Omega}^{-1} (\mathbf{Y}_S - \mathbf{X}\beta) \end{aligned} \quad (3.49)$$

where  $c_0$  and  $c_1$  are constant terms. The maximum likelihood (ML) estimators of  $\beta$  and  $\sigma^2$  are

$$\hat{\beta}_{ml} = (\mathbf{X}^T \hat{\mathbf{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{\Omega}}^{-1} \mathbf{Y}_S \quad (3.50)$$

and

$$\hat{\sigma}_{ml}^2 = \frac{1}{n} (\mathbf{Y}_S - \mathbf{X}\hat{\beta}_{ml})^T \hat{\mathbf{\Omega}}^{-1} (\mathbf{Y}_S - \mathbf{X}\hat{\beta}_{ml}) \quad (3.51)$$

where  $\hat{\mathbf{\Omega}} = \mathbf{\Omega}(\hat{\theta}_{ml})$  is an estimator of  $\mathbf{\Omega}$  (under the parameterization  $\mathbf{\Omega}(\theta)$ ), and  $\hat{\theta}_{ml}$  is the ML estimator of  $\theta$ . No analytic solution for  $\hat{\theta}_{ml}$  exists, so it must be obtained from a (constrained) maximization of Equation 3.49.

A disadvantage of ML estimation is that the estimator of  $\sigma^2$  (Equation 3.51) is biased in small samples (Judge *et al.* 1985). The ML estimator tends to underestimate  $\sigma^2$ , especially when

the number of fixed-effect parameters  $p$  is large relative to  $n$ . This bias can be attributed to the fact that this estimator does not account for the loss of degrees of freedom incurred in estimating  $\beta$ . To rectify this problem, Patterson and Thompson (1971, 1974) proposed the REML approach.

The basis of the REML approach is to: (1) estimate the variance parameters from a restricted portion of the likelihood that is confined to the  $n - p$  set of linearly independent error contrasts (*i.e.*, contrasts with expectation zero); and then (2) estimate  $\beta$  using the estimated generalised least squares (EGLS) formula. Harville (1974) showed that the restricted likelihood required for variance parameter estimation is

$$L^*(\sigma^2, \theta | \mathbf{Y}_S) = c_3 - \frac{n-p}{2} \ln \sigma^2 - \frac{1}{2} \ln |\boldsymbol{\Omega}| - \frac{1}{2} \ln |\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}| - \frac{1}{2\sigma^2} \mathbf{Y}_S^T \mathbf{Q} \mathbf{Y}_S \quad (3.52)$$

where  $\mathbf{Q} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1}$  and  $c_3$  is a constant.

The REML estimators of the variance parameters are those values of  $\sigma^2$  and  $\theta$  that maximize Equation 3.52. There is no closed form solution for the REML estimator of  $\theta$  but the value of  $\sigma^2$ , say  $\tilde{\sigma}^2$ , for which Equation 3.52 is a maximum is given by

$$\tilde{\sigma}^2 = \frac{1}{n-p} \mathbf{Y}_S^T \mathbf{Q} \mathbf{Y}_S \quad (3.53)$$

(Harville 1977). Substituting Equation 3.53 into Equation 3.52 yields

$$L_c^*(\theta | \mathbf{Y}_S) = c_4 - \frac{n-p}{2} \ln (\mathbf{Y}_S^T \mathbf{Q} \mathbf{Y}_S) - \frac{1}{2} \ln |\boldsymbol{\Omega}| - \frac{1}{2} \ln |\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}| \quad (3.54)$$

where  $L_c^*(\theta)$  is the concentrated restricted log-likelihood function (up to a constant  $c_4$ ). The REML estimator of  $\theta$  is then obtained by maximizing Equation 3.54.

Maximization of Equation 3.54 is a constrained optimization problem ( $\theta \in \Theta \subset [0, 1]$ ). However,  $L_c^*(\cdot)$  is a function of  $\theta$  alone. This greatly facilitates the implementation of numerical search algorithms. In general, search algorithms requiring a user-specified gradient vector and Hessian matrix are preferable (the first and second partial derivatives of Equation 3.54 are given in Appendix A). One impediment to the successful application of search algorithms is the fact that Equation 3.54 can have multiple modes. Under the SAR model, modes will exist for  $L_c^*(\theta)$  at all values of  $\theta$  corresponding to the eigenvalues of  $\mathbf{W}$  when  $\theta$  is not restricted to  $[0, 1/\omega(1)]$ . Dietrich (1991) showed that the multimodality of  $L_c^*(\theta)$  increased up to approximately 10% with increasing sample size under the GAU model. His study also found that  $L_c^*(\theta)$  was consistently unimodal under the EXP model. In practice, because the concentrated log-likelihood is a univariate function and  $\theta$  has a relatively narrow range ( $\Theta \subseteq [0, 1]$ ), visual inspection of the modality of  $L_c^*(\theta)$  is

feasible. This would ensure that the global maximum is found. A second option is to repeat the optimization procedure using different starting points.

Once  $\hat{\theta}_{reml}$  is found, the REML estimator of  $\sigma^2$  is obtained by replacing  $\Omega$  in Equation 3.53 with  $\hat{\Omega} = \Omega(\hat{\theta}_{reml})$ . This substitution yields

$$\hat{\sigma}_{reml}^2 = \frac{1}{n-p} \mathbf{Y}_S^T \hat{\mathbf{Q}} \mathbf{Y}_S \quad (3.55)$$

where  $\hat{\mathbf{Q}} = \hat{\Omega}^{-1} - \hat{\Omega}^{-1} \mathbf{X} (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\Omega}^{-1}$ . The REML estimator of  $\beta$  is subsequently obtained from the EGLS formula

$$\hat{\beta}_{reml} = (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\Omega}^{-1} \mathbf{Y}_S \quad (3.56)$$

Since the REML estimators of  $\sigma^2$  and  $\theta$  are based on a transformation of the log-likelihood, they have several properties in common with ML estimators. In particular, the REML estimators are consistent and asymptotically normally distributed (Cressie and Lahiri 1996). The asymptotic variance matrix is given by the inverse of the matrix

$$\begin{aligned} \mathcal{J}(\sigma^2, \theta) &= -\mathbf{E} \begin{bmatrix} \frac{\partial^2}{\partial \sigma^4} L^*(\sigma^2, \theta) & \frac{\partial^2}{\partial \sigma^2 \partial \theta} L^*(\sigma^2, \theta) \\ \frac{\partial^2}{\partial \sigma^2 \partial \theta} L^*(\sigma^2, \theta) & \frac{\partial^2}{\partial \theta^2} L^*(\sigma^2, \theta) \end{bmatrix} \\ &= \begin{bmatrix} \frac{n-p}{2\sigma^4} & \frac{1}{2\sigma^2} \text{tr} \left( \mathbf{Q} \frac{\partial \Omega}{\partial \theta} \right) \\ \frac{1}{2\sigma^2} \text{tr} \left( \mathbf{Q} \frac{\partial \Omega}{\partial \theta} \right) & \frac{1}{2} \text{tr} \left( \mathbf{Q} \frac{\partial \Omega}{\partial \theta} \mathbf{Q} \frac{\partial \Omega}{\partial \theta} \right) \end{bmatrix} \end{aligned} \quad (3.57)$$

(Cressie and Lahiri 1996; Appendix A). In a spatial setting these asymptotic results are conditional on certain assumptions about the form of the joint pdf of  $\mathbf{Y}_S$  and the relative spatial distribution of the data as  $n \rightarrow \infty$ . The assumptions concerning the joint pdf of  $\mathbf{Y}_S$  are the usual regularity conditions required for ML estimation (see Cassella and Berger 1990). But Cressie and Lahiri (1996) showed that the consistency and asymptotic normality of the REML estimators in a spatial setting also require that the minimum separation distance between observations does not approach 0 as the sample size grows. Finally, note that  $\hat{\sigma}_{reml}^2 = \frac{n}{n-p} \hat{\sigma}_{ml}^2$  for fixed  $\hat{\theta}$ , and that  $\hat{\sigma}_{reml}^2$  also reduces to the OLS estimator given by Equation 3.15 when  $\Omega = \mathbf{I}_n$  (*i.e.*, when  $\theta = 0$ ). Although  $\hat{\sigma}_{reml}^2$  is asymptotically less precise than  $\hat{\sigma}_{ml}^2$ , it has superior finite-sample bias properties.

It is difficult to specify the exact distribution of  $\hat{\beta}_{reml}$  from Equation 3.56 because  $\mathbf{Y}_S$  is correlated with  $\hat{\Omega}$ . However, Breusch (1980) showed that under the assumption of multivariate normality,  $\hat{\beta}_{reml}$  is unbiased (provided the expectation exists) regardless of whether the parameterization of  $\Omega$  is correct. The asymptotic variance of  $\hat{\beta}_{reml}$  is

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\beta}_{reml}) = \sigma^2 (\mathbf{X}^T \Omega^{-1} \mathbf{X})^{-1} \quad (3.58)$$

(Judge *et al.* 1985) and a common estimator of the variance is

$$\widehat{\text{Var}}(\hat{\beta}_{reml}) = \hat{\sigma}_{reml}^2 (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1} \quad (3.59)$$

However, Kenward and Roger (1997) noted that Equation 3.59 yields a biased estimator of  $\text{Var}(\hat{\beta}_{reml})$  because (1) Equation 3.58 does not account for the effect of estimating  $\theta$  and  $\sigma^2$  on the variance of  $\hat{\beta}_{reml}$ ; and (2)  $\widehat{\text{Var}}(\hat{\beta}_{reml})$  is a biased estimator of  $\lim_{n \rightarrow \infty} \text{Var}(\hat{\beta}_{reml})$ . In general, Equation 3.59 will tend to overestimate the precision of  $\hat{\beta}_{reml}$ .

Finally, when all other factors in the model are fixed, a treatment  $F$ -statistic can be computed as

$$F_{(t-1), (n-p)} = \frac{\hat{\tau}_{reml}^T \mathbf{L}^T \left( \mathbf{L} \lambda_t^T (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1} \lambda_t \mathbf{L}^T \right)^{-1} \mathbf{L} \hat{\tau}_{reml}}{(t-1) \hat{\sigma}_{reml}^2} \quad (3.60)$$

where  $\hat{\Omega} = \Omega(\hat{\theta}_{reml})$ . However, the treatment  $F$ -statistic given in Equation 3.60 will not strictly follow an  $F_{(t-1), (n-p)}$  distribution because this statistic does not account for estimation of  $\theta$ . Kenward and Roger (1997) discuss several approximations that could be used in place of Equation 3.60, the properties of which depend on several factors including the model covariance structure.

### 3.4 Model Selection

Spatial analysis methods based on the CE models outlined in Section 3.3 offer potential gains in efficiency and validity relative to ANOVA/OLS. However, these gains are only realized if the data are spatially autocorrelated and if this autocorrelation is appropriately modelled. Legitimate problems are ascertaining when spatial autocorrelation is significant, and choosing among competing spatial CE models.

Since spatial autocorrelation is defined differently by different models, measures of spatial autocorrelation in a data set are contingent on the identification of a suitable model. The most commonly used tests for spatial autocorrelation are based on NN models. For example, the Moran  $I$  statistic (Cliff and Ord 1981) is a ratio of cross products reflecting the degree of association defined by a spatial weight matrix relative to the overall sum of squares:

$$I = \frac{n \mathbf{e}_{ols}^T \mathbf{W} \mathbf{e}_{ols}}{z \mathbf{e}_{ols}^T \mathbf{e}_{ols}} \quad (3.61)$$

where  $\mathbf{W} = \{w_{il}\}$  is the spatial weight matrix ( $w_{ii} = 0$ ),  $\mathbf{e}_{ols}$  is the  $n \times 1$  vector of ANOVA/OLS residuals, and  $z = \sum_{i=1}^n \sum_{l=1}^n w_{il}$  is a constant. This statistic can be used to test the null hypothesis of independence against the alternative hypothesis of non-zero (*i.e.*, positive or negative)

spatial autocorrelation; if the magnitude of  $I$  is sufficiently large, the null hypothesis is rejected. However, the exact distribution of  $I$  under the null hypothesis is complicated by autocorrelation among the ANOVA/OLS residuals. Also, the  $I$  statistic is conditional on the spatial weight matrix providing a suitable reflection of the pairwise dependence among observations but does not account for estimation of  $\mathbf{W}$ .

A more broadly applicable measure for spatial autocorrelation is the likelihood ratio (LR) statistic. The LR statistic ( $\Lambda$ ) is defined as

$$\Lambda = 2 \left[ L(\boldsymbol{\beta}, \sigma^2, \theta \mid \mathbf{Y}_S) - L(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{Y}_S, \theta = 0) \right] \quad (3.62)$$

where  $L(\boldsymbol{\beta}, \sigma^2, \theta \mid \mathbf{Y}_S)$  is the unconstrained log-likelihood of  $\mathbf{Y}_S$  and  $L(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{Y}_S, \theta = 0)$  is the log-likelihood constrained by the condition  $\theta = 0$ . An LR statistic based on the ratio of the restricted log-likelihoods ( $\Lambda^*$ ) can be defined similarly

$$\Lambda^* = 2 \left[ L^*(\sigma^2, \theta \mid \mathbf{Y}_S) - L^*(\sigma^2 \mid \mathbf{Y}_S, \theta = 0) \right] \quad (3.63)$$

where  $L^*(\cdot)$  is the restricted log-likelihood of  $\mathbf{Y}_S$ . The restricted LR statistic given by Equation 3.63 is easier to compute than  $\Lambda$  when REML estimation is used; the unconstrained function  $L^*(\sigma^2, \theta \mid \mathbf{Y}_S)$  is readily available from the REML estimation procedure and  $L^*(\sigma^2 \mid \mathbf{Y}_S)$  is simply a function of  $\hat{\sigma}_{ols}^2$ . The restricted LR statistic converges in distribution to a  $\chi^2$  random variable with a single degree of freedom when  $\theta = 0$  (Beckers 1997). This statistic can therefore be used to test the null hypothesis of no spatial autocorrelation ( $\theta = 0$ ) against an alternative hypothesis, the interpretation of which depends upon the spatial CE model under consideration.

A drawback of the restricted LR test is that it can only be used as a model selection tool when comparing nested models (*i.e.*, a spatial CE model *vs.* ANOVA). This test cannot be used as a basis for model selection when several different spatial CE models are being considered. Model selection should incorporate the principle of parsimony both in terms of the model itself and in terms of the data given the model. The Akaike Information Criterion (AIC)

$$\text{AIC} = -L^*(\sigma^2, \theta \mid \mathbf{Y}_S) + q \quad (3.64)$$

where  $q$  is the number of variance parameters in the analysis model (Akaike 1974), is often used for model selection in time-series settings. The idea is to select the analysis method that yields the smallest AIC. In the context of the analysis methods discussed in this chapter, the AIC penalizes the spatial analysis methods for fitting the additional covariance parameter  $\theta$ .

# Chapter 4

## Methods

### 4.1 Simulation Study

The second objective of this thesis was to evaluate the efficiency and validity of spatial analysis methods for forestry Nelder experiments. When data are spatially autocorrelated and ANOVA/OLS is used, treatment effects are not estimated efficiently and variance estimators are biased. Analysis methods based on spatial CE models and REML estimation may be asymptotically valid and efficient but finite-sample properties can not be established analytically. Therefore, the distributions of the REML estimators, and the efficiency and validity of spatial analysis methods, were evaluated by means of a simulation study.

The simulation study consisted of four trials. In each trial, a different spatial CE model was used to simulate an autocorrelated series at one of several levels of spatial autocorrelation. The models used were the EXP, GAU, SAR, and CAR models discussed in Section 3.3. Different magnitudes of  $\theta$  were simulated in each trial because the level of the spatial autocorrelation implied by  $\theta$  is model-specific. Also, this parameter has a restricted range in the two NN models. In each trial, at every level of  $\theta$ , 500 Nelder experiments were simulated. Since estimation of  $\theta$  is invariant to both  $\sigma^2$  and  $\mu$  (Breusch 1980) these parameters were fixed throughout the study. Every simulated Nelder experiment was analysed using both ANOVA/OLS and a spatial analysis method consisting of the spatial model used in the simulation with REML estimation (Table 4.1).

The simulated spatial lattice and treatment structures were taken directly from the spacing wheel design of the MKN experiment (Section 2.3). Only a single spacing wheel was simulated, as differences among spacing wheels can be modelled by the mean structure. The spatial weight matrix given by Equation 3.34 was used in the SAR and CAR simulations. In each simulation, 113 normally distributed random deviates were generated using the multiplicative congruential random

Table 4.1: Simulation models and analysis methods used in the simulation study. The simulated mean, variance, and spatial autocorrelation are given by  $\mu$ ,  $\sigma^2$ , and  $\theta$ , respectively.

| Trial | Simulation model | $\mu$ | $\sigma^2$ | $\theta$ | Number of simulations | Analysis methods       |
|-------|------------------|-------|------------|----------|-----------------------|------------------------|
| 1     | EXP              | 0     | 1          | 0        | 500                   | ANOVA/OLS,<br>EXP/REML |
|       |                  |       |            | 0.2      | 500                   |                        |
|       |                  |       |            | 0.4      | 500                   |                        |
|       |                  |       |            | 0.6      | 500                   |                        |
| 2     | GAU              | 0     | 1          | 0        | 500                   | ANOVA/OLS,<br>GAU/REML |
|       |                  |       |            | 0.2      | 500                   |                        |
|       |                  |       |            | 0.4      | 500                   |                        |
|       |                  |       |            | 0.6      | 500                   |                        |
| 3     | SAR              | 0     | 1          | 0        | 500                   | ANOVA/OLS,<br>SAR/REML |
|       |                  |       |            | 0.1      | 500                   |                        |
|       |                  |       |            | 0.2      | 500                   |                        |
| 4     | CAR              | 0     | 1          | 0        | 500                   | ANOVA/OLS,<br>CAR/REML |
|       |                  |       |            | 0.1      | 500                   |                        |
|       |                  |       |            | 0.2      | 500                   |                        |

number generator in SAS/IML (SAS Institute Inc. 1999). These independent normal deviates were then assigned to 25 levels of tree density and transformed through the root of the simulated variance matrix to produce an autocorrelated series. For the SAR model trial, the root of the variance matrix was obtained directly. For the other three simulation models, the Cholesky decomposition (Cressie 1993) was applied to the simulated variance matrix to obtain the necessary transformation. The effects of tree density ( $\tau$ ) were assigned by dividing the 25 levels of tree density into three classes. The 8 highest density treatments were assigned a value of  $-0.2$ , the 8 lowest density treatments a value of  $+0.2$ , and the 9 intermediate treatments a value of zero (Table 4.2). This provided the

Table 4.2: Simulated treatments effects by tree density class.

| Class | Tree density (stems/ha) | Simulated treatment effect ( $\tau$ ) |
|-------|-------------------------|---------------------------------------|
| 1     | > 2500                  | $-0.2$                                |
| 2     | 800-2500                | 0.0                                   |
| 3     | < 800                   | $+0.2$                                |

means to evaluate the effects of spatial autocorrelation on both the power and size of treatment  $F$ -tests for different analysis methods.

REML estimation of covariance parameters was performed using the Newton-Raphson line search algorithm in SAS/IML (SAS Institute Inc. 1999). The range of  $\hat{\theta}_{reml}$  under both NN models was restricted to the interval  $[0, 0.275]$ , as 0.275 corresponded to the reciprocal of the largest eigenvalue of the spatial weight matrix. For the RV models, estimates of  $\theta$  were restricted to the range  $[0, 0.99]$  since the variance matrix is singular at  $\theta = 1$ .

The efficiency and validity of each spatial analysis method relative to ANOVA/OLS was assessed following the procedure outlined by Besag and Kempton (1986). Under this scheme, the average empirical variance (EMP) of the set of pairwise treatment contrasts is used to compare the efficiency of each pair of analysis methods in a simulation trial at each level of  $\theta$ . The empirical variance ( $\text{Var}_E$ ) of an estimated treatment contrast ( $\delta_{\langle i,j \rangle}^T \hat{\tau} = \hat{\tau}_i - \hat{\tau}_j$ ) is

$$\text{Var}_E(\delta_{\langle i,j \rangle}^T \hat{\tau}) = \frac{1}{500} \sum_{k=1}^{500} \left[ \left( \delta_{\langle i,j \rangle}^T \hat{\tau}_{[k]} \right) - \left( \delta_{\langle i,j \rangle}^T \tau \right) \right]^2 \quad (4.1)$$

where  $\delta_{\langle i,j \rangle}$  is a  $t \times 1$  vector with a +1 and a -1 in the  $i$ th and  $j$ th rows ( $i, j = 1, 2, \dots, t$ ), respectively, and  $\hat{\tau}_{[k]}$  is the estimate of  $\tau$  from the  $k$ th simulation. Equation 4.1 is the average squared deviation of  $\delta_{\langle i,j \rangle}^T \hat{\tau}$  around its expected value,  $\delta_{\langle i,j \rangle}^T \tau$  ( $\hat{\tau}$  is unbiased). EMP is obtained from Equation 4.1 by averaging over all pairwise treatment contrasts

$$\text{EMP} = \frac{2}{t(t-1)} \sum_{i=1}^t \sum_{j>i}^t \text{Var}_E(\delta_{\langle i,j \rangle}^T \hat{\tau}) \quad (4.2)$$

Smaller values of EMP correspond to greater precision. Therefore, in each trial, the relative efficiency of the spatial analysis method (relative to ANOVA/OLS) could be evaluated at each level of  $\theta$  through the EMP statistics. The relative empirical variance (REV) of an analysis method was calculated as

$$\text{REV} = \frac{\text{EMP}_{\mathcal{M}}}{\text{EMP}_{\text{ANOVA/OLS}}} \times 100\% \quad (4.3)$$

where  $\text{EMP}_{\mathcal{M}}$  is the EMP of the analysis method and  $\text{EMP}_{\text{ANOVA/OLS}}$  is the EMP of ANOVA/OLS. For example, in trial 1, the REV of EXP/REML would be

$$\text{REV} = \frac{\text{EMP}_{\text{EXP/REML}}}{\text{EMP}_{\text{ANOVA/OLS}}} \times 100\% \quad (4.4)$$

for a given  $\theta$ . By this convention, the REV of ANOVA/OLS is always 100%.

The validity of an analysis method can be assessed by comparing EMP to its estimated value, the average predicted variance (PRE) of the set of all pairwise treatment contrasts. The average predicted variance of a treatment contrast is

$$\widehat{\text{Var}}(\delta_{\langle i,j \rangle}^T \hat{\tau}) = \frac{1}{500} \sum_{k=1}^{500} \hat{\sigma}_{[k]}^2 (\delta_{\langle i,j \rangle}^T \lambda_t (\mathbf{X}^T \hat{\Omega}_{[k]}^{-1} \mathbf{X})^{-1} \lambda_t \delta_{\langle i,j \rangle}) \quad (4.5)$$

where  $\hat{\Omega}_{[k]} = \Omega(\hat{\theta}_{[k]})$ ;  $\hat{\sigma}_{[k]}^2$  and  $\hat{\theta}_{[k]}$  are the estimates of  $\sigma^2$  and  $\theta$  from the  $k$ th simulation, respectively; and  $\lambda_t$  defines the linear transformation such that  $\tau = \lambda_t^T \beta$ . PRE is the average of  $\widehat{\text{Var}}(\delta_{(i,j)}^T \hat{\tau})$  over the  $t(t-1)/2$  pairwise treatment contrasts

$$\text{PRE} = \frac{2}{t(t-1)} \sum_{i=1}^t \sum_{j>i}^t \widehat{\text{Var}}(\delta_{(i,j)}^T \hat{\tau}) \quad (4.6)$$

If an analysis method is valid, estimates of precision are unbiased and PRE should be approximately equal to EMP. If EMP-PRE is positive (negative), it is an indication that the variance of treatment estimates is underestimated (overestimated) (Brownie and Gumpertz 1997). In each trial, the relative variance bias (RVB) of an analysis method at a given level of  $\theta$  is

$$\text{RVB} = \frac{\text{EMP}_{\mathcal{M}} - \text{PRE}_{\mathcal{M}}}{\text{EMP}_{\text{ANOVA/OLS}}} \times 100\% \quad (4.7)$$

where  $\text{EMP}_{\mathcal{M}}$  and  $\text{PRE}_{\mathcal{M}}$  are the EMP and PRE of the analysis method and  $\text{EMP}_{\text{ANOVA/OLS}}$  is the EMP of ANOVA/OLS. The RVB of ANOVA/OLS corresponds to the percent difference between EMP and PRE since  $\text{EMP}_{\text{ANOVA/OLS}}$  is the denominator of Equation 4.7.

The distributions of  $\hat{\theta}$ ,  $\hat{\sigma}^2$ , and the treatment  $F$ -statistics for both the spatial analysis method and ANOVA/OLS were also of interest. In particular, the distributions of the  $F$ -statistics can be used to assess the power and size of resultant  $F$ -tests. The treatment effect structure (Table 4.2) provided the opportunity to examine the power and size of different treatment  $F$ -tests. The overall treatment effect was partitioned into an among-classes component and a within-classes component (Table 4.3). The significance of each of these components was tested against the experimental error. The empirical rejection rate of the among-classes  $F$ -test provided an assessment of its power (*i.e.*, the probability of rejecting the false null hypothesis). Similarly, the empirical rejection rate of the within-classes  $F$ -test provided an assessment of its size (*i.e.*, the probability of not rejecting the true null hypothesis). The power of the ANOVA/OLS overall treatment  $F$ -test under this treatment effect structure and zero spatial autocorrelation is approximately 0.1 at the 0.05 significance level.

In each trial, the empirical rejection rate of the restricted LR test at the 0.1 significance level was calculated for each simulated level of  $\theta$ . At  $\theta = 0$ , this rejection rate indicated the size of this test; for  $\theta \neq 0$ , the rejection rate indicated the power of the test.

For each trial, and each level of  $\theta$ , the proportion of simulations or rate at which

$$\text{AIC}_S \leq \text{AIC}_{\text{ANOVA/OLS}} \quad (4.8)$$

where  $\text{AIC}_S$  was the AIC for the spatial analysis method and  $\text{AIC}_{\text{ANOVA/OLS}}$  was the AIC for ANOVA/OLS, was also tracked across the simulated ranges of  $\theta$ . This provided an indication of the value of this comparison with respect to model identification and selection.

Table 4.3: Analysis of variance table for the simulated Nelder experiments.

| Source of Variation | Degrees of freedom | Mean Square | $F$ -test           |
|---------------------|--------------------|-------------|---------------------|
| Treatments          | 24                 | $MS_{TRT}$  | $MS_{TRT}/MS_{ERR}$ |
| Among-classes       | 2                  | $MS_A$      | $MS_A/MS_{ERR}$     |
| Within-classes      | 22                 | $MS_B$      | $MS_B/MS_{ERR}$     |
| Error               | 88                 | $MS_{ERR}$  |                     |
| Total               | 112                |             |                     |

## 4.2 Analysis of the MKN Experiment

The third objective of this thesis was to assess the practicality of spatial analysis methods for forestry Nelder experiments. ANOVA/OLS, and each of the four spatial analysis methods evaluated in the simulation study (Section 4.1), were therefore applied to data from the MKN experiment (Section 2.3).

The analysis of variance table for the MKN experiment age 15 diameter outside bark at 1.3 m above ground (DBH15) data set is given in Table 4.4. The difference between the two

Table 4.4: Analysis of variance table for the MKN experiment DBH15 data set.

| Source of Variation                 | Degrees of freedom | Mean Square | $F$ -test           |
|-------------------------------------|--------------------|-------------|---------------------|
| Spacing wheel                       | 1                  | $MS_{SW}$   | $MS_{SW}/MS_{ERR}$  |
| Tree density                        | 24                 | $MS_{TRT}$  | $MS_{TRT}/MS_{ERR}$ |
| Spacing wheel $\times$ tree density | 24                 | $MS_{INT}$  | $MS_{INT}/MS_{ERR}$ |
| Error                               | 163                | $MS_{ERR}$  |                     |
| Total                               | 212                |             |                     |

spacing wheels was modeled as a fixed effect. As might be expected, the DBH15 data set was incomplete; 3 observations were missing, all from the same spacing wheel. Missing data is a major concern in spacing experiments, because it implies that the original treatment structure has been compromised. It is straightforward to modify an analysis to accommodate missing data when it is known that their absence is due to errors made during collection or recording. But if data may be missing because trees have died, the analysis must be modified to account for concomitant changes in tree density. In the analyses undertaken here, the assumption was made that changes in tree density resulting from tree mortality would not have had significant effects on tree growth outside the immediate vicinity of the dead trees. Therefore, at each of the three points in the

MKN experiment with missing DBH15 values, the nearest inter-spoke and intra-spoke trees were identified and excluded from the analyses. As a result, all five analyses were based on a total of only 213 trees: 113 trees from one spacing wheel, 100 from the other.

The spatial analysis methods modeled spatial autocorrelation only within spacing wheels. The SAR/REML and CAR/REML analysis methods used the spatial weight function given by Equation 3.34. The restricted likelihood functions for the spatial analysis methods were based on a multivariate normal distribution with variance matrix

$$\mathbf{V} = \sigma^2 \mathbf{\Omega} = \sigma^2 \begin{bmatrix} \mathbf{\Omega}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_2 \end{bmatrix} \quad (4.9)$$

where  $\mathbf{\Omega}_j$  is the covariance matrix for the  $j$ th spacing wheel ( $j = 1, 2$ ). REML estimation was based on the Newton-Raphson line search algorithm in SAS/IML (SAS Institute Inc. 1999). The range of possible values for  $\hat{\theta}_{reml}$  was restricted to the interval  $[0, 0.339]$  for both SAR/REML and CAR/REML and to  $[0, 0.99]$  for EXP/REML and GAU/REML.

# Chapter 5

## Results and Discussion

### 5.1 Results

The second and third objectives of this thesis were addressed by means of the simulation study and MKN DBH15 analysis described in Chapter 4. The results of these two applications are presented below.

#### 5.1.1 Simulation Study

The REV and RVB of ANOVA/OLS and the spatial analysis method used in each simulation trial are presented in Figure 5.1. Each graph in Figure 5.1 is a plot of RVB *vs.* REV for the two analysis methods at the levels of spatial autocorrelation simulated in one of the four trials. The trial numbers and the spatial CE simulation models are indicated above each graph. The simulated level of autocorrelation is referenced by its decimal digit (*e.g.*,  $\theta = 0.2$  is indicated by a 2).

The REV of EXP/REML and of GAU/REML in trials 1 and 2, respectively, tended to decrease with the magnitude of the spatial correlation (Figure 5.1). At the highest levels of spatial autocorrelation ( $\theta = 0.6$ ) the EMP statistics for these two spatial analysis methods were approximately 2% smaller than  $EMP_{ANOVA/OLS}$ . In contrast, the REV of the two analysis methods based on NN models, SAR/REML and CAR/REML, increased with increasing spatial autocorrelation in trials 3 and 4. ANOVA/OLS was more efficient than these two spatial analysis methods at all levels of  $\theta$ .

The RVB of ANOVA/OLS increased with the level of spatial autocorrelation in all four trials (Figure 5.1). In all trials,  $PRE_{ANOVA/OLS}$  underestimated  $EMP_{ANOVA/OLS}$  when  $\theta$  was greater than zero. The RVB of EXP/REML in trial 1, and the RVB of GAU/REML in trial 2, were consistently below zero. The RVB of EXP/REML was lowest for  $\theta = 0.6$  ( $-3.4\%$ ) while the RVB of

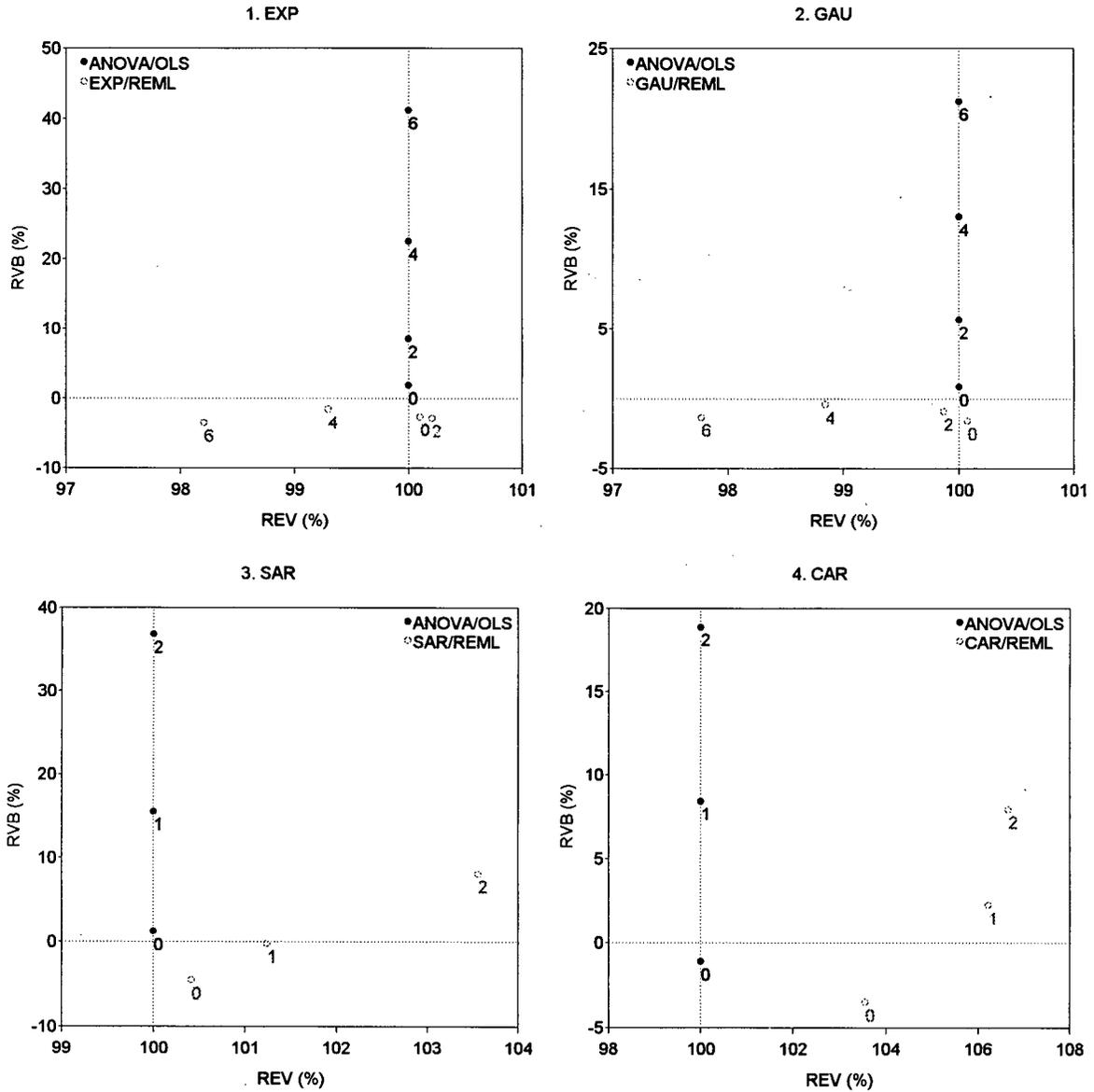


Figure 5.1: Relative variance bias (RVB) and relative empirical variance (REV) of the two analysis methods used in each simulation trial. The simulated level of spatial autocorrelation ( $\theta$ ) is indicated by its decimal digit (e.g.,  $\theta = 0.2$  is indicated by a 2).

EXP/REML was lowest for  $\theta = 0$  ( $-1.5\%$ ). The RVB of SAR/REML and the RVB of CAR/REML increased with the level of spatial autocorrelation in trials 3 and 4, respectively. For these two spatial analysis methods, PRE overestimated EMP at  $\theta = 0$  and underestimated EMP at  $\theta = 0.2$ .

The means and standard deviations of the variance parameter estimators are given in Table 5.1 (recall that  $\sigma^2 = 1$  in all simulations). In trials 1 and 2, the bias in  $\hat{\sigma}_{ols}^2$  increased with the level of spatial autocorrelation but the mean of the REML estimator of  $\sigma^2$  was consistently close to 1. However, the precision of  $\hat{\sigma}_{reml}^2$  was less than that of  $\hat{\sigma}_{ols}^2$  and decreased as  $\theta$  increased in these

Table 5.1: Mean and standard deviations of the OLS and REML estimators of  $\sigma^2$  and  $\theta$  at each level of spatial autocorrelation ( $\theta$ ), by simulation trial (500 replications).

| Trial  | $\theta$ | $\hat{\sigma}_{ols}^2$ |          | $\hat{\sigma}_{reml}^2$ |          | $\hat{\theta}_{reml}$ |          |
|--------|----------|------------------------|----------|-------------------------|----------|-----------------------|----------|
|        |          | Mean                   | Std.Dev. | Mean                    | Std.Dev. | Mean                  | Std.Dev. |
| 1. EXP | 0        | 0.998                  | 0.152    | 1.011                   | 0.157    | 0.086                 | 0.123    |
|        | 0.2      | 0.978                  | 0.147    | 1.013                   | 0.161    | 0.207                 | 0.160    |
|        | 0.4      | 0.922                  | 0.142    | 1.001                   | 0.175    | 0.383                 | 0.150    |
|        | 0.6      | 0.831                  | 0.142    | 1.013                   | 0.230    | 0.593                 | 0.114    |
| 2. GAU | 0        | 1.000                  | 0.150    | 1.007                   | 0.152    | 0.073                 | 0.103    |
|        | 0.2      | 0.983                  | 0.150    | 1.002                   | 0.156    | 0.209                 | 0.129    |
|        | 0.4      | 0.963                  | 0.152    | 1.003                   | 0.160    | 0.409                 | 0.092    |
|        | 0.6      | 0.923                  | 0.150    | 1.000                   | 0.165    | 0.603                 | 0.048    |
| 3. SAR | 0        | 0.992                  | 0.155    | 0.995                   | 0.155    | 0.034                 | 0.047    |
|        | 0.1      | 0.989                  | 0.151    | 0.983                   | 0.148    | 0.095                 | 0.064    |
|        | 0.2      | 1.062                  | 0.177    | 0.982                   | 0.145    | 0.185                 | 0.052    |
| 4. CAR | 0        | 1.002                  | 0.161    | 1.003                   | 0.162    | 0.054                 | 0.076    |
|        | 0.1      | 0.996                  | 0.150    | 0.991                   | 0.150    | 0.110                 | 0.094    |
|        | 0.2      | 0.998                  | 0.148    | 0.981                   | 0.145    | 0.164                 | 0.090    |

two trials. In trials 3 and 4, the mean of the OLS estimator of  $\sigma^2$  was close to 1 at all levels of  $\theta$ . Also, in these last two trials the bias in  $\hat{\sigma}_{reml}^2$  increased, though its precision decreased, with the level of spatial autocorrelation. Overall, the biases in  $\hat{\theta}_{reml}$  were greatest at  $\theta = 0$  and, in trials 3 and 4, were also notable at  $\theta = 0.2$ . The standard deviation of  $\hat{\theta}_{reml}$  was always greater towards the centre of the sampled range of  $\theta$  than at the extremities.

The empirical rejection rates for among- and within-class treatment  $F$ -tests are summarized in Figure 5.2. Each graph displays the rejection rates for the null hypothesis of no differences among treatment classes on the vertical axis and for the null hypothesis of no differences within treatment classes on the horizontal axis. The trial and simulation model are indicated above each graph, and the results are colour-coded by analysis method. As in Figure 5.1, the level of spatial autocorrelation is indicated by its decimal digit (*e.g.*,  $\theta = 0$  is referenced by the number 0). The rejection rates differed by analysis method within simulation trials (*e.g.*, ANOVA/OLS *vs.* EXP/REML) because the two analysis methods used different estimates of  $\tau$  and of  $\mathbf{V}$ .

The size of the ANOVA/OLS within-class treatment  $F$ -test was badly controlled (Figure 5.2) when  $\theta > 0$  in all four trials. At the 0.05 significance level, the true null hypothesis was rejected under ANOVA/OLS in over 40% of the simulations in trial 1 at  $\theta = 0.6$ . By comparison, the rejection

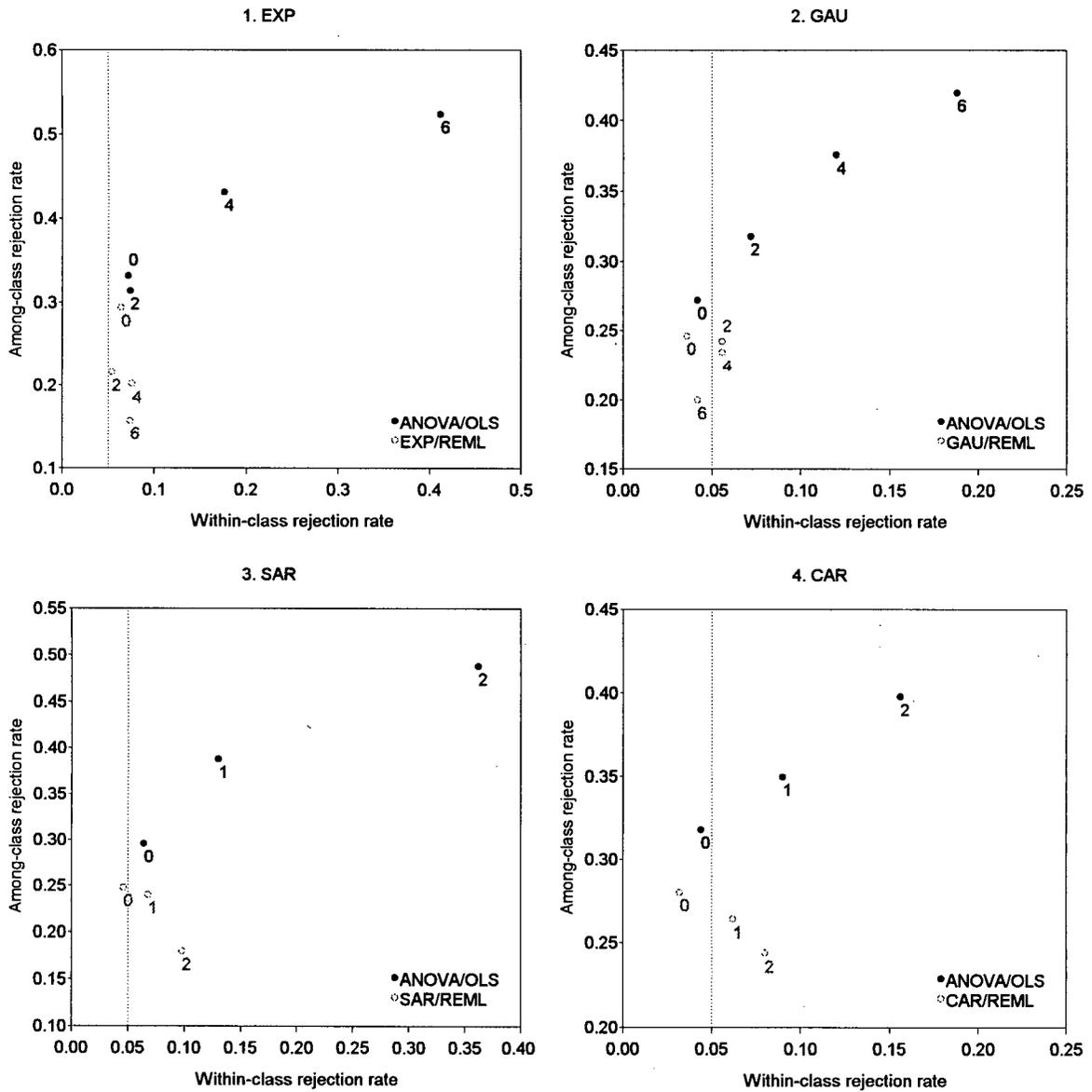


Figure 5.2: Empirical rejection rates (at the 0.05 significance level) of among- and within-class treatment  $F$ -tests by analysis method, level of spatial autocorrelation, and simulation trial. The level of spatial autocorrelation ( $\theta$ ) is indicated by its decimal digit (e.g.,  $\theta = 0.2$  is referenced by the number 2).

rates of the same test for the spatial analysis methods were lower and consistently below 0.1.

The empirical rejection rates of the false null hypothesis of no differences among treatment classes were higher for ANOVA/OLS in all trials at all levels of spatial autocorrelation (Figure 5.2). This rejection rate for ANOVA/OLS also increased with the level of spatial autocorrelation. The rejection rates of the same test for the spatial analysis methods decreased with the level of spatial autocorrelation.

The empirical rejection rate of the restricted LR test, and the proportion of simulations in which  $AIC_{ANOVA/OLS}$  exceeded the AIC of the spatial analysis method, are shown in Figure 5.3. The null hypothesis of  $\theta = 0$  (no spatial autocorrelation) and the 0.1 significance level were used for the restricted LR test.

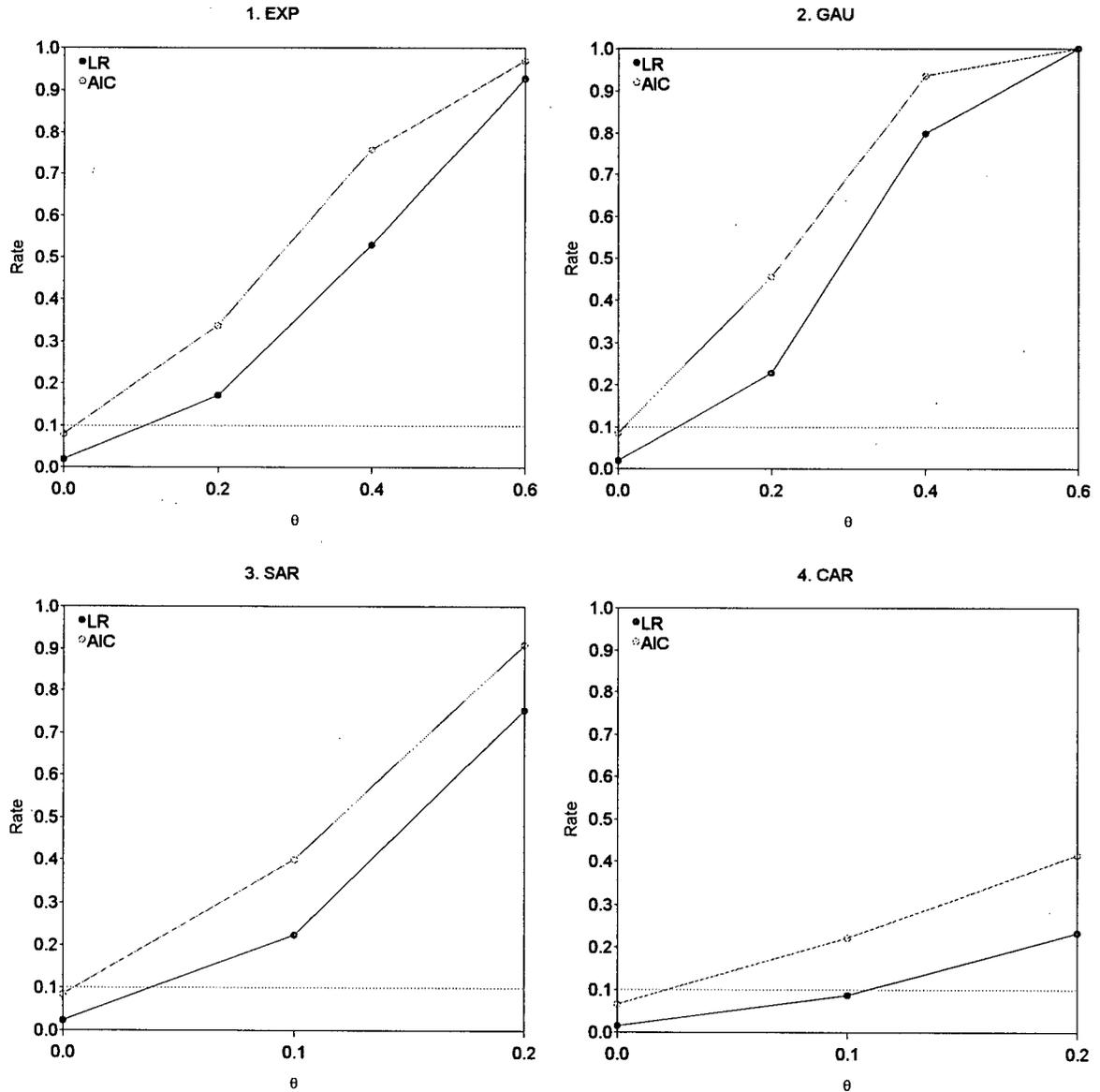


Figure 5.3: Empirical rejection rates (at the 0.1 significance level) of the restricted LR test for zero spatial autocorrelation, and the rates at which the AIC for ANOVA/OLS exceeded the AIC for the spatial analysis method, in each simulation trial as a function of the spatial autocorrelation ( $\theta$ ).

The proportion of simulations in which  $AIC_{ANOVA/OLS}$  exceeded the AIC of the spatial analysis method was at least as high as the rejection rate for the restricted LR test (Figure 5.3). Also, when  $\theta = 0$  the restricted LR test consistently rejected the null hypothesis in fewer than 10%

of the simulations. Although the rejection rate of the restricted LR test increased with  $\theta$  more rapidly in some trials than in others, the scales of the horizontal axes are not strictly comparable. The same value of  $\theta$  does not imply the same magnitude of spatial autocorrelation in different models.

### 5.1.2 MKN Analysis

The ANOVA/OLS analysis of variance table for the DBH15 data set is given in Table 5.2. Note that differences among spacing wheels and differences within tree density treatments between spacing wheels were modeled as fixed effects. Under ANOVA/OLS, interaction and treatment effects were both significant at the 0.05 significance level. The experimental error (*i.e.*,  $MS_{ERR} = \hat{\sigma}_{ols}^2$ ) corresponded to an estimated coefficient of variation of 22.7% ( $\bar{y} = 129.0$ ).

Table 5.2: ANOVA/OLS analysis of variance table for the MKN experiment DBH15 data set.

| Source of Variation                 | Degrees of freedom | Mean square | $F$   | $p^1$   |
|-------------------------------------|--------------------|-------------|-------|---------|
| Spacing wheel                       | 1                  | 1621.3      | 1.91  | 0.168   |
| Tree density                        | 24                 | 8673.9      | 10.24 | < 0.001 |
| Spacing wheel $\times$ tree density | 24                 | 2114.4      | 2.50  | < 0.001 |
| Error                               | 163                | 846.9       |       |         |
| Total                               | 212                |             |       |         |

<sup>1</sup> Probability of observing an  $F$ -statistic at least as great under the null hypothesis of no differences among factor effect levels.

The frequency distribution of the ANOVA/OLS standardized residuals (*i.e.*,  $\mathbf{e}_{ols}/\hat{\sigma}_{ols}$ ; Figure 5.4) was not significantly different from a normal distribution at the 0.05 significance level (Shapiro-Wilk statistic 0.979;  $p$ -value 0.251). The plot of these standardized residuals against the predicted DBH15 value did not reveal any heteroskedastic trends.

The estimates of the scale and spatial autocorrelation parameters obtained from each analysis method are given in Table 5.3. Convergence was obtained for the REML estimators of  $\theta$  for SAR/REML and CAR/REML. Estimates of  $\theta$  for both EXP/REML and GAU/REML were too close to zero to obtain absolute convergence. The concentrated restricted log-likelihood was examined across the potential ranges of  $\hat{\theta}$  for each spatial analysis method to verify that the REML estimates were in the neighbourhood of a global maximum.

The restricted LR statistic was used to test the null hypothesis of zero spatial autocorrelation ( $\theta = 0$ ) for each spatial analysis method (Table 5.3). The restricted LR tests indicated that the

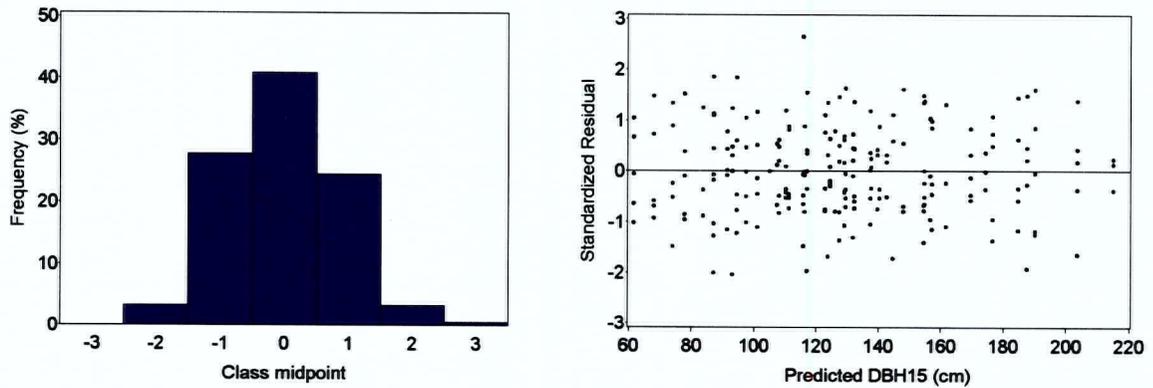


Figure 5.4: Frequency distribution and residual plot for the standardized ANOVA/OLS MKN DBH15 residuals.

DBH15 data did not exhibit significant positive spatial autocorrelation, as defined by any of the four simulation models applied. The restricted LR statistic for CAR/REML was negative, indicating that the resultant CAR model was a worse fit for the data than the i.i.d. ANOVA model. ANOVA/OLS also had the smallest AIC among the five analysis methods applied (Table 5.3).

Table 5.3: Estimates of scale ( $\sigma^2$ ) and spatial autocorrelation ( $\theta$ ) parameters, AIC, and restricted LR ( $\Lambda^*$ ) statistics (with the associated  $p$ -values) for the MKN DBH15 data set, by analysis method.

| Analysis method | $\hat{\sigma}^2$ | $\hat{\theta}$         | AIC    | $\Lambda^*$ | $p^1$ |
|-----------------|------------------|------------------------|--------|-------------|-------|
| ANOVA/OLS       | 846.9            | -                      | 781.72 | -           | -     |
| EXP/REML        | 846.9            | $1.36 \times 10^{-18}$ | 782.72 | 0.000       | 1.000 |
| GAU/REML        | 846.9            | $9.88 \times 10^{-19}$ | 782.72 | 0.000       | 1.000 |
| SAR/REML        | 855.7            | 0.045                  | 782.40 | 0.645       | 0.422 |
| CAR/REML        | 873.5            | 0.106                  | 782.99 | -0.522      | N/A   |

<sup>1</sup> Probability of observing an  $\Lambda^*$  statistic at least as great under the null hypothesis of  $\theta = 0$ .

Estimates of the effects of tree density differed little among the analysis methods applied. Rankings of the tree density effects as estimated by ANOVA/OLS, EXP/REML, GAU/REML, and SAR/REML were identical (Table 5.4). However, rankings under the CAR/REML analysis method were different. Only the average treatment effect rankings are listed in Table 5.4. Due to the significant interaction effect for tree density by spacing wheel, the within spacing wheel treatment effect rankings may be different.

Table 5.4: Rankings of estimated tree density effects for the MKN DBH15 dataset, by analysis method (largest effect is rank 1; ranks different from ANOVA/OLS are in bold).

| Tree density<br>(stems/ha) | Estimated effect rank |          |          |          |           |
|----------------------------|-----------------------|----------|----------|----------|-----------|
|                            | ANOVA/OLS             | EXP/REML | GAU/REML | SAR/REML | CAR/REML  |
| 17,001                     | 23                    | 23       | 23       | 23       | 23        |
| 12,321                     | 16                    | 16       | 16       | 16       | 16        |
| 8,931                      | 24                    | 24       | 24       | 24       | 24        |
| 6,475                      | 25                    | 25       | 25       | 25       | 25        |
| 4,693                      | 18                    | 18       | 18       | 18       | 18        |
| 3,401                      | 22                    | 22       | 22       | 22       | 22        |
| 3,089                      | 7                     | 7        | 7        | 7        | 7         |
| 3,042                      | 14                    | 14       | 14       | 14       | 14        |
| 2,464                      | 17                    | 17       | 17       | 17       | 17        |
| 2,348                      | 21                    | 21       | 21       | 21       | 21        |
| 1,787                      | 20                    | 20       | 20       | 20       | <b>19</b> |
| 1,607                      | 13                    | 13       | 13       | 13       | 13        |
| 1,597                      | 11                    | 11       | 11       | 11       | 11        |
| 1,295                      | 15                    | 15       | 15       | 15       | 15        |
| 949                        | 19                    | 19       | 19       | 19       | <b>20</b> |
| 875                        | 10                    | 10       | 10       | 10       | 10        |
| 850                        | 8                     | 8        | 8        | 8        | <b>9</b>  |
| 690                        | 12                    | 12       | 12       | 12       | 12        |
| 504                        | 9                     | 9        | 9        | 9        | <b>8</b>  |
| 450                        | 4                     | 4        | 4        | 4        | 4         |
| 368                        | 5                     | 5        | 5        | 5        | 5         |
| 267                        | 6                     | 6        | 6        | 6        | 6         |
| 242                        | 3                     | 3        | 3        | 3        | 3         |
| 134                        | 2                     | 2        | 2        | 2        | 2         |
| 131                        | 1                     | 1        | 1        | 1        | 1         |

## 5.2 Discussion

The simulation study described in Section 4.1 was undertaken to answer three questions. The first was whether the spatial analysis methods described in Chapter 3 were more efficient than ANOVA/OLS when applied to spatially autocorrelated Nelder experiments. The second question was whether the estimates of precision supplied by the spatial analysis methods were unbiased. The last question was whether the restricted LR test, or the AIC, could be used to determine the presence or significance of spatial autocorrelation in a data set. The analyses of the MKN DBH15 data set were undertaken to assess the practical properties of spatial analysis methods relative to ANOVA/OLS.

### 5.2.1 Simulation Study

The two spatial analysis methods based on RV models, EXP/REML and GAU/REML, were slightly less efficient than ANOVA/OLS at  $\theta = 0$  (Figure 5.1). EXP/REML was also less efficient than ANOVA/OLS at  $\theta = 0.2$ . As the spatial autocorrelation was increased, however, the relative efficiency of these two spatial analysis methods increased. This result was consistent with findings from other published studies, but the observed degree of difference in precision was relatively minor, and deserves some attention.

Neither the EXP/REML nor GAU/REML analysis methods offered much more than a 2% gain over ANOVA/OLS in the efficiency of treatment effect estimators over the range of simulated autocorrelation (Figure 5.1). Zimmerman and Harville (1991) reported much larger gains in the relative efficiency of spatial analysis methods based on RV models. Their study was based on agricultural uniformity trial<sup>1</sup> data containing strong large-scale spatial trends. They surmised that some of the apparent efficiency of the spatial analysis methods relative to ANOVA/OLS in their study could be attributed to the potential for these analysis models to “soak up” some of this large-scale spatial variation. No large-scale fixed trends were simulated in this study.

Brownie and Gumpertz (1997) also reported more appreciable increases in the efficiency of treatment effect estimators of analysis methods based on RV models relative to ANOVA/OLS with increasing spatial autocorrelation. Their study was simulation-based, but the simulated field experiment was constructed on a regular spatial grid and treatments were randomly allocated to experimental units. Their study showed that EXP/REML could be as much as 40% more efficient than ANOVA/OLS. The discrepancy between the results of the present study and those of Brownie and Gumpertz may be due in part to the differences between the irregularly spaced but more symmetrical systematic Nelder experiment simulated here, and the randomized rectangular field experiment simulated by Brownie and Gumpertz. This point is considered further below.

The simulation study also indicated that the analysis methods based on NN models are relatively less efficient than ANOVA/OLS (Figure 5.1). Furthermore, the relative efficiency of SAR/REML and CAR/REML decreased with the magnitude of the spatial autocorrelation. This was a surprising and somewhat disconcerting result given the popularity of these two analysis methods. As Wall (2000) noted, these analysis methods have achieved sufficient popularity that the statistical software package S-Plus<sup>®</sup> has built-in functions for fitting SAR and CAR models. The efficiency of ANOVA/OLS relative to these two spatial analysis methods in the present study

---

<sup>1</sup>A field experiment conducted with an uniformly applied treatment.

may have been due in part to the irregular spatial distribution of the observations in the simulated Nelder experiments. These NN models were originally conceived of as models for processes distributed over regular, infinite lattices (Whittle 1954, Besag 1974). The poor performance of these analysis methods may also have been due to the complex nature of the correlation structure implied by the SAR and CAR models. Because the correlation structure is modelled through the inverse variance matrix, the resultant variance matrices do not generally have constant diagonal elements or even similar covariances among pairs of observations separated by the same distance (Wall 2000). This inherent lack of stationarity was built into the SAR and CAR simulations.

Regardless of their relative efficiency, however, unless estimates of the precision of treatment contrasts are unbiased, more complex analysis methods are of limited value. The simulation study demonstrated that, as expected, ANOVA/OLS is not a valid analysis method when data are spatially autocorrelated. ANOVA/OLS underestimated the average precision of the treatment contrast estimators when  $\theta$  was greater than zero, and the bias increased with the spatial autocorrelation.

Brownie and Gumpertz (1997) also examined the validity of ANOVA/OLS through the distribution of EMP and PRE statistics. In their study,  $EMP_{ANOVA/OLS} - PRE_{ANOVA/OLS}$  was negative when the data were spatially autocorrelated, implying that the variance estimators were positively biased. This seemingly contradictory result can again be explained by the different treatment structure imposed in their simulations. Brownie and Gumpertz noted that  $PRE_{ANOVA/OLS}$  was greater than  $EMP_{ANOVA/OLS}$  because of an induced block effect. This induced block effect resulted from the simulated spatial autocorrelation and a restricted randomization of treatments to experimental units. Since no block effects were included in their analysis models, the induced block effect was allocated to the experimental error. In the present study, spatial trends may have been induced by the simulated spatial autocorrelation. However, because the simulated treatment structure was constant throughout, the spatial trend was divided into variability within and among treatments.

ANOVA/OLS was not valid in any of the simulation trials when  $\theta$  was greater than zero, but the validity of the spatial analysis methods varied by trial (Figure 5.1). In trials 1 and 2, application of the spatial analysis method corresponding to the simulation model resulted in relatively minor differences between EMP and PRE. These results were consistent with those of Brownie and Gumpertz (1997). The REML estimators of  $\theta$  and  $\sigma^2$  also performed well in the first two trials (Table 5.1). In trials 3 and 4, the spatial analysis methods were not valid and the REML estimators of  $\sigma^2$  were biased. In these two trials,  $\hat{\sigma}_{ols}^2$  was a better estimator of  $\sigma^2$  than  $\hat{\sigma}_{reml}^2$ .

The simulated spacing wheel is an example of a case in which REML estimation of the scale

parameter is preferable to ML estimation. The number of fixed effect parameters ( $p = 25$ ) in the analysis models was large relative to total number of observations ( $n = 113$ ). In this case, the finite sample bias of the ML estimator of  $\sigma^2$

$$E(\hat{\sigma}_{ml}^2) - \sigma^2 = -\frac{p}{n}\sigma^2 = -\frac{25}{113}\sigma^2 \quad (5.1)$$

became relatively large. The REML estimator for  $\sigma^2$  is unbiased and, in the absence of spatial autocorrelation, reduces to the usual OLS estimator.

The REML estimators of  $\theta$  were consistently biased at  $\theta = 0$  (Table 5.1). In a few of these simulations, the REML estimation optimization procedure failed to converge because the Newton-Raphson line search algorithm employed in SAS/IML (SAS Institute Inc. 1999) did not strictly adhere to the restrictions placed on the range of possible values for  $\hat{\theta}$ . As the estimate of  $\theta$  approached zero, the algorithm would sometimes attempt to evaluate the gradient or Hessian at values of  $\theta < 0$ . Also, in some simulations, the algorithm did not possess sufficient sensitivity to optimize the objective function because the log-likelihood was very flat in the neighbourhood of 0. Notwithstanding, the unit range and unidimensional nature of the objective function permit more direct search operations for any single experimental data set.

Although the differences between EMP and PRE were used to determine the validity of the analysis methods, validity is more commonly understood in terms of the distribution of the treatment  $F$ -statistics. The empirical rejection rate or size of the ANOVA/OLS within treatment class  $F$ -test was badly controlled when spatial autocorrelation was simulated (Figure 5.2). The true null hypothesis of no differences among treatments within classes was rejected by the ANOVA/OLS within-class  $F$ -test in an increasing proportion of the simulations as the value of  $\theta$  was increased, regardless of the simulation model. This result is directly related to the underestimation of the scale parameter ( $\sigma^2$ ; Table 5.1) and of EMP (Figure 5.1). These biases are also responsible for the increasing empirical rejection rates of the among-classes treatment  $F$ -test for ANOVA/OLS as the spatial autocorrelation was increased (Figure 5.2). The false null hypothesis of no differences among treatment classes was rejected more often than it should have been (given the actual treatment and correlation structures) because the variance of treatment contrasts was underestimated. Figure 5.2 should not be taken to imply that the ANOVA/OLS treatment  $F$ -tests are more powerful when data are spatially autocorrelated: the rejection rate increased with  $\theta$  because the variance estimators were increasingly biased. The rejection rate for the among-classes treatment  $F$ -test would increase regardless of the simulation structure (*i.e.*, even if  $\tau = \mathbf{0}$ ). Unfortunately, the rejection rate of the within-class treatment  $F$ -test for ANOVA/OLS was not 0.05 (the significance level used) when

$\theta = 0$  (Figure 5.2). This was an artefact of the simulation study and suggests that the number of simulations was too small.<sup>2</sup>

The empirical rejection rate of the within-classes treatment  $F$ -test for the spatial analysis methods was relatively well-controlled at around 0.05 (the significance level used; Figure 5.2). This was particularly true in the first two trials. What was more interesting was that the rejection rate of the among-classes treatment  $F$ -test for all four spatial analysis methods decreased as  $\theta$  increased. The power of the treatment  $F$ -test is generally regarded as a function of the variance ( $\sigma^2$ ); the effect size

$$\frac{\tau_{(1)} - \tau_{(t)}}{\sigma} \quad (5.2)$$

where  $\tau_{(j)}$  is the  $j$ th largest treatment effect; the sample size ( $n$ ); and the number of observations per treatment (Nemec 1991). Clearly, this is an oversimplification; all these factors were held constant in this study. First, the measure of effect size given by Equation 5.2 is too liberal when data are autocorrelated. The variability of an autocorrelated series can not be adequately described by  $\sigma^2$  alone. Secondly, as Griffiths (1992) pointed out, spatial autocorrelation impacts on the information content or dimensionality of a sample of a given size. A more positively autocorrelated sample can therefore be interpreted as having a smaller effective sample size.<sup>3</sup> This will impact on the degrees of freedom of the treatment  $F$ -test.

In the simulation study, the efficiency and validity of the spatial analysis methods were evaluated only where the true model for the spatial autocorrelation was known and exploited. In practice, it will be necessary to ascertain which model is most appropriate for a given data set. The simulation study indicated that both the restricted LR test and the AICs provide useful measures for spatial autocorrelation and model selection (Figure 5.3). However, in this simulation study, the comparison of AICs is essentially the same as the restricted LR test. The proportion of simulations in which  $AIC_{ANOVA/OLS}$  exceeded the AIC of the spatial analysis method was always at least as high as the empirical rejection rate for the restricted LR tests because of the manner in which the simulation study was constructed. When the AICs of nested models are compared, the comparison can be written as a ratio of the restricted likelihoods, that is, if

$$AIC_S - AIC_{ANOVA/OLS} \leq 0$$

<sup>2</sup>This discrepancy between nominal and empirical rejection rates may alternatively be ascribed to the imperfections of the random number generator. However, the quality of the random number generator in SAS/IML (SAS Institute Inc. 1999) has been well substantiated by Fishman and Moore (1982).

<sup>3</sup>Interestingly, the reverse is also true; negatively autocorrelated data of dimension  $n$  carry a greater information content and have an effective sample size greater than  $n$  (Griffiths 1992).

where  $AIC_S$  is the AIC of one of the spatial analysis methods used in the simulation study, then from Equation 3.64

$$-L^*(\sigma^2, \theta | \mathbf{Y}_S) + 2 + L^*(\sigma^2 | \mathbf{Y}_S, \theta = 0) - 1 \leq 0 \quad (5.3)$$

and from Equation 3.63

$$\begin{aligned} -0.5\Lambda^* + 1 &\leq 0 \\ \Lambda^* &\geq 2 \end{aligned} \quad (5.4)$$

Therefore, an AIC comparison among nested models is equivalent to a restricted LR test with a significance level of approximately 0.16. The value of the AIC lies more in its potential for comparing non-nested models for spatial autocorrelation.

Overall, the simulation study indicated that the analysis methods based on RV models were efficient and valid relative to ANOVA/OLS for the simulated Nelder experiment. The study also indicated that SAR/REML and CAR/REML were neither efficient nor valid for this same application. However, it must be emphasized that these results are conditional on the application of the correct model for spatial autocorrelation. Also, as mentioned above, the fact that the empirical rejection rates of the within treatment class  $F$ -tests for ANOVA/OLS deviated from the nominal  $\alpha$  level of 0.05 indicated that it may have been useful generate more simulations. Even with such an improvement, however, Ripley (1981; p.17) has pointed out that "a problem with using simulation to find empirical distributions is that one is usually interested in the tails of the distribution and most of the simulations provide very little relevant information." It can thus be argued that uniformity trial data would have provided a more secure basis for evaluating different analysis methods. For the purposes of this thesis, however, uniformity trial data are unobtainable; it is impossible to conduct a Nelder experiment, or any other spacing experiment, without imposing the effects of differential tree density levels. Simulation experiments are the only option available for evaluating the statistical properties of different analysis methods for Nelder experiments.

Of all the considerations that should be weighed in evaluating the results of this study, perhaps the most important is the fact that all simulations were based on a single experimental design and layout. Judge *et al.* (1985) noted that simulation results are often highly conditional on the design matrix ( $\mathbf{X}$  in Equation 3.4). This restriction in the scope of this study is of added concern here for two reasons. First, the effects of spatial autocorrelation will certainly be a function of the spatial scale of the experiment. If the simulation study were repeated with the same values of  $\theta$  but different dimensions for the simulated spacing wheel, the effects of spatial autocorrelation on the efficiency and validity of a given analysis method could be quite different. Second, Nelder

spacing wheels are relatively symmetrical, and the symmetry of a design can have a major impact on the relative efficiency of ANOVA/OLS. Prior to undertaking this study, a simple type Ia spacing wheel (Figure 2.2) was considered as a candidate experimental design for the simulations. However, it was soon discovered that as a consequence of the perfect symmetry in the type Ia (or type Ib) spacing wheel, the OLS and GLS estimators of  $\beta$  were equivalent under all the isotropic spatial CE simulation models used.<sup>4</sup>

The equivalence of the OLS and GLS estimators of  $\beta$  in symmetrical designs is due to the form of the GLS estimator. This estimator is essentially a weighted least-squares estimator. When the data are spatially autocorrelated under one of the isotropic spatial CE models described in Chapter 3, the GLS estimator weights the observations by their relative isolation; observations that have many close neighbours are given a smaller weight than observations that are relatively remote. But in a spatially symmetrical design like the type Ia spacing wheel, all observations assigned to a given treatment have exactly the same relative spatial position. Thus, all observations on a treatment level must be weighted equally and the resulting GLS estimator of  $\tau$  will be equivalent to the OLS estimator. It follows that under these circumstances the OLS estimator of  $\tau$  is efficient and attains the Cramér-Rao lower bound. The OLS estimators of variance parameters will still be biased, however.

The spacing wheels used in the MKN experiment were not perfectly symmetrical, but were still highly symmetrical relative to the designs simulated in other published studies. This may partially explain why the efficiencies of EXP/REML and GAU/REML were not found to be as large, relative to ANOVA/OLS, as those reported by Brownie and Gumpertz (1997).

### 5.2.2 MKN Analysis

Considerable variability was observed in the MKN DBH15 data set, but the data did not appear to exhibit spatial autocorrelation. Based on the empirical distribution of the ANOVA/OLS residuals (Figure 5.4), the normal distribution appeared to be a suitable distribution on which to base REML estimation of variance parameters for the spatial analysis methods.

For the MKN DBH15 data set, the REML estimation procedure for the EXP/REML and GAU/REML analysis methods failed to attain convergence. Visual inspection of the restricted log-likelihood functions confirmed that the estimates of  $\theta$  were in the neighbourhood of a global maximum. The estimates of  $\theta$  for these two spatial analysis methods were very close to zero,

---

<sup>4</sup>A formal proof of this assertion is given in Appendix B.

so the corresponding estimates for  $\sigma^2$  were nearly identical to those given by ANOVA/OLS (Table 5.3). Also, both these spatial analysis methods produced the same treatment effect rankings as ANOVA/OLS (Table 5.4). Since the estimates of spatial autocorrelation were minimal for both EXP/REML and GAU/REML, all observations received approximately the same weight (*i.e.*, as  $\hat{\theta}_{reml} \rightarrow 0$ ,  $\hat{\Omega} \rightarrow \mathbf{I}_n$ ) and  $\hat{\tau}_{reml} \approx \hat{\tau}_{ols}$ .

The REML estimation procedure did converge for the SAR/REML and CAR/REML analysis methods. Visual inspection of the restricted log-likelihood functions confirmed that the positive estimates for  $\theta$  were in the neighbourhood of a global maximum. Based on the restricted LR tests, the estimates of  $\theta$  for these two spatial analysis methods were not significantly different from zero (Table 5.3). However, the value of  $\hat{\theta}_{reml}$  for CAR/REML was sufficiently large that the resulting estimates of  $\tau$  were different from ANOVA/OLS to an extent that different overall rankings were produced (Table 5.4).

The restricted LR tests indicated that none of the spatial analysis methods proved any more useful than ANOVA/OLS in explaining the variation in this dataset (Table 5.3). Based on the minimum AIC rule, ANOVA/OLS provided the most concise stochastic description of the data (Tables 5.3).

Application of the four spatial analysis methods was a much more involved procedure than the application of ANOVA/OLS. Although the spatial coordinates of the trees in a Nelder spacing wheel are relatively easy to obtain, the construction of the spatial weight matrix ( $\mathbf{W}$ ) for NN models can be a difficult task, especially for large asymmetric designs with missing observations. Also, the spatial analysis methods require the derivation and differentiation of the appropriate restricted log-likelihood functions. Finally, although numerical search procedures are built into many statistical software packages, REML estimation will generally require some programming skills. Nonetheless, for the spatial analysis methods examined here, the actual REML estimation procedure is quite efficient in practice. There is only a single parameter estimate to search for, and a direct inspection of the result is facilitated by the restricted range for  $\hat{\theta}_{reml}$ .

## Chapter 6

# Conclusion

Spatial analysis methods are most useful for data that exhibit autocorrelation when variable values are ordered in geographic space. Classical analysis methods rely on the device of experimental randomization to neutralize the effects of autocorrelation over repeated trials. But in any single field experiment, the application of spatial analysis methods to spatially autocorrelated data can provide more efficient estimates of treatment effects.

Nelder (1962) spacing wheels are economical designs for studying tree density-growth relationships. The formal geometric structure of these designs also lends them to spatial analysis applications. In this thesis, a simulation study was used to assess the efficiency and validity of four spatial analysis methods relative to ANOVA/OLS. In the simulation study, the two spatial analysis methods based on RV models and REML estimation tended to be both valid and more efficient than ANOVA/OLS when spatial autocorrelation was simulated. The reverse was true for the two spatial analysis methods based on NN models.

The application of the spatial analysis methods to the MKN DBH15 data did not yield any additional information about the variability in the variable of interest or of the experimental material itself. Estimates of spatial autocorrelation parameters for all four spatial analysis methods were low relative to the ranges of spatial autocorrelation used in the simulation trials. However, this application did demonstrate the complexities of spatial modeling and likelihood-based estimation for field experiments with large numbers of irregularly distributed observations. These issues present new challenges, but given the power and flexibility of modern statistical software, it is certainly feasible to implement spatial analytic techniques.

Whether spatial autocorrelation will have a significant impact on the precision of treatment effect estimates, or on the accuracy of the estimates of their precision, in actual forestry Nelder experiments could not be determined. However, the levels of spatial autocorrelation estimated for

the MKN DBH15 data should not necessarily be taken as typical or representative of the levels found in a more general class of Nelder experiments. Galinski *et al.* (1994) reported significant spatial dependence in Scots pine (*Pinus sylvestris* L.) height growth patterns in a Nelder experiment in Sweden. Significant spatial autocorrelation has also been reported in agricultural trials operated at spatial scales in line with the Nelder experiments examined here (*e.g.*, Ball *et al.* 1993; Brownie *et al.* 1993). The presence and significance of spatial autocorrelation in Nelder experiments must be determined on a case-by-case basis.

This thesis has provided some indication that LR tests or information criteria may be useful tools for discerning the significance of spatial autocorrelation in a data set and for selecting an appropriate spatial CE model. Most importantly, the results of this study indicate that RV models may be the most suitable class of spatial CE models for forestry field experiments. The spatial analysis methods based on NN models studied in this thesis were neither valid nor efficient, and were more difficult to implement and interpret.

# Literature Cited

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* AC-19: 716-723.
- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer. 284 pp.
- Antle, A.N. and P.L. Marshall. 1996. Filling in missing forestry data: exploring autocorrelation techniques. *In Proceedings Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium, May 21-23, Fort Collins, CO. USDA For. Ser. Gen. Tech. Rep. RM-GTR-277. pp. 593-600.*
- Azaïs, J.-M., H. Monod, and R.A. Bailey. 1998. The influence of design on validity and efficiency of neighbour methods. *Biometrics* 54: 1374-1387.
- Ball, S.T., D.J. Mulla, and C.F. Konzak. 1993. Spatial heterogeneity affects variety trial interpretation. *Crop Sci.* 33: 931-935.
- Beckers, F. 1997. *On the Spatial and Space-Time Analysis of Field Experiments*. Ph.D. thesis, Université Catholique de Louvain, Belgique. 178 pp.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statist. Soc. B* 36: 192-236.
- Besag, J.E. and R.A. Kempton. 1986. Statistical analysis of field experiments using neighbouring plots. *Biometrics* 42: 231-251.
- Breusch, T.S. 1980. Useful invariance results for generalized regression models. *J. Econometrics* 13: 327-340.
- Brownie, C., D.T. Bowman, and J.W. Burton. 1993. Estimating spatial variation in analysis of data from yield trials: a comparison of methods. *Agron. J.* 85: 1244-1253.

- Brownie, C. and M.L. Gumpertz. 1997. Validity of spatial analyses for large field trials. *J. Agric. Biol. Environ. Stat.* 2:1-23.
- Casella G. and R.L. Berger. 1990. *Statistical Inference*. Belmont, CA: Duxbury. 650 pp.
- Chacko, V.J. 1965. Designs for forestry spacing and thinning experiments. *In Proceedings of the Conference of Advisory Group of Forest Statisticians of the International Union of Forest Research Organisations Section 25, 27th Sept. to 1st Oct., Stockholm, Sweden. Paper No.5.*
- Chung, K.L. 1974. *A Course in Probability Theory*. 2nd ed. New York: Academic Press. 365 pp.
- Cliff, A.D. and J.K. Ord. 1981. *Spatial Processes: Models and Applications*. London: Pion. 266 pp.
- Cole, E.C. and M. Newton. 1987. Fifth-year response of Douglas-fir to crowding and nonconiferous competition. *Can. J. For. Res.* 17: 181-186.
- Cressie, N. 1990. The origins of kriging. *Math. Geol.* 22: 239-252.
- Cressie, N.A.C. 1993. *Statistics for Spatial Data*. New York: Wiley. 900 pp.
- Cressie, N. and S.N. Lahiri. 1996. Asymptotics for REML estimation of spatial covariance parameters. *J. Statist. Plann. Inference* 50: 327-341.
- Dietrich, C.R. 1991. Modality of the restricted likelihood for spatial Gaussian random fields. *Biometrika* 78: 833-839.
- Evert, F. 1971. Spacing studies - A review. Forest Management Institute Information Report FMR-X-37. Canadian Forestry Service: Ottawa. 95 pp.
- Evert, F. 1984. An update (1970/71-82) of the annotated bibliography on initial tree spacing. Petawawa National Forestry Institute Information Report PI-X-44. Canadian Forestry Service: Chalk River, Ontario. 157 pp.
- Fishman, G.S. and L.R. Moore. 1982. A statistical evaluation of multiplicative congruential random number generators with modulus  $2^{31} - 1$ . *J. Amer. Statist. Assoc.* 77: 129-136.
- Freeman, G.H. 1964. The use of a systematic design for a spacing trial with a tropical tree crop. *Biometrics* 20: 200-203.

- Galinski, W., J. Witowski, and M. Zwieniecki. 1994. Non-random height pattern formation in even aged scots pine (*Pinus sylvestris* L.) Nelder plots as affected by spacing and site quality. *Forestry* 67: 49-61.
- Gould, P. 1970. Is *statistix inferens* the geographical name for a wild goose? *Econ. Geog.* 46: 439-448.
- Griffith, D.A. 1992. What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *L'Espace Géographique* 3: 265-280.
- Hansen, M.H., W.G. Madow, and B.J. Tepping. 1983. An evaluation of model-dependent and probability sampling inferences in sample surveys. [With comments and rejoinder.] *J. Amer. Statist. Assoc.* 78: 776-807.
- Harville, D.H. 1974. Bayesian inference for variance components using only error constraints. *Biometrika* 61: 383-385.
- Harville, D.H. 1975. Experimental randomization: Who needs it? *Amer. Statist.* 29: 27-31.
- Harville, D.H. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* 72: 320-338.
- Hooper, P.M. 1989. Experimental randomization and the validity of normal-theory inference. *J. Amer. Statist. Assoc.* 84: 576-587.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl, T.-C. Lee. 1985. *The Theory and Practice of Econometrics*. 2nd ed. New York: Wiley. 1019 pp.
- Kenward, M.G. and J.H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53: 983-997.
- Loo-Dinkins, J.A. and C.G. Tauer. 1987. Statistical efficiency of six progeny test field designs on three loblolly pine (*Pinus taeda* L.) site types. *Can. J. For. Res.* 17: 1066-1070.
- Magnussen, S. 1993. Bias in genetic variance estimates due to spatial autocorrelation. *Theoret. App. Gen.* 86: 349-355.
- Milne, A. 1959. The centric systematic area-sample treated as a random sample. *Biometrics* 15: 270-297.

- Mora, A. 1996. Método de ajuste de un modelo rendimiento-densidad para observaciones provenientes del diseño sistemático de Nelder. [In Spanish.] *Rev. Facultad de Agronomía* 22: 31-36.
- Namkoong, G. 1966. Applications of Nelder's designs in tree improvement research. *In Proceedings 8th Southern Conference on Forest Tree Improvement, June 16-17, 1965, Savannah, Georgia.* pp. 24-37.
- Nelder, J.A. 1962. New kinds of systematic designs for spacing experiments. *Biometrics* 18: 283-307.
- Nemec, A.F.L. 1991. Power analysis handbook for the design and analysis of forestry trials. B.C. Ministry of Forests Biometrics Information Handbook 2. Victoria, B.C.: Crown Publications. 26 pp.
- Nyland, R.D. 1996. *Silviculture: Concepts and Applications.* New York: McGraw-Hill. 633 pp.
- Paltzat, R. 1984. Performance of red alder in Nelder plot designs at age 7 at the U.B.C. Research Forest. B.S.F. thesis, University of British Columbia, Vancouver, B.C. 45 pp.
- Papadakis, J.S. 1937. Méthode statistique pour des expériences sur champ. [In French.] *Bull. de l'Institut d'Amélioration des Plantes à Salonique* 23: 13-27.
- Patterson, H.D. and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545-554.
- Patterson, H.D. and R. Thompson. 1974. Maximum likelihood estimation of components of variance. *In Proceedings of the 8th International Biometrics Conference, Aug. 25-30, Constanta, Romania.* pp. 197-207.
- Radosevich, S.R. 1987. Methods to study interactions among crops and weeds. *Weed Tech.* 1: 190-198.
- Reukema, D.L. 1979. Fifty-year development of Douglas-fir stands planted at various spacings. *USDA For. Serv. Pap. PNW-253.* 21 pp.
- Ripley, B.D. 1981. *Spatial Statistics.* New York: Wiley. 252 pp.
- Rustagi, K.P. 1984. A modified Nelder design to achieve square or triangular spacing. *Indian Forester* 110: 521-528.

- SAS Institute Inc. 1999. SAS OnlineDoc. V.8. SAS Institute Inc.: Cary, NC.
- Smith, J.H.G. 1959. Comprehensive and economical designs for studies of spacing and thinning. *For. Sci.* 5: 237-245.
- Smith, J.H.G. 1978. Design factors from Nelder and other spacing trials to age 20. *Com. For. Rev.* 57: 109-119.
- Student. 1914. The Elimination of Spurious Correlation due to Position in Time or Space. *Biometrika* 10: 179-180.
- Student. 1937. Comparison between balanced and random arrangements of field plots. *Biometrika* 29: 363-379.
- Thorntott, M.L. 1982. The role of randomization in model-based inference. *Aust. J. Stat.* 24: 137-145.
- Toumey, J.W. and C.F. Korstian. 1943. *Seeding and Planting in the Practice of Forestry*. 3rd ed. Wiley, New York. 520 pp.
- Wall, M.M. 2000. A close look at the spatial correlation structure implied by the CAR and SAR models. Res. Rep. No. 2000-022, Division of Biostatistics, University of Minnesota, Minneapolis, MN. 16 pp.
- Wackernagel, H. 1998. *Multivariate Geostatistics*. 2nd ed. Berlin: Springer. 291 pp.
- Whittle, P. 1954. On stationary processes in the plane. *Biometrika* 41: 434-449.
- Zedaker, S.H. 1982. Growth and development of young Douglas-fir in relation to intra- and inter-specific competition. Ph.D. thesis, Oregon State University, Corvallis, Oregon. 175 pp.
- Zimmerman, D.L, and D.A. Harville. 1991. A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47: 223-239.
- Zyskind, G. 1967. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Stat.* 38: 1092-1109.

## APPENDICES

# Appendix A

## REML Estimation

Consider the statistical model for the  $n \times 1$  random vector  $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\text{A.1})$$

where  $\mathbf{X}$  is a  $n \times p$  matrix of full column rank,  $\boldsymbol{\varepsilon}$  is a  $n \times 1$  vector of random errors,  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$ , and  $\mathbf{V}$  is a  $n \times n$  positive definite matrix. Further, let  $\mathbf{V} = \sigma^2 \boldsymbol{\Omega}$  where  $\sigma^2$  is a general scale parameter ( $\sigma^2 > 0$ ) and  $\boldsymbol{\Omega} \equiv \boldsymbol{\Omega}(\theta)$  is a  $n \times n$  positive definite matrix known up to a parameter  $\theta$ .

The restricted log-likelihood is

$$L^*(\sigma^2, \theta | \mathbf{Y}) = c - \underbrace{\frac{(n-p)}{2} \ln \sigma^2}_A - \underbrace{\frac{1}{2} \ln |\boldsymbol{\Omega}|}_B - \underbrace{\frac{1}{2} \ln |\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}|}_C - \underbrace{\frac{1}{2\sigma^2} \mathbf{Y}^T \mathbf{Q} \mathbf{Y}}_D \quad (\text{A.2})$$

where  $\mathbf{Q} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1}$ ,  $c$  is a constant term, and  $L^*(\cdot)$  is divided into the components  $A$ ,  $B$ ,  $C$ , and  $D$ , for ease of reference below (Harville 1974; Section 3.3.2).

### Differentiation of the Restricted Log-Likelihood

The first two partial derivatives of the restricted log-likelihood with respect to the scale parameter ( $\sigma^2$ ) are

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} L^*(\sigma^2, \theta | \mathbf{Y}) &= -\frac{\partial}{\partial \sigma^2} A - \frac{\partial}{\partial \sigma^2} D \\ &= -\frac{(n-p)}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbf{Y}^T \mathbf{Q} \mathbf{Y} \end{aligned} \quad (\text{A.3})$$

and

$$\frac{\partial^2}{\partial \sigma^4} L^*(\sigma^2, \theta | \mathbf{Y}) = \frac{(n-p)}{2\sigma^4} - \frac{1}{\sigma^6} \mathbf{Y}^T \mathbf{Q} \mathbf{Y} \quad (\text{A.4})$$

Differentiation of  $L^*(\cdot)$  with respect to the covariance parameter ( $\theta$ ) requires partial differentiation of the inverse and determinant of the matrix  $\boldsymbol{\Omega}$ . The subsequent derivations make use of the following results

1.  $\frac{\partial}{\partial \theta} \Omega^{-1} = -\Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1}$
2.  $\frac{\partial}{\partial \theta} \ln |\Omega| = \text{tr} \left[ \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \right]$
3.  $\frac{\partial}{\partial \theta} \mathbf{Y}^T \mathbf{Q} \mathbf{Y} = -\mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \Omega \right) \mathbf{Q} \mathbf{Y}$

the latter being due to Beckers (1997). Making use of these results, the first partial derivative of  $L^*(\cdot)$  with respect to  $\theta$  can be obtained from

$$\frac{\partial}{\partial \theta} A = 0$$

$$\frac{\partial}{\partial \theta} B = \frac{\partial}{\partial \theta} \left[ \frac{1}{2} \ln |\Omega| \right] = \frac{1}{2} \text{tr} \left[ \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \right]$$

$$\begin{aligned} \frac{\partial}{\partial \theta} C &= \frac{\partial}{\partial \theta} \left[ \frac{1}{2} \ln |\mathbf{X}^T \Omega^{-1} \mathbf{X}| \right] \\ &= \frac{1}{2} \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \frac{\partial}{\partial \theta} \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right) \right] \\ &= -\frac{1}{2} \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] \end{aligned}$$

$$\frac{\partial}{\partial \theta} D = \frac{\partial}{\partial \theta} \left[ \frac{1}{2\sigma^2} \mathbf{Y}^T \mathbf{Q} \mathbf{Y} \right] = -\frac{1}{2\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \Omega \right) \mathbf{Q} \mathbf{Y}$$

The first partial derivative of the  $L^*(\cdot)$  with respect to  $\theta$  is then

$$\begin{aligned} \frac{\partial}{\partial \theta} L^*(\sigma^2, \theta | \mathbf{Y}) &= -\frac{1}{2} \text{tr} \left[ \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \right] \\ &\quad + \frac{1}{2} \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] \\ &\quad + \frac{1}{2\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \Omega \right) \mathbf{Q} \mathbf{Y} \end{aligned} \quad (\text{A.5})$$

The second partial derivative of the restricted likelihood with respect to  $\theta$  can be obtained from

$$\begin{aligned} \frac{\partial}{\partial \theta} \text{tr} \left[ \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \right] &= -\text{tr} \left[ \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \right] + \text{tr} \left[ \Omega^{-1} \left( \frac{\partial^2}{\partial \theta^2} \Omega \right) \right] \\ \frac{\partial}{\partial \theta} \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] &= \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] \\ &\quad - 2 \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] \\ &\quad + \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial^2}{\partial \theta^2} \Omega \right) \Omega^{-1} \mathbf{X} \right] \\ &= \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] \\ &\quad - 2 \text{tr} \left[ \left( \mathbf{X}^T \Omega^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \left( \frac{\partial}{\partial \theta} \Omega \right) \Omega^{-1} \mathbf{X} \right] \end{aligned}$$

$$\begin{aligned}
& + \text{tr} \left[ \left( \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \left( \frac{\partial^2}{\partial \theta^2} \boldsymbol{\Omega} \right) \boldsymbol{\Omega}^{-1} \mathbf{X} \right] \\
& = -\text{tr} \left[ \left( \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \boldsymbol{\Omega}^{-1} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \boldsymbol{\Omega}^{-1} \mathbf{X} \right] \\
& \quad + \text{tr} \left[ \left( \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \left( \frac{\partial^2}{\partial \theta^2} \boldsymbol{\Omega} \right) \boldsymbol{\Omega}^{-1} \mathbf{X} \right]
\end{aligned}$$

$$\frac{\partial}{\partial \theta} \frac{1}{\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \mathbf{Q} \mathbf{Y} = -\frac{2}{\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \mathbf{Q} \mathbf{Y} + \frac{1}{\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial^2}{\partial \theta^2} \boldsymbol{\Omega} \right) \mathbf{Q} \mathbf{Y}$$

Combining these three results according to Equation A.5, the second partial derivative of the restricted log-likelihood function with respect to  $\theta$  is

$$\begin{aligned}
\frac{\partial^2}{\partial \theta^2} L^*(\sigma^2, \theta | \mathbf{Y}) & = -\frac{1}{\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \mathbf{Q} \mathbf{Y} + \frac{1}{2\sigma^2} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial^2}{\partial \theta^2} \boldsymbol{\Omega} \right) \mathbf{Q} \mathbf{Y} \\
& \quad + \frac{1}{2} \text{tr} \left[ \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \boldsymbol{\Omega}^{-1} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \right] \\
& \quad - \frac{1}{2} \text{tr} \left[ \mathbf{Q} \left( \frac{\partial^2}{\partial \theta^2} \boldsymbol{\Omega} \right) \right]
\end{aligned} \tag{A.6}$$

Finally, the second partial derivative of the  $L^*(\cdot)$  with respect to both  $\sigma^2$  and  $\theta$  can be obtained from partial differentiation of Equation A.3 with respect to  $\theta$

$$\begin{aligned}
\frac{\partial^2}{\partial \sigma^2 \partial \theta} L^*(\sigma^2, \theta | \mathbf{Y}) & = \frac{\partial}{\partial \theta} \frac{1}{2\sigma^4} \mathbf{Y}^T \mathbf{Q} \mathbf{Y} \\
& = -\frac{1}{2\sigma^4} \mathbf{Y}^T \mathbf{Q} \left( \frac{\partial}{\partial \theta} \boldsymbol{\Omega} \right) \mathbf{Q} \mathbf{Y}
\end{aligned} \tag{A.7}$$

### Asymptotic Covariance Matrix

The asymptotic covariance matrix of the REML estimators of  $\sigma^2$  and  $\theta$  is given by the inverse of the information matrix

$$\mathcal{J}(\sigma^2, \theta) = -\mathbf{E} \begin{bmatrix} \frac{\partial^2}{\partial \sigma^4} L^*(\sigma^2, \theta) & \frac{\partial^2}{\partial \sigma^2 \partial \theta} L^*(\sigma^2, \theta) \\ \frac{\partial^2}{\partial \sigma^2 \partial \theta} L^*(\sigma^2, \theta) & \frac{\partial^2}{\partial \theta^2} L^*(\sigma^2, \theta) \end{bmatrix} \tag{A.8}$$

(Cressie and Lahiri 1996). The elements in the right-hand side of Equation A.8 are given by

$$\begin{aligned}
-\mathbf{E} \left( \frac{\partial^2}{\partial \sigma^4} L^*(\sigma^2, \theta) \right) & = -\frac{(n-p)}{2\sigma^4} + \mathbf{E} \left( \frac{1}{\sigma^6} \mathbf{Y}^T \mathbf{Q} \mathbf{Y} \right) \\
& = -\frac{(n-p)}{2\sigma^4} + \frac{1}{\sigma^6} \mathbf{E} \left( \text{tr} \left[ \mathbf{Q} \mathbf{Y} \mathbf{Y}^T \right] \right) \\
& = -\frac{(n-p)}{2\sigma^4} + \frac{1}{\sigma^6} \left( \sigma^2 \text{tr} \left[ \mathbf{Q} \boldsymbol{\Omega} \right] + \text{tr} \left[ \mathbf{Q} \mathbf{X} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{X}^T \right] \right) \\
& = -\frac{(n-p)}{2\sigma^4} + \frac{1}{\sigma^4} (n-p) + \frac{1}{\sigma^6} (0) \\
& = \frac{(n-p)}{2\sigma^4}
\end{aligned}$$



## Appendix B

# On the Equivalence of OLS and GLS

Consider the general linear model of the  $n \times 1$  response vector ( $\mathbf{Y}_S$ ) for a Nelder experiment consisting of a single type Ia spacing wheel

$$\mathbf{Y}_S = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_S \quad (\text{B.1})$$

where  $\mathbf{S}$  is a  $n \times 2$  matrix that defines the relative spatial locations of the trees,  $\mathbf{X}$  is a  $n \times p$  design matrix,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of unknown parameters,  $\boldsymbol{\varepsilon}_S \sim N(\mathbf{0}_{n \times 1}, \mathbf{V})$  is a  $n \times 1$  vector of random errors and  $\mathbf{V}$  is an  $n \times n$  positive definite matrix. The relative spatial location of the  $i$ th tree is completely determined by the radius of its arc ( $r_k; k = 1, 2, \dots, t$ ) and the angle of its spoke ( $\varpi_l; l = 1, 2, \dots, m$ ), that is,

$$\mathbf{s}_i = \begin{bmatrix} r_k \\ \varpi_l \end{bmatrix} = \begin{bmatrix} r_k \\ l\varpi \end{bmatrix} \quad (\text{B.2})$$

where  $\mathbf{s}_i^T$  is the  $i$ th row of  $\mathbf{S}$  ( $i = 1, 2, \dots, n$ ) and  $\varpi$  is the angle between successive spokes in the design.

**Theorem 1** *If the random vector  $\boldsymbol{\varepsilon}_S$  for a Nelder experiment consisting of a single complete type Ia spacing wheel with  $m$  spokes and  $t$  experimental arcs is defined as an isotropic process, then the ordinary least squares (OLS) and generalized least squares (GLS) estimators of  $\boldsymbol{\beta}$ ,*

$$\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_S \quad (\text{B.3})$$

and

$$\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}_S \quad (\text{B.4})$$

respectively, are equivalent, and  $\hat{\boldsymbol{\beta}}_{ols}$  is the minimum-variance unbiased estimator of  $\boldsymbol{\beta}$ .

Zyskind (1967) outlined the general conditions under which OLS and GLS estimators of  $\beta$  are equivalent. In particular, Zyskind demonstrated that a necessary and sufficient condition for the equivalence of the OLS and GLS estimators is that  $\mathbf{V}$  can be decomposed as

$$\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{D} \quad (\text{B.5})$$

where  $\mathbf{X}$  is the  $n \times p$  design matrix and  $\mathbf{D}$  is any  $p \times p$  matrix. It is not required that  $\mathbf{X}$  be of full column rank. The following two lemmas will be used to show that the identity defined by Equation B.5 applies under the conditions of Theorem 1.

**Lemma 1:** If  $\varepsilon_{\mathbf{S}}$  is an isotropic stochastic process, then the elements of  $\mathbf{V} = \{v_{i,j}\}$ , arranged lexicographically by spoke and arc, are constrained by the relation

$$\sum_{j=1}^n v_{(h+(a-1)t),j} = \sum_{j=1}^n v_{h,j} \quad (\text{B.6})$$

for all  $a = 1, 2, \dots, m$ ; and  $h = 1, 2, \dots, t$  (i.e.,  $(h + (a - 1)t) = 1, 2, \dots, n$ ).

Proof:

$$\begin{aligned} \sum_{j=1}^n v_{(h+(a-1)t),j} &= \sum_{j=1}^n f \left( \left\| \mathbf{s}_{(h+(a-1)t)} - \mathbf{s}_j \right\| \right) \\ &= \sum_{k=1}^t \sum_{l=1}^m f \left( \left\| \begin{bmatrix} r_h \\ \varpi_a \end{bmatrix} - \begin{bmatrix} r_k \\ \varpi_l \end{bmatrix} \right\| \right) \\ &= \sum_{k=1}^t \sum_{l=1}^m f \left( r_h^2 + r_k^2 - 2r_h r_k \cos(a\varpi - l\varpi) \right) \\ &= \sum_{k=1}^t \sum_{c=1}^m f \left( r_h^2 + r_k^2 - 2r_h r_k \cos(\varpi - c\varpi) \right) \quad [c = 1 + l - a] \\ &= \sum_{k=1}^t \sum_{c=1}^m f \left( \left\| \begin{bmatrix} r_h \\ \varpi_1 \end{bmatrix} - \begin{bmatrix} r_k \\ \varpi_c \end{bmatrix} \right\| \right) \\ &= \sum_{j=1}^n f \left( \left\| \mathbf{s}_h - \mathbf{s}_j \right\| \right) \\ &= \sum_{j=1}^n v_{h,j} \end{aligned}$$

where the first equality follows from the condition of isotropy.

**Lemma 2:** If  $\varepsilon_{\mathbf{S}}$  is an isotropic stochastic process, then the elements of  $\mathbf{V} = \{v_{i,j}\}$ , arranged lexicographically by spoke and arc, are constrained by the relation

$$\sum_{l=1}^m v_{(h+(a-1)t),(h'+(l-1)t)} = \sum_{l=1}^m v_{h,(h'+(l-1)t)} \quad (\text{B.7})$$

for all  $a = 1, 2, \dots, m$ ;  $h = 1, 2, \dots, t$ ; and  $h' = 1, 2, \dots, t$ .

Proof:

$$\begin{aligned}
\sum_{l=1}^m v_{(h+(a-1)t), (h'+(l-1)t)} &= \sum_{l=1}^m f \left( \left\| \mathbf{s}_{(h+(a-1)t)} - \mathbf{s}_{(h'+(l-1)t)} \right\| \right) \\
&= \sum_{l=1}^m f \left( \left\| \begin{bmatrix} r_h \\ \varpi_a \end{bmatrix} - \begin{bmatrix} r_{h'} \\ \varpi_l \end{bmatrix} \right\| \right) \\
&= \sum_{l=1}^m f \left( r_h^2 + r_{h'}^2 - 2r_h r_{h'} \cos(a\varpi - l\varpi) \right) \\
&= \sum_{c=1}^m f \left( r_h^2 + r_{h'}^2 - 2r_h r_{h'} \cos(\varpi - c\varpi) \right) \quad [c = 1 + l - a] \\
&= \sum_{c=1}^m f \left( \left\| \begin{bmatrix} r_h \\ \varpi_1 \end{bmatrix} - \begin{bmatrix} r_{h'} \\ \varpi_c \end{bmatrix} \right\| \right) \\
&= \sum_{c=1}^m f \left( \left\| \mathbf{s}_h - \mathbf{s}_{(h'+(c-1)t)} \right\| \right) \\
&= \sum_{c=1}^m v_{h, (h'+(c-1)t)}
\end{aligned}$$

from which Equation B.7 follows.

It is now possible to show that the conditions of Theorem 1 imply the identity given by Equation B.5. Since  $\mathbf{X}$  need not be of full column rank, it is possible to write

$$\mathbf{VX} = \mathbf{V} [[\mathbf{1}_{t \times 1} \mathbf{I}_t] \otimes \mathbf{1}_{m \times 1}] \quad (\text{B.8})$$

where  $\otimes$  is the direct (Kronecker) product operator. Therefore,

$$\mathbf{VX} = \begin{bmatrix} \sum_{j=1}^n v_{1,j} & \sum_{r=1}^m v_{1,1+(r-1)t} & \cdots & \sum_{r=1}^m v_{1,t+(r-1)t} \\ \sum_{j=1}^n v_{2,j} & \sum_{r=1}^m v_{2,1+(r-1)t} & \cdots & \sum_{r=1}^m v_{2,t+(r-1)t} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n v_{n,j} & \sum_{r=1}^m v_{n,1+(r-1)t} & \cdots & \sum_{r=1}^m v_{n,t+(r-1)t} \end{bmatrix} \quad (\text{B.9})$$

$$= \begin{bmatrix} a_1 & b_{1,1} & \cdots & b_{1,t} \\ a_2 & b_{2,1} & \cdots & b_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_n & b_{n,1} & \cdots & b_{n,t} \end{bmatrix} \quad (\text{B.10})$$

From lemmas 1 and 2, respectively,  $a_h = a_{h+(l-1)t}$  and  $b_{(h+(l-1)t),(h'+(l-1)t)} = b_{h,(h'+(l-1)t)}$  ( $h = 1, 2, \dots, t; h' = 1, 2, \dots, t; l = 1, 2, \dots, m$ ). Thus,

$$\mathbf{VX} = \begin{bmatrix} a_1 & b_{1,1} & \cdots & b_{1,t} \\ a_2 & b_{2,1} & \cdots & b_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_t & b_{t,1} & \cdots & b_{t,t} \end{bmatrix} \otimes \mathbf{1}_{(m \times 1)} \quad (\text{B.11})$$

$$= \begin{bmatrix} \mathbf{1}_{(t \times 1)} \mathbf{I}_t & \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_1 & b_{1,1} & \cdots & b_{1,t} \\ a_2 & b_{2,1} & \cdots & b_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_t & b_{t,1} & \cdots & b_{t,t} \end{bmatrix} \end{bmatrix} \otimes \mathbf{1}_{(m \times 1)} \quad (\text{B.12})$$

$$= \left[ \left[ \mathbf{1}_{(t \times 1)} \mathbf{I}_t \right] \otimes \mathbf{1}_{(m \times 1)} \right] \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_1 & b_{1,1} & \cdots & b_{1,t} \\ a_2 & b_{2,1} & \cdots & b_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_t & b_{t,1} & \cdots & b_{t,t} \end{bmatrix} \quad (\text{B.13})$$

$$= \mathbf{XD} \quad (\text{B.14})$$

Since  $\mathbf{VX} = \mathbf{XD}$  is a necessary and sufficient condition for the equivalence of the OLS and GLS estimators of  $\beta$  (Zyskind 1967), Theorem 1 is proved.

The two lemmas define a type of symmetry that is implied by the condition of isotropy assumed in Theorem 1; the elements of the rows of  $\mathbf{V}$  corresponding to same arc, but different spokes, are simply permutations of one another. This symmetry in  $\mathbf{V}$  is a consequence of the symmetry in the treatment structure in the type Ia design. This symmetry, and the equivalence of the OLS and GLS estimators of  $\beta$ , is lost if trees have died or observations are missing from the analysis.

Theorem 1 can also be generalized to Nelder experiments consisting of several type Ia spacing wheels, or combinations of several type Ia and type Ib spacing wheels, if differences among spacing wheels are modelled as fixed effects.