INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

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

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

SPATIAL STATISTICS USING QUASI-LIKELIHOOD METHODS WITH APPLICATIONS

By

DAVID M. DOLAN, M.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

SPATIAL STATISTICS USING QUASI-LIKELIHOOD METHODS

Doctor of Philosophy (1999)
(Mathematics)

McMaster University
Hamilton, Ontario

TITLE: Spatial Statistics using Quasi-Likelihood Methods with Applications

AUTHOR: David M. Dolan, M.A. (University of Michigan)

SUPERVISOR: Professor A.H. El-Shaarawi

NUMBER OF PAGES: xvi, 217

ABSTRACT

Spatial Statistics have been applied to many types of problems in the environmental sciences, mostly dealing with continuously distributed data from Gaussian or near-Gaussian processes. There is a need for methods capable of handling discrete, non-Gaussian data, such as species counts from biological processes. This thesis applies the method of quasi-likelihood from general linear models to the problem of spatial prediction of benthic invertebrate counts. These organisms are important elements of the aquatic food chain and are indicators of pollutant impacts. Predictions of their abundance are needed as clean up targets in areas where remedial actions are being considered. The proposed method is illustrated using an example data set from Great Lakes reference sites.

The applicability of the method is first illustrated by re-analysis of examples from the literature. Variogram models are fitted to quasi-likelihood residuals with two alternative distance metrics. The models are compared using cross-validation and predictions are made using the classical estimator of the variogram and distance determined from a Geographic Information System (GIS). Asymptotic normality of quasi-likelihood parameter estimates is shown to

hold when spatial dependence is accounted for by an exponential variogram model. A brief simulation study is included that verifies the applicability of asymptotic results to the estimation of model parameters.

ACKNOWLEDGEMENTS

The author wishes to thank Professor A.H. El-Shaarawi for his role as Supervisory Committee Chair and, in general, for fostering interest in environmental statistics. The author is also grateful to Professors R. Viveros and H. Heinig who served on the Supervisory Committee. Thanks go to T. B. Reynoldson and K. E. Day, of Environment Canada, for the use of their data for the example application in this thesis. Thanks also go to S. Hedtke and S. Lozano, of the U. S. Environmental Protection Agency, who made it possible for the author to continue his education. Finally, the author wishes to thank his wife, Mary Ginnebaugh, for encouragement and support during the long process of completion of the dissertation.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

TABLE OF CONTENTS (CONT'D)

TABLE OF CONTENTS (CONT'D)

LIST OF FIGURES

LIST OF FIGURES (CONT'D)

LIST OF FIGURES (CONT'D)

LIST OF TABLES

LIST OF TABLES (CONT'D)

LIST OF TABLES (CONT'D)

LIST OF TABLES (CONT'D)

CHAPTER 1: INTRODUCTION

1.1 Introduction

Geostatistics has been developed for use in the fields of mining, metallurgical engineering and geology to estimate unknown quantities of ore and to predict likely locations of mineral deposits. The basic approach is to take advantage of the correlation structure among observations distributed in space to perform estimation, prediction, hypothesis testing and other statistical procedures. The similarity of these applications to some agricultural, sociological and environmental applications is obvious, especially for large, continuous resources such as lakes and forests. In fact, Matheron (1963) defines Geostatistics as: "The application of the formalism of random functions to the reconnaissance and estimation of natural phenomena." Many environmental problems would benefit from a geostatistical treatment.

In the Great Lakes, one of the most serious problems is the contaminated sediment that lies below the water in many polluted areas, especially in harbours and bays. When this material is dredged for navigational purposes, it must be "confined" because it is often toxic to aquatic life or contains

persistent chemicals that bioaccumulate throughout the food chain. Confinement facilities are expensive, usually only temporary, and often create environmental problems of their own. In areas where navigational dredging does not occur, contaminated sediments often do not provide the quality of habitat required for a healthy aquatic ecosystem. These sediments may need to be removed, isolated or treated *in situ*, depending on the results of assessment procedures.

Current assessment procedures for sediment quality are chemically based, ie. they prescribe a lowest effect level for each chemical found in the sediment that must not be exceeded. However, natural variations of a geological and biological nature confound these procedures, leading to non-optimal decisions. Recently, advances (Bailey, *et al*, 1995; Moss, *et al*, 1987) have been made in biologically based assessment procedures that avoid some of these pitfalls. The basic concern of these procedures is biological response, not chemical concentrations. A problem with this approach is that biological responses have inherent spatial variability due to habitat changes and therefore generic guidelines cannot be established. Methods for establishing site-specific guidelines are critical if biological responses are to be used as sediment criteria.

Reynoldson *et al.* (1995) have taken over 300 sediment samples from the Great Lakes nearshore areas and have analyzed them biologically as well as

chemically. All of the samples have been taken from "clean areas", ie. those relatively unimpacted by human activity. It is intended that this data set be used as a training data set to predict what type of biological community should be found in an area assuming it to be clean. This can then be used as an assessment tool (Corkum, 1989; Johnson and Wiederholm, 1989; Wright *et al.*, 1984; Zarull and Reynoldson, 1992) to identify which areas require remedial action. Alternatively, the data set could be used as the nominal condition, against which other areas are compared when reporting the state of the sediment habitat. While this is a promising approach, the significance of differences between observed and predicted biological communities is an area that requires further development. The communities could be represented in ordination space and compared using multivariate statistical methods. Current multivariate methods use discriminant function analysis to select site specific responses. This requires groups to be formed that represent biological communities. However, while statistically different, these groups may simply represent response continua and therefore alternate methods that do not require groups may be preferable.

Preliminary analysis of this data set indicates strong spatial clustering (Bailey, *et al.* 1995), that may be related to natural phenomena such as geology, climate, and hydrology. They identified three distinct benthic communities based on Euclidian distances calculated from mean densities of 15 major community

members. Sites that were geographically "close" were often also "close" in ordination space. In order to take advantage of all of the information contained in a sample, as well as make the best prediction possible, the methods of spatial statistics should be explored as a possible approach for analysis of these data.

1.2 The Problem

The problem posed here is that of using the methods of spatial statistics to predict the value of a count variable given a location and a set of explanatory variables. Methods exist for prediction of continuous variables with known distributions using Geostatistics, but this work extends existing methodology to the case of a spatially autocorrelated regression model with a discrete variable with unknown distribution. This work is applied in nature: examples from the literature are re-analyzed using new combinations of existing methods. The example described in Section 1.4 is then analyzed using these procedures.

One of the major issues in spatial statistics is the estimation of the covariance function. This is made considerably easier when the data are multivariate normal, but some data of interest are not distributed in this way. In particular, binary responses and counts often occur in the biological,

environmental and health sciences. Methods (e.g. autologistic, autobinomial, auto-Poisson and auto-Pascal models) have been developed that take into account various dependencies in the data, with different assumptions about the underlying distribution of the data (Haining, 1990, p. 99-101). Each assumption has its drawbacks and the degree of success achieved depends on the nature of the problem. The use of quasi-likelihood methods avoids assumptions about the distribution of the data, but these methods have not been used frequently with spatial statistics.

1.3 The Approach

A general review of quasi-likelihood methods, spatial statistics and relevant environmental applications from the literature is presented in Chapter 2. Existing methods and studies that are directly related to the current work are reviewed in subsequent chapters.

As will be discussed in Chapter 2, Zeger (1988) presented an approach to modelling dependency in time-series data of counts. He used quasi-likelihood methods and estimating equations based on the score function to solve for unknown coefficients for the explanatory variables and account for time-dependency. The approach here will be to extend Zeger's work to the case of

spatial dependence. There are many obvious analogies between time-series and spatial statistics. If time-series data are thought of as unidirectional, where the value of a random variable at time t is some function of a random shock and all previous observations of the variable, then spatial data may be thought of as multidirectional, where the value of the random variable at a site **s** is some function of a random shock and all other observations. Zeger accounted for time dependency through the variance-covariance matrix in the score function. There is no reason why this matrix could not be structured to account for spatial dependency. This is the topic of Chapter 3.

The identification of the form of the dependency is a major aspect of this work. Since the large-scale variation is modelled by a function of explanatory variables, the spatial dependency will be among residuals. Various parameterizations of the spatial covariance function are examined and procedures for estimation are employed. Chapter 4 contains this material.

Because the goal of this work is to identify optimal prediction equations, the different types of kriging that permit accurate projections from spatial data are discussed and evaluated relative to example data sets in Chapter 5. Cross-validation of different models is used to select the "best" model and predictions are made for two examples.

Finally, simulations are performed that check the validity of the results and the applicability of inferences based on quasi-likelihood. This is the topic of Chapter 6.

1.4 Example - Benthic Counts at Great Lakes Reference Sites

Reynoldson *et al.* (1995) have reported the initial results of the extensive sampling of the sediments of nearshore areas of the Great Lakes. The purpose of their study was to develop sediment quality criteria based on ecological response. Stratified samples were taken at reference sites throughout the Great Lakes Basin. The sites were stratified by ecoregions which are regions of similar climate, vegetation, bedrock geology, flora and fauna. The reference sites were taken to be representative of "unpolluted conditions" and were selected according to the following rules:

1) further than 10 km away from known discharges
2) less than 2 km from shore
3) less than 30 metres in depth
4) known or suspected to have fine-grain substrate

The resultant sites for the Great Lakes are shown in Figure 1.1. Various physical

Reference Sites

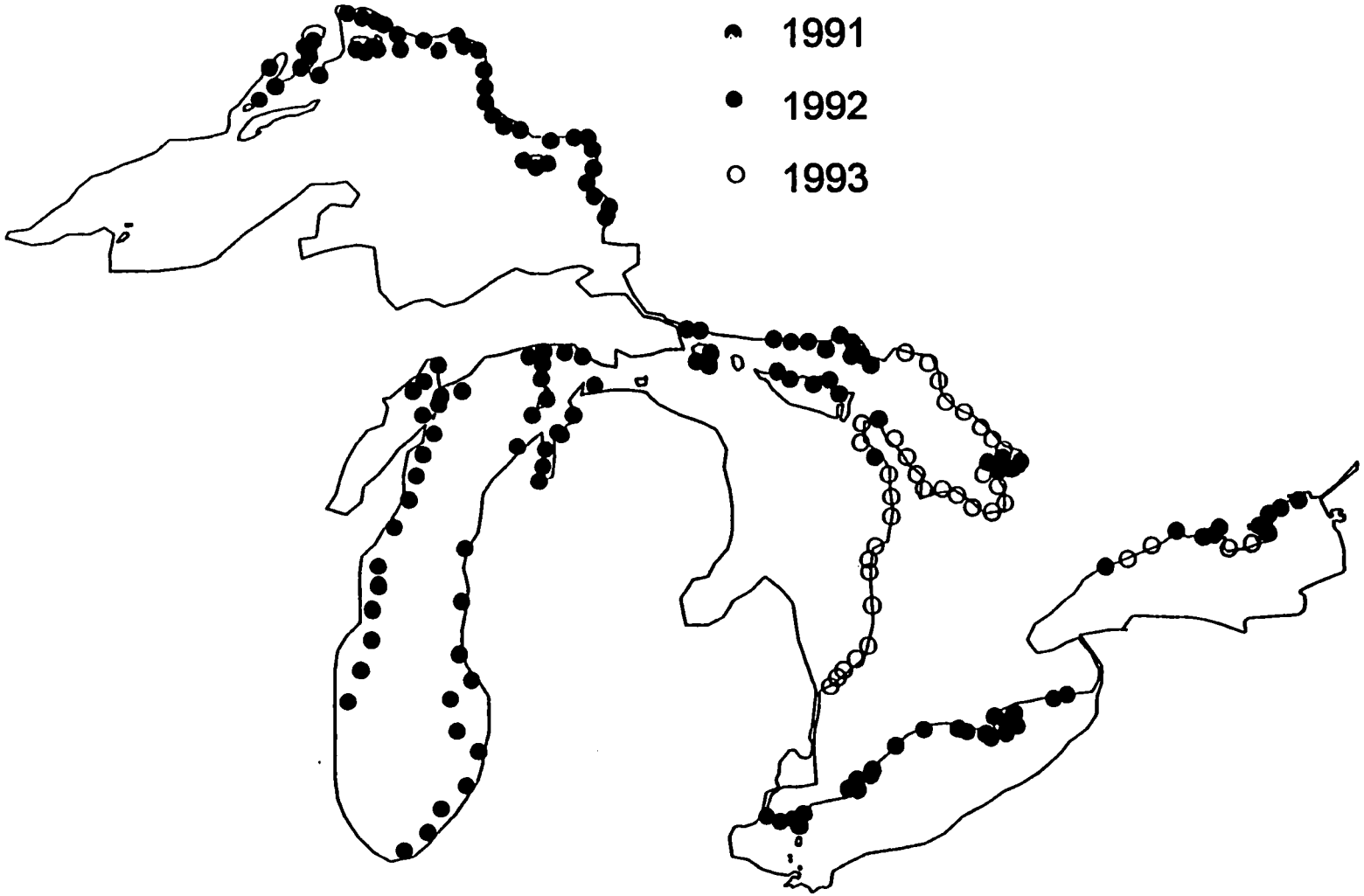

FIGURE 1.1 Great Lakes Reference Sites

and chemical measurements were made on samples from a box-core or a "mini-box core" which are devices for taking up to a 50 cm by 50 cm section of the sediment at the site to a depth of about 25 to 30 cm. The material in the box core was then subsampled using five small plexiglass tubes (inner diameter 6 cm) to obtain organisms for benthic community structure evaluations. A separate sampling device, called a mini-ponar, was used to collect five replicate sediment samples for laboratory bioassays. Samples for analysis of nutrients in the overlying water were also taken.

Benthic community structure was reported by mean species count per site. Typically, only those taxa that made up 0.05% or more of the sample were included in subsequent analyses. *Stylodrilus herringianus, Diporeia hoyi, Procladius sp.* and *Limnodrilus hoffmeisteri* were among the more common and abundant species. Forty-three environmental variables were measured on the sediment and overlying water including nutrients, metals, and physical parameters of particles (Table 1.1). The authors of this work have made their entire data set available for use in this dissertation; however, only two benthic species are examined in detail and the geographical area has been restricted to Georgian Bay and the North Channel of Lake Huron.

TABLE 1.1

Data Available from Example Problem

SITE INFORMATION:

Location information (latitude, longitude), sample dates, and ecodistrict.

FIELD DATA:

Water depth, field measurements, such as temperature, dissolved oxygen and conductivity.

WATER CHEMISTRY:

Alkalinity, nutrients (including nitrogen and phosphorous).

SEDIMENT CHEMISTRY:

Minerals, metals, percent gravel, sand, silt, and clay, and organic content.

BENTHIC INVERTEBRATE COUNTS:

ARTHROPODS - 80 Species

MOLLUSKS - 93 Species

FAMILY - Arthropods and Mollusks aggregated to 45 families

ORDER - Arthropods and Mollusks aggregated to 18 orders.

The rationale for the choice of geographic area and benthic species to model is included in Chapter 2. *Diporeia hoyi* are important indicators of oligotrophic conditions in lakes and are a key link in the aquatic food chain. They make up as much as 60% of the benthic biomass in Lake Huron. The Great Lakes Water Quality Agreement includes a quantitative objective for *Diporeia hoyi* abundance in Lake Superior. *Procladius* are ubiquitous in most areas of the Great Lakes, but are sensitive to pollution.

CHAPTER 2: LITERATURE REVIEW

2.1 Quasi-likelihood Methods

2.1.1 Definition

McCullagh and Nelder (1983, Ch. 1) describe how Gauss abandoned the assumption of normally distributed errors in the classical regression model in favor of the weaker assumption of constant variance only. They then describe how this weaker assumption was in turn generalized by Wedderburn (1974) to the case of generalized linear models using the concept of quasi-likelihood. In standard likelihood methods, the form of the distribution of the random variable must be specified. In the quasi-likelihood method, a relationship between the mean and variance of the random variable is all that need be specified. The resultant quasi-likelihood function can then be used for estimation in exactly the same manner as the log-likelihood function.

The quasi-likelihood function, $K(z_i, \mu_i)$, for each observation, z_i ($i = 1,\ldots,n$), is defined by the relation:

$$\partial K(z_i, \mu_i)/\partial \mu_i = (z_i - \mu_i)/V(\mu_i) \qquad (2.1)$$

where z_i are independent observations

μ_i are the expectations $E(Z_i)$. Each μ_i is a known, differentiable function of a set of parameters $\beta_1,\ldots,\beta_p$

and $V(\mu_i)$ is the finite variance $Var(Z_i)$.

2.1.2 Properties of Quasi-likelihoods

THEOREM 1. (Wedderburn, 1974) Let z, μ, $V(\mu)$ and K be defined as above (with the subscripts dropped). For each observation, μ is some known function of a set of parameters $\beta_1,\ldots,\beta_p$. Then K has the following properties:

(i) $E(\partial K/\partial\mu) = 0$

(ii) $E(\partial K/\partial\beta_j) = 0$

(iii) $E(\partial K/\partial\mu)^2 = - E(\partial^2 K/\partial\mu^2) = 1/V(\mu)$

(iv) $E((\partial K/\partial\beta_j)(\partial K/\partial\beta_k)) = - E(\partial^2 K/\partial\beta_j\partial\beta_k)$

$= (1/V(\mu))(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_k)$.

These properties are similar to those of log likelihoods. The theorem is proved below, following the proof sketched by Wedderburn.

Proof. (i) From the definition of K,

$E((Z-\mu)/ V(\mu)) = E(Z / V(\mu)) - E(\mu / V(\mu)) = 0$

(ii) Noting that $\partial K/\partial\beta_j = (\partial K/\partial\mu)(\partial\mu/\partial\beta_j)$,

$$E((\partial K/\partial\mu)(\partial\mu/\partial\beta_j)) = \partial\mu/\partial\beta_j \; E(\partial K/\partial\mu) = 0$$

(iii) A special case of (iv). See below.

(iv) $$E((\partial K/\partial\beta_j)(\partial K/\partial\beta_k)) = E\{((\partial K/\partial\mu)(\partial\mu/\partial\beta_j))((\partial K/\partial\mu)(\partial\mu/\partial\beta_k))\}$$

$$= E\{(\partial K/\partial\mu)^2\}(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_k)$$

$$= E\{((Z-\mu)/V(\mu))^2\}(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_k)$$

Since $V(\mu) = Var(Z)$, $$= (1/V(\mu))(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_k).$$

Also, $$- E(\partial^2 K/\partial\beta_j\partial\beta_k) = - E\{\partial/\partial\beta_k((\partial K/\partial\mu)(\partial\mu/\partial\beta_j))\}$$

$$= - E\{(Z-\mu)\partial/\partial\beta_k((1/V(\mu))(\partial\mu/\partial\beta_j)\}$$

$$- E\{(1/V(\mu))(\partial\mu/\partial\beta_j)(-\partial\mu/\partial\beta_k)\}$$

$$= 0 + (1/V(\mu))(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_k).$$

To prove (iii), $E(\partial K/\partial\mu)^2 = E((\partial K/\partial\beta_j)/(\partial\mu/\partial\beta_j))^2$ as in (ii), and apply (iv) with i=j.

COROLLARY. If the distribution of z is specified in terms of μ, so that the log likelihood, l, can be defined (for each observation), then

$$- E(\partial^2 K/\partial\mu^2) \leq - E(\partial^2 l/\partial\mu^2).$$

Proof. From (iii), $- E(\partial^2 K/\partial\mu^2) = 1/V(\mu) = 1/Var(Z)$.

From the Cramer-Rao inequality,

$$\text{Var}(Z) \geq -1/E(\partial^2 l/\partial\mu^2).$$

So, $-1/E(\partial^2 K/\partial\mu^2) = \text{Var}(Z) \geq -1/E(\partial^2 l/\partial\mu^2)$

and $-E(\partial^2 K/\partial\mu^2) \leq -E(\partial^2 l/\partial\mu^2)$.

Wedderburn (1974) has also shown that the above inequality becomes an equality when the observations are distributed as a one-parameter exponential family. Further, he argues that $-E(\partial^2 K/\partial\mu^2)$ is a measure of the information z gives concerning μ when only the mean-variance relationship is known and that $E(\partial^2 (K-l)/\partial\mu^2)$ is a measure of the additional information provided by knowing the distribution of Z.

For a one-parameter exponential family, the log-likelihood is the same as the quasi-likelihood. To see this, write the log-likelihood for one observation from a distribution belonging to a one-parameter exponential family:

$$l(\mu(\beta),z) = z\theta - g(\theta) + f(z)$$

where θ is some function $\theta(\mu)$.

Differentiate with respect to μ, then:

$$\partial l(\mu(\beta),z)/\partial\mu = (z - g'(\theta))(d\theta/d\mu)$$

$$= (z - \mu)/V(\mu)$$

$$= \partial K(z,\mu)/\partial\mu$$

since $\mu = g'(\theta)$ and $d\theta/d\mu = 1/V(\mu)$. Hence the log-likelihood and the quasi-likelihood are the same. Therefore a one-parameter exponential family can be assumed as a weak sort of distributional assumption. The variance, $Var(Z_i)$, of each observation, z_i, is assumed to be either equal to, or proportional to, some known function of its expectation, μ_i, ie. $Var(Z_i) \propto V(\mu)$.

2.1.3 Score Function of Quasi-likelihood

The score function, **U**, for quasi-likelihood is defined analogously to that of likelihood with more than one observation (ie. it is the first partial derivative of the log quasi-likelihood with respect to β):

$$U(\beta) = \sum_{t=1}^{n} (\partial\mu_t/\partial\beta)(z_t - \mu_t)/V(\mu_t)$$

where the conditions of equation 2.1 apply to each observation, z_t. The parameter estimates, $\hat{\beta}$, are the solutions of the set of equations resulting when the expectation of the score function is set equal to zero. McCullagh (1983) shows that parameter estimates based on this function are consistent, asymptotically Gaussian and optimal in an extended Gauss-Markov sense. First, the expectation, $E(\hat{\beta}-\beta)$ is shown to be of order n^{-1}. This satisfies the definition of consistency since the expectation tends to zero as n increases to infinity, as

does the variance, Var($\hat{\beta}$). Next, McCullagh shows that the quantity $\sqrt{n}(\hat{\beta}-\beta)$ has an asymptotic normal distribution as n tends to infinity. Finally, the parameter estimates are shown to be asymptotically optimal in that they are unbiased and have uniformly minimum variance as n increases to infinity. This is an extension of the Gauss-Markov theorem (Harvey, 1981, pp. 44-46) under the assumption of local instead of global linearity and asymptotic instead of exact unbiasedness.

2.1.4 Quasi-likelihood Function for Dependent Observations

Zeger (1988) generalized the score function of the quasi-likelihood to the dependent case:

$$\mathbf{U}(\beta) = (\partial\boldsymbol{\mu}/\partial\beta)\, \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}\, (\mathbf{z} - \boldsymbol{\mu}) = 0$$

where $\hat{\theta}$ are $\sqrt{n}$-consistent estimates of the parameters of the variance-covariance matrix (These are treated separately and not estimated with weighted least squares)

$\mathbf{z}$ and $\boldsymbol{\mu}$ are now vectors of the observations and means, respectively,

and $\mathbf{V}$ is the variance-covariance matrix of the observations.

For the independent case, $\mathbf{V}$ is a diagonal matrix. The dependent case can include additional parameters in the off-diagonal elements that describe the

covariance of the observations (see Section 2.2.5 for the conditions for positive definiteness of **V**). Zeger developed the method for autocorrelated counts.

Autocorrelation can be introduced to the model either by observation-driven or parameter-driven log-linear models (Cox, 1981). In the observation-driven case, the conditional expectation is $E(Z_t|Z_{t-1}) = f(z_{t-1}, z_{t-2}, \ldots, z_1)$, where *f* is a function of all previous observations (Zeger and Qaqish, 1988). A special case is $E(Z_t|Z_{t-1}) = f(z_{t-1})$ where $\{Z_t\}$ is a Markov process. The autoregressive time series models described in the next section are also examples of observation-driven models.

In the parameter-driven case, a latent process, ϵ_t, is used to introduce autocorrelation:

$$\theta_t = \log \mu_t = \theta(\epsilon_t, \mu_{t-1}, \ldots, \mu_1).$$

A special case is $\theta_t = \theta(\epsilon_t, \mu_{t-1})$ where $\{\mu_t\}$ is now a Markov process.

Zeger gives the first two marginal moments of a time series z_t that is Poisson, given the latent process ϵ_t. The derivations follow:

Conditional on ϵ_t, let z_t be a sequence of independent counts with

mean, $u_t = E(Z_t|\epsilon_t) = \exp(\mathbf{x}_t\beta)\epsilon_t$ and variance, $w_t = Var(Z_t|\epsilon_t) = u_t$. Assume that ϵ_t is an unobserved stationary process with $E(\epsilon_t) = 1$ and $Cov(\epsilon_t, \epsilon_{t+\tau}) = \sigma^2\rho_\epsilon(\tau)$. Then the marginal moments of z_t are found by:

$$\begin{aligned}\mu_t = E(Z_t) &= E[\,E(Z_t|\epsilon_t)] \\ &= \int_{-\infty}^{\infty} \exp(\mathbf{x}_t\beta)\epsilon_t\, g(\epsilon_t)\, d(\epsilon_t) \\ &= \exp(\mathbf{x}_t\beta) \int_{-\infty}^{\infty} \epsilon_t\, g(\epsilon_t)\, d(\epsilon_t) \\ &= \exp(\mathbf{x}_t\beta)\, E(\epsilon_t) = \exp(\mathbf{x}_t\beta),\end{aligned}$$

where $g(\epsilon_t)$ is the marginal density of ϵ_t.

Also by Woodroofe (1975, p. 282):

$$\begin{aligned}v_t = Var(Z_t) &= E[Var(Z_t|\epsilon_t)] + Var[E(Z_t|\epsilon_t)] \\ &= \mu_t + Var[\mu_t\epsilon_t] \\ &= \mu_t + \mu_t^2 Var[\epsilon_t] \\ &= \mu_t + \mu_t^2\sigma^2 = \mu_t(1 + \sigma^2\mu_t), \text{ and}\end{aligned}$$

$$\begin{aligned}\rho_z(t,\tau) &= corr(Z_t, Z_{t+\tau}) \\ &= Cov(Z_t, Z_{t+\tau})/((Var(Z_t)Var(Z_{t+\tau}))^{1/2} \\ &= \{E[Cov(Z_t, Z_{t+\tau}|\epsilon_t, \epsilon_{t+\tau}] + \\ &\qquad Cov[E(Z_t|\epsilon_t), E(Z_{t+\tau}|\epsilon_{t+\tau})]\}/((Var(Z_t)Var(Z_{t+\tau}))^{1/2} \\ &= \{0 + Cov[\mu_t\epsilon_t, \mu_{t+\tau}\epsilon_{t+\tau}]\}/(Var(Z_t)Var(Z_{t+\tau}))^{1/2} \\ &= \{\mu_t\mu_{t+\tau}\, Cov[\epsilon_t, \epsilon_{t+\tau}]\}/(Var(Z_t)Var(Z_{t+\tau}))^{1/2}\end{aligned}$$

$$= \{\mu_t \mu_{t+\tau} \sigma^2 \rho_e(\tau)\} / (\mathrm{Var}(Z_t)\mathrm{Var}(Z_{t+\tau}))^{1/2}$$

$$= \{\rho_e(\tau)\} / (\{1 + [\sigma^2 \mu_t]^{-1}\}\{1 + [\sigma^2 \mu_{t+\tau}]^{-1}\})^{1/2}.$$

Zeger then uses an estimating equation approach analogous to quasi-likelihood to estimate β and applies it to the example of a time series of counts of polio cases. Since **V** can be quite large its inversion may be a problem and hence Zeger gives approximate methods for estimation of β and shows via simulation that they are adequate.

2.1.5 Asymptotic Properties of Quasi-likelihood Parameters

McCullagh (1983) has shown that the estimates of quasi-likelihood parameters are asymptotically Gaussian when the observations are independent (see Section 2.1.3). Moore (1986) generalized this result for the case of overdispersed counts and proportions. He assumed a model similar to that of McCullagh (equation 2.1), but allowed for overdispersion:

$$E(Z_i) = \mu_i \,, \; \mathrm{var}(Z_i) = \nu_1(\mu_i, n_i) + \sigma^2 \nu_2(\mu_i, n_i)$$

where ν_1 and ν_2 are functions of the means μ_i and possibly of n_i,

and $g(\boldsymbol{\mu}) = \eta = \mathbf{X}\beta$, where g is the link function.

Moore used the Lindeberg Central Limit Theorem and the Inverse Function

Theorem (see Foutz, 1977) to prove that $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2)$ estimated from quasi-likelihood equations is asymptotically multivariate normal. He also stated a corollary that showed that the asymptotic covariance of $\hat{\beta}$ is not affected by the estimation of $\hat{\sigma}^2$.

Liang and Zeger (1986) and Zeger and Liang (1986) extended the estimating equation approach to longitudinal data analysis. They showed that the parameter estimates from the generalized estimating equations, which are closely related to quasi-likelihood, are asymptotically Gaussian (see Section 6.2). Kaufmann (1987) generalized regression models for stochastically independent categorical observations to the case of nonstationary categorical time series. He presented theorems that showed the asymptotic properties of the maximum likelihood estimator. Zeger and Qaqish (1988) used the results of Kaufmann to establish the asymptotic normality of estimators for parameters in quasi-likelihood Markov models. Zeger (1988) developed a parameter-driven regression model for time series of counts (see section 2.1.4). He showed that the parameter estimates from quasi-likelihood estimating equations were asymptotically multivariate Gaussian. The generalized estimating equation approach of Liang and Zeger was extended by Prentice (1988) to allow joint estimation of parameters of both the marginal response probabilities and the pairwise correlation in correlated binary regression models. He showed that the

joint asymptotic distribution of these parameters was multivariate Gaussian. Thall and Vail (1990) presented a family of covariance models for longitudinal counts with predictive covariates. Their philosophy of parameter estimation was to obtain a unified distribution theory for regression parameters and covariance parameters as Moore (1986) and Prentice (1988) did. They showed that the $\hat{\theta}$ from Moore (above) will be asymptotically normal when the estimating equations for $\hat{\beta}$ are augmented by a second set of moment equations for the parameters $\hat{\alpha}$ that arise in the formulation of the covariance matrix of the observations (ie. $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$ is jointly multivariate Gaussian).

2.2 Spatial Statistics

2.2.1 Introduction

Since its beginnings in the early 1960s, Geostatistics has evolved into the field of Spatial Statistics. With the availability of powerful computers for simulation and graphical display, the field has seen applications in many other disciplines besides geology. According to Ripley (1981, p. 1), Spatial Statistics includes spatial data analysis and confirmatory data analysis. Spatial data analysis is "the reduction of spatial patterns to a few clear and useful summaries." Confirmatory data analysis goes beyond this by comparing these

summaries to what might be expected from theories of how the patterns might have developed and originated.

As mentioned in Chapter 1, strong spatial clustering has been observed in the example data set. Basically, samples that are "neighbours" geographically tend to cluster together. Although cluster analysis does not use geographic distance explicitly, there is a strong connection with Spatial Statistics through the spatial autocorrelation function. In the univariate case, define the "squared distance", d_{ij}^2, between two sites i and j as:

$$d_{ij}^2 = \sum_{k=1}^{n} (x_{ik} - x_{jk})^2 / n$$

where x_{ik} is the kth sample value at the ith site

and n is the number of samples taken at each site.

This distance is determined in relation to the metric of the sample observation, ie. # of organisms per square metre or milligrams of contaminant per litre of water. If X_i has mean μ_i and variance σ_i^2, and if the covariance of X_i and X_j is denoted by σ_{ij}, d_{ij}^2 is an estimate of $E[(X_i - X_j)^2]$, where

$$E[(X_i - X_j)^2] = E[X^2_i - 2X_iX_j + X^2_j]$$

$$= Var(X_i) + [E(X_i)]^2 - 2Cov(X_i,X_j)$$

$$-2E(X_i)E(X_j) + Var(X_j) + [E(X_j)]^2$$

$$= \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij} + (\mu_i - \mu_j)^2. \qquad (2.2)$$

To see this, expand d_{ij}^2 and compare terms with (equation 2.2).

In cluster analysis, sites with small d_{ij} cluster together, usually forming a cluster if d_{ij} is less than some minimum D_0. However, a small d_{ij} requires that both σ_{ij} be large and that $(\mu_i-\mu_j)^2$ be small. Some of the techniques of Spatial Statistics, discussed in Section 2.2.5, can be used to estimate σ_{ij}, the spatial covariance function, or ρ_{ij}, the spatial autocorrelation function. These two functions are related by:

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}.$$

Agterberg (1970), describes a typical geostatistical prediction problem that can be solved by a method known as kriging, named after D. G. Krige, a South African mining engineer. If P_i, i = 1, ..., n, are irregularly distributed points in a geographic area and values x_i, i = 1, ..., n, for a given attribute (i.e., data from samples) are known for the points, and if P_0 is a point with arbitrary

coordinates, the problem is to predict a value $\hat{x}_0$ for the point P_0 from the known values in the "neighbourhood". The method of kriging estimates the value of x_0 at the point P_0 by:

$$\hat{x}_0 = \mu_0 + \sum_{i=1}^{n} \hat{a}_{0,i}(x_i - \mu_i),$$

where the coefficients, $\hat{a}_{0,i}$, are given by:

$$\begin{bmatrix} \hat{a}_{0,1} \\ \hat{a}_{0,2} \\ \cdot \\ \cdot \\ \hat{a}_{0,n} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdot & \cdot & \cdot & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \cdot & \cdot & \cdot & \rho_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdot & \cdot & \cdot & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{01} \\ \rho_{02} \\ \cdot \\ \cdot \\ \rho_{0n} \end{bmatrix},$$

and μ_0 is the expected value of the attribute at point P_0,

μ_i is the value of the mean of the process at point P_i,

and ρ_{ij} is the spatial autocorrelation function discussed above.

This method is known as simple kriging (Matheron, 1971) and requires that the mean function, μ_i and ρ_{ij} be known. If the data, x_i, are Gaussian, then $\hat{x}_0$ will be the optimal predictor if the criterion is minimization of $E(x_0 - \hat{x}_0)^2$.

Cressie (1993, p. 106) defines kriging as "a minimum-mean-squared-error method of spatial prediction that (usually) depends on the second-order properties of the process $Z(\cdot)$", where $Z(\cdot)$ is a random process. There are other types of kriging, such as ordinary kriging and universal kriging. These are described in more detail in Section 2.2.6.

2.2.2 Analogy with Time Series

In time series analysis, the two most basic models are the first order autoregressive model, AR(1), and the first order moving average, MA(1). These models may be compared with their spatial counterparts, simultaneous autoregressive (SAR) and spatial moving average (MA) as described by Haining (1990, pp. 81-84):

AR(1) $\quad Z_t = \mu + \phi(Z_{t-1} - \mu) + \epsilon_t$

SAR $\quad Z_i = \mu_i + \sum_{j=1}^{n} \phi_{ij}(Z_j - \mu_j) + \epsilon_i$ (2.3)

MA(1) $\quad Z_t = \mu + \epsilon_t - \theta\epsilon_{t-1}$

Spatial MA $Z_i = \mu_i + \sum_{j=1}^{n} \theta_{ij}\epsilon_j + \epsilon_i$

where $Z_{(\cdot)}$ are random variables at time t=1,..., or site i=1,..., from a stationary (in time) stochastic process,

μ is the mean of the process,

$\mu_{(\cdot)}$ is the mean of the spatial process at a site

ϕ is the autoregressive parameter,

ϕ_{ij} are spatial autoregressive parameter with i and j indexing sites that interact,

θ is the moving average parameter,

θ_{ij} are spatial moving average parameters with i and j indexing sites that interact,

$\epsilon_{(\cdot)}$ are random shocks at time t=1,..., or at sites i=1,..., from a white noise process

and n is the number of neighbouring observations that influence Z_i.

The spatial models presented here are for two dimensions, but can be generalized to higher dimensions. If the sites are on a regular grid or lattice, distances among sites will be uniform and the idea of lags from time series can be extended to the spatial case. For example, in time series, observations z_t and z_{t-1} are said to be one lag apart. If observations z_i and z_j are neighbours on a spatial grid, they are said to be one lag apart.

2.2.3 The Variogram

A principal tool in describing spatial autodependence in Geostatistics is the variogram. Matheron (1963) describes the variogram in detail. It is a curve in R^2 which displays the mean of the square of differences between the same characteristic (ie. temperature, concentration, number of organisms) at two points a distance, r, from each other versus that distance. The variogram and its parameterization is an increasing function of distance since, in general, the further apart the two samples are, the greater is likely to be their difference. The variogram, or more conveniently, the semivariogram, $\gamma(r)$, is related to the covariance function (ACF or correlogram if standardized by dividing by the variance of the process), at least for stationary random fields, by the following relationship:

$$C(r) = \sigma^2 - \gamma(r)$$

where σ^2 is the variance

$\gamma(r)$ is the semivariogram

and $C(r)$ is the covariance function (see Section 2.2.5).

To see this, let $Z(\mathbf{s})$ and $Z(\mathbf{u})$ be two realizations of a process with the same mean determined at locations **s** and **u**, a distance, r, apart. By the definition of the variogram,

$$2\gamma(r) = \mathrm{Var}\{Z(\mathbf{s}) - Z(\mathbf{u})\}$$

$$= \mathrm{Var}\{Z(\mathbf{s})\} + \mathrm{Var}\{Z(\mathbf{u})\} - 2\mathrm{Cov}\{Z(\mathbf{s}),Z(\mathbf{u})\}$$

$$= 2\sigma^2 - 2C(r) \qquad (2.4)$$

or $C(r) = \sigma^2 - \gamma(r)$.

Cox (1981) has noted that the sample estimate of the variogram may be more useful than the corresponding sample estimate for the covariance function in time series. One of the advantages is that a linear trend added to a stationary series barely affects the variogram for small lags (distances) whereas large shifts can occur in an autocorrelation plot. Upton and Fingleton (1985, Ch. 3) use correlograms to identify spatial autocorrelation. As mentioned above, these are tools from time series extended to the spatial case. Cressie (1993, p. 73) notes that despite possible advantages, statisticians probably do not use the variogram because of lack of familiarity.

Monestiez, *et al.* (1990), describe the classical definition of three parts of the semivariogram curve (Figure 2.1):

1) discontinuity at the origin. This is often described as the "nugget effect" and is discussed below.

2) monotonic increase with distance. This means that the closer two samples are, the more likely they are to have similar values (ie. dependence).

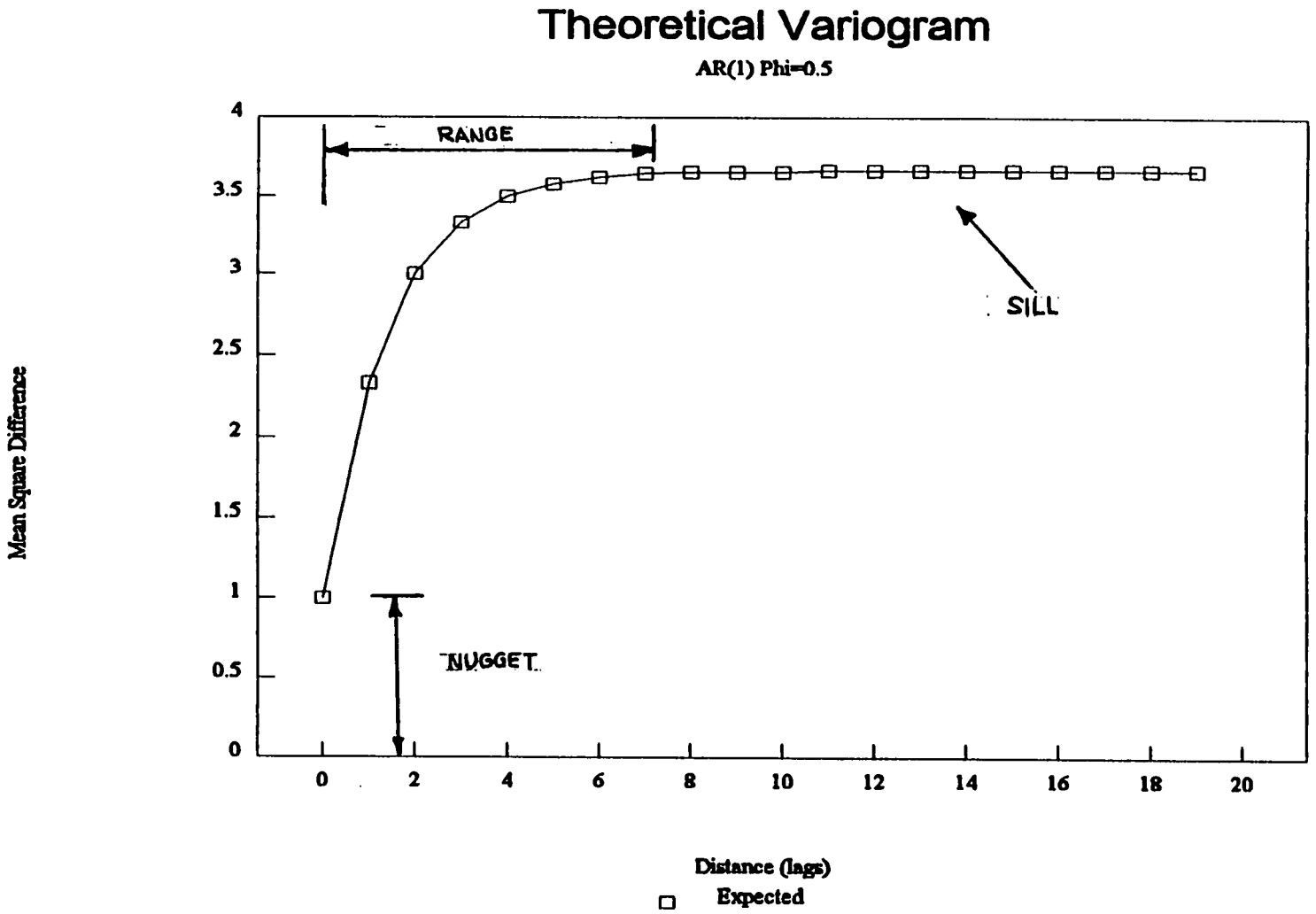

FIGURE 2.1 Theoretical Variogram

3) the steady value or sill that is reached by the function. The value of the sill is just $2\sigma^2$ (or just σ^2 for the semivariogram) which is a consequence of the covariance going to zero at greater distances (equation 2.4).

Christensen (1991, pp. 274-275) has shown that the "nugget effect", or non-zero variance at the origin of the semivariogram, can be reproduced by a measurement error model. Thus for each realization of a second order stationary process, the value, Z(**s**), is made up of three components:

$$Z(\mathbf{s}) = \mu + e(\mathbf{s}) + e_M(\mathbf{s})$$

where μ is the mean value of the process

$e(\mathbf{s})$ is the stochastic error process

and $e_M(\mathbf{s})$ is the measurement error process.

This measurement error is the only process active when r=0 on the semivariogram and so $\gamma(0) = \sigma_m^2$, the measurement error variance.

Monestiez, *et al.* (1990), note that the value of r in the semivariogram at which the function reaches the sill, or steady value, is called the range and is basically the limit of spatial dependence for the process. For higher values of r, the sill can fluctuate. This may be due to a non-regular phenomenon at larger distances or lower accuracy in the estimation of mean square difference due to a

smaller number of pairs at large distances. In time series analysis a rule of thumb is to look at only the first 25% of lags (Box and Jenkins, 1970, p. 33). For spatial autocorrelation, this could be modified by defining the maximum distance on the x-axis of the semivariogram to be 25% of the maximum possible distance. Journel and Huijbregts (1978, p. 194) recommend that no more than 50% of the maximum possible distance be used depending on the number of pairs available.

2.2.4 AR(1) Variograms

Continuing the analogy with time series, comparisons between simulated covariance functions and semivariograms are now made. Abraham and Ledolter (1983, p. 195) give the sample autocorrelation function (SACF) as:

$$\hat{\rho}(r)= \sum_{t=r+1}^{n} (z_t - \bar{z})(z_{t-r} - \bar{z}) / \sum_{t=1}^{n} (z_t - \bar{z})^2 \qquad r = 0,1,2,...$$

Also, the variance of these sample autocorrelations can be estimated by:

$$\mathrm{Var}(\hat{\rho}_r) \approx (1 + 2\hat{\rho}_1^2 + ... + 2\hat{\rho}_q^2)/n$$

where q is the lag at which $\rho_r = 0$ for $r > q$.

The classical estimator of the semivariogram, from Matheron (1962, as cited by Cressie, 1993, p. 69), for one dimension and in the notation of time series is:

$$2\hat{\gamma}(r) = (1/n(r)) \sum_{n(r)} (z_t - z_{t-r})^2 \qquad (2.5)$$

where n(r) is the number of pairs of **z** that have lag r.

The expected semivariogram for an AR(1) model with ϕ=0.5 is shown in Figure 2.2, along with the results of three simulations. Figure 2.3 shows the same model with a Poisson error term. Note that the expected semivariogram is the same as for the normal case, but the variability is greater.

2.2.5 The Spatial Covariance Function

Ripley (1981, p. 10) defines the covariance between Z(**s**) and Z(**u**) for two points **s** and **u** by:

$$C(\mathbf{s},\mathbf{u}) = E[\{Z(\mathbf{s}) - E(Z(\mathbf{s}))\}\{Z(\mathbf{u}) - E(Z(\mathbf{u}))\}].$$

Further, the stochastic process that generates Z is considered to be stationary under translations or homogeneous if C depends only on the vector $\mathbf{h} = \mathbf{u}-\mathbf{s}$. If C depends only on the distance d(**s**,**u**) between **s** and **u**, then the process is isotropic, or stationary under rotations about the origin. Homogeneous covariance functions are identified by C(**h**), while isotropic covariance functions are identified by C(r), where r = d(**s**,**u**).

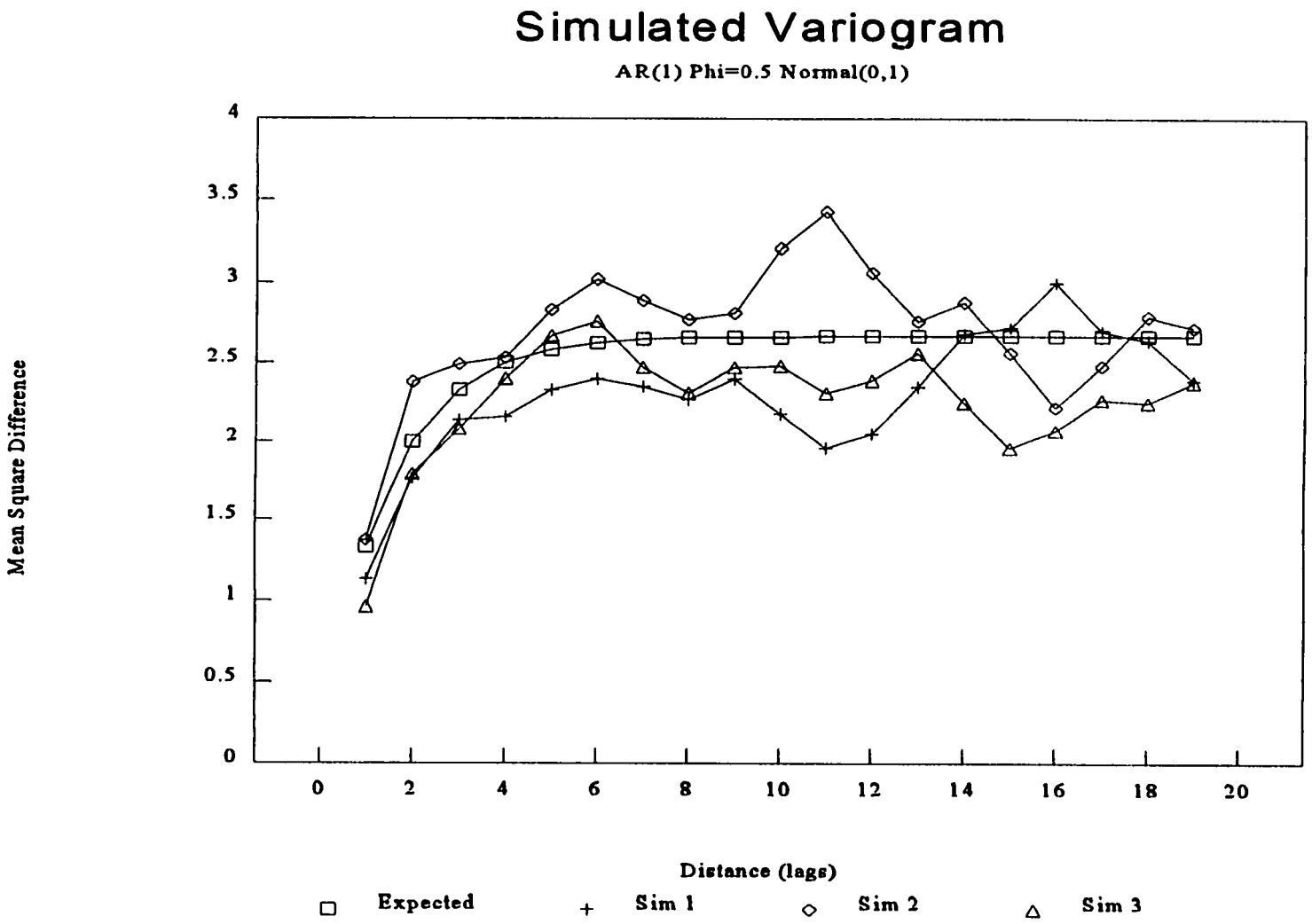

FIGURE 2.2 Expected Variogram and Three Simulations (Normal Case)

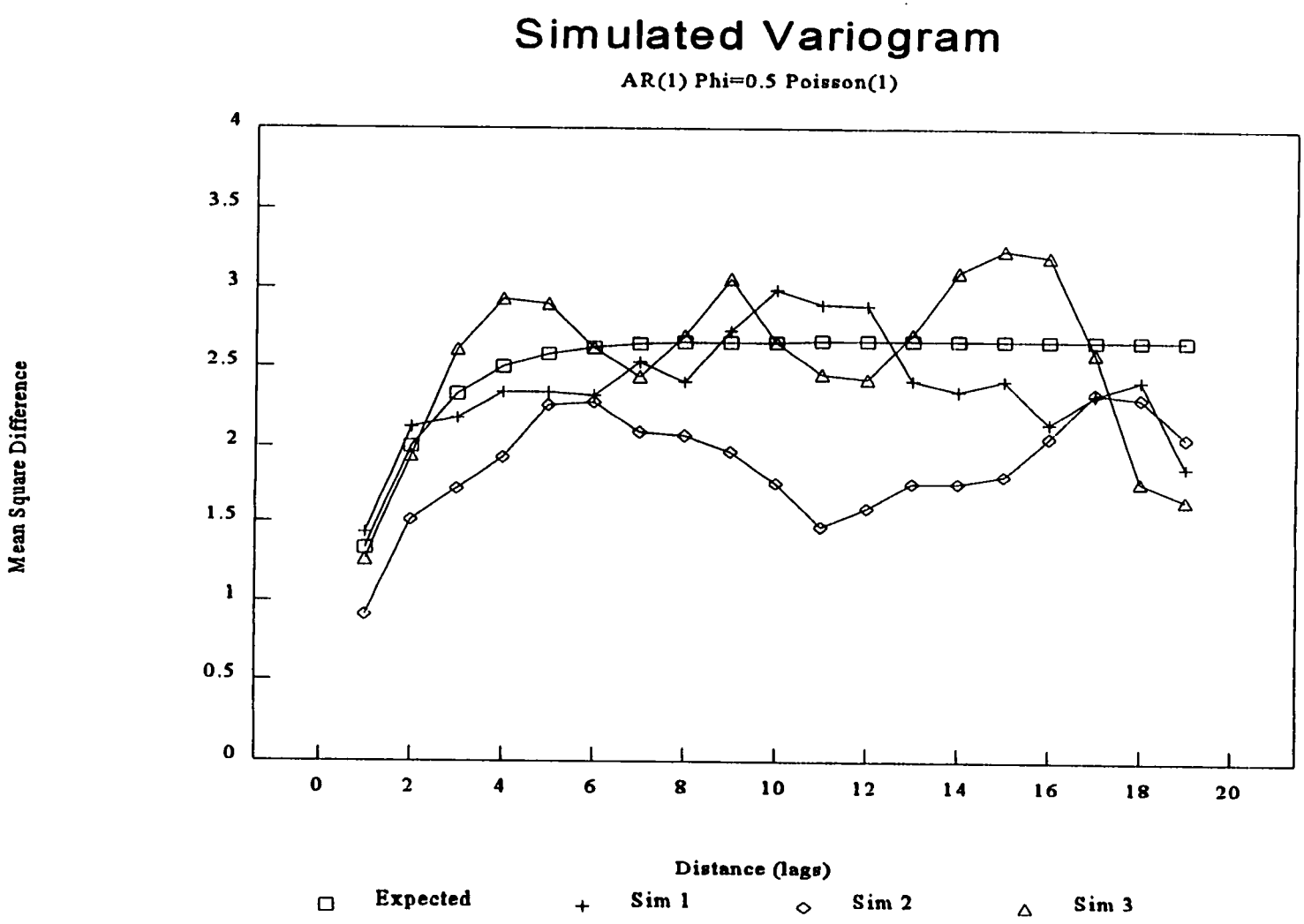

FIGURE 2.3 Expected Variogram and Three Simulations (Poisson Case)

Haining (1990, pp. 90-94) describes examples of isotropic covariance functions that can be used to specify relevant variance-covariance matrices **V**. Since **V** has to be invertible, the covariance function $C(\cdot)$ must preserve the symmetry and positive definite property of **V**. So:

(i) $C(0) \geq 0$ (variance must be non-negative)

(ii) $C(r) = C(-r)$ (symmetry)

(iii) $|C(r)| \leq C(0)$.

Further, relevant covariance functions for social and environmental data often have these common features:

(i) $C(r) \rightarrow 0$ as r increases

(ii) $|C(r_1)| < |C(r_2)|$ if $r_1 > r_2$

(iii) $C(r) \geq 0$ for all r.

The following example covariance functions are described by Haining (1990, p. 91):

(i) Triangular correlation function

$$C(r) = \sigma^2(1 - \gamma r) \quad \text{for } r \leq 1/\gamma$$

$$= 0 \quad \text{for } r > 1/\gamma$$

(ii) Spherical model

$$C(r) = \sigma^2\{1 - (3|r|/2\alpha) + (|r|^3/2\alpha^3)\} \qquad |r| \leq \alpha$$

$$= 0 \qquad |r| > \alpha$$

(iii) Exponential function

$$C(r) = \sigma^2 \exp[-ar^{\beta}] \qquad a > 0;\ \beta > 0 \qquad (2.6)$$

(iv) Whittle's model (1954)

$$C(r) = \sigma^2\ \eta r K_1(\eta r) \qquad \eta > 0$$

where $K_1(x)$ is a modified Bessel function of the second kind of order 1 (Murphy, 1960, p.339).

Guttorp and Sampson (1994) refer to the Gaussian covariance function which is the same as equation 2.6 with $\beta = 2$:

(v) Gaussian function:

$$C(r) = \sigma^2 \exp[-ar^2] \qquad a > 0.$$

Ecker and Gelfand (1997) have added the Cauchy covariance function to the choices of parametric models:

(vi) Cauchy correlation model:

$$C(r) = \sigma^2\ [1/(1 + ar^2)] \qquad a > 0. \qquad (2.7)$$

A variety of methods exist for estimating the parameters of the variance-covariance function. Zeger (1988) uses only two parameters to specify this function, σ^2 and ρ, the first order autocorrelation. These are estimated by the method of moments, but he indicates that other methods may be necessary, especially for small sample sizes and large $|\rho|$. Kitanidis (1983) presents and applies three methods: maximum likelihood, minimum variance unbiased quadratic and minimum norm (least squares). Cressie (1985) shows that weighted least squares and generalized least squares can be used to estimate these parameters at the cost of increasing complexity. He also presents a robust estimator which is a compromise between simplicity and accuracy. Zimmerman and Zimmerman (1991) compare the performance of seven different versions of the above estimators and then go on to compare the performance of the ordinary kriging predictor corresponding to each of the seven estimators. They conclude that the computationally simpler ordinary least squares and weighted least squares perform as well or nearly as well as the more complex methods.

The spatial covariance functions considered so far are appropriate for stationary, isotropic processes. The nonparametric approach of Sampson and Guttorp (1992) assumes neither spatial stationarity or isotropy. In this approach, a random function $Z(x,t)$ is observed repeatedly at times t_i at a finite number of sampling stations x_i. The spatial dispersions $\text{var}(Z(x_i,t) - Z(x_j,t))$ are analyzed

instead of the semivariogram. Multidimensional scaling and thin-plate splines are used to obtain nonparametric estimates of var($Z(x_i,t) - Z(x_j,t)$). The resulting nonparametric model corresponds to a valid covariance model as described by Haining (1990, p. 90).

2.2.6 Properties of Semivariogram Parameters

Cressie (1993, p.99) noted that the distributional properties of the semivariogram parameters defined above "are not well understood for any of the methods proposed..." These methods include all of the estimators compared by Zimmerman and Zimmerman (1991). The weighted least squares procedure from their comparison would yield asymptotically normal parameter estimates if the semivariogram estimates, $\hat{\gamma}(\mathbf{h})$ at each lag **h**, were independent (Jennrich, 1969). As Armstrong (1984) pointed out, since these estimates are from squared differences of the observed data, the assumption of independence is untenable. Switzer (1984) proposed that this problem be addressed by linear transformations of the data to uncorrelated contrasts. This would allow confidence interval estimates to be generated for the semivariogram parameters using Jennrich's (1969) asymptotic theory for nonlinear least squares.

Simulation often offers a reasonable alternative to complex distribution theory. Cressie (1993, p. 477) noted several small simulation studies that "examine the finite-sample properties of different types of estimators..." of both semivariogram parameters, θ and the coefficients of explanatory variables, β. These studies have typically been for Gaussian data on a lattice. In general, they conclude that while β can be estimated with low bias and asymptotic variances are good approximations of exact variances, θ has to be bias corrected and asymptotic variances are not good approximations unless the spatial dependence is weak.

The Bayesian approach to variogram modelling offers advantages if previous studies are available that yield prior information about spatial covariance functions. Cui *et al.* (1995) used semivariograms from different, yet comparably polluted, areas to establish prior semivariograms with empirically determined prior distributions for semivariogram parameters. The inverse of the sill of the semivariogram was found to have a χ^2 distribution and the range of the semivariogram had an exponential distribution. Thus, finite sample inference was possible for the semivariogram parameters. Ecker and Gelfand (1997) used discrete mixtures of Bessel functions to nonparametrically model variograms within a Bayesian framework. They point out that this approach yields posterior distributions for each semivariogram parameter which allows inference without

the possibly inappropriate reliance on approximate normality assumptions.

Shapiro and Botha (1991) also used a nonparametric method for semivariogram estimation with mixtures of Bessel functions. Cherry (1997) used the method to generate nonparametric estimates of the sill of the semivariogram which he found to be badly biased using simulated random fields. He then developed a bias correction for the sill using a penalized fitting routine which resulted in essentially unbiased estimates of the sill.

2.2.7 Prediction

As mentioned above, the main approach to prediction in Spatial Statistics is kriging. The method introduced above (Section 2.2.1) is known as simple kriging. Ordinary kriging is similar to simple kriging except that the mean function is assumed to be unknown but estimable. This method also assumes that the spatial data is jointly Gaussian, a constraint not often met in practice, especially for environmental data. Noting that simple transforms to normality may be difficult to identify in the presence of outliers, Hawkins and Cressie (1984) proposed the method of robust kriging. The idea is to model the noise process as a mixture of distributions in which most of the data is Gaussian, but a fraction of it is from some heavy-tailed "contaminating distribution". This distribution could

also be normal but with large variance to account for outliers.

There are currently many forms of kriging which are detailed in Cressie (1993, Ch. 3). Besides ordinary kriging and robust kriging, the two most important types are universal kriging and median-polish kriging. In universal kriging, the expectation of the spatial process E[Z(**s**)] is no longer assumed to be constant, but rather an unknown linear combination of known functions. These functions, in turn, could just be values of explanatory variables at the point **s**. Median-polish kriging assumes that E[Z(**s**)] is unknown, but decomposes additively into directional components.

Other types of predictors exist and have been used in spatial prediction problems. Laslett (1994) mentioned two alternatives: smooth interpolators and nonparametric spatial regression. From these, the method of splines is prominent. In fact, there has been an ongoing debate since the early beginnings of Geostatistics as to what method of prediction to use. This has led to studies of the performances of various alternatives to kriging (Matheron, 1967; Delfiner and Delhomme, 1975; Laslett, 1994) but that method has always remained popular. Ordinary kriging is used in the present study and no conclusions are drawn about other methods, except to note that most environmental applications seem to involve its use.

2.3 Environmental Applications

2.3.1 Spatially Distributed Environmental Data

The variance-mean relationship known as Taylor's power law (Taylor, 1961) has been used by biologists and ecologists to characterize the spatial distribution of individual organisms. The relationship is:

$$s^2 = a\, m^b$$

where s^2 is the sampling variance

m is the sample mean

a is a scaling factor

and b is an empirically determined index of spatial distribution characteristic of a species.

Banerjee (1976) has used this law to analyze the distribution of *Trigoniulus lumbricinus*, an lulid diplopod. He has shown that there are intra-specific variations in b that depend on the sex of the organism. Taylor and Taylor (1977) have applied this law to terrestrial insects and found that the specific value of b that characterizes a species hold at spatial scales of 10^2 km.

King (1981) related the macroinvertebrate distribution in a small river in South Africa to changes in the physico-chemical environment using stepwise

multiple discriminant analysis. The species composition of various benthic communities were described. The communities were identified using cluster analysis. Changes in the communities occurred with changes in substrate and environmental variables such as pH and total alkalinity. Corkum and Currie (1987) have observed distributional patterns of black flies (Diptera) throughout northwestern North America. Using multiple discriminant analysis, they have shown that latitude and distance from stream source were the most useful for separating black fly groups. Similarly, Corkum (1989) used cluster analysis to determine the spatial distribution of benthic invertebrate assemblages at 100 river sites in northwestern North America. She found that mean current velocity and mean depth at site were most useful in delineating site groupings.

2.3.2 The Importance of Individual Benthic Species

While the use of multivariate techniques to analyze biological species assemblages has become quite popular (see Section 2.3.1), the reporting of individual abundance and univariate analysis of species counts is justified if a single organism is dominant. The amphipod *Pontoporeia hoyi* has been found to be dominant in areas of Lake Superior unimpacted by pollution, but was absent in nearshore areas impacted by copper mining activity (Kraft, 1979; Kraft and Sypniewski, 1981). Parker (1980) has noted that *Pontoporeia hoyi* is the

dominant macrobenthic organism in the Great Lakes and is an important food source for immature salmonids and other fish species. In nearshore, southeastern Lake Michigan, Nalepa and Quigley (1983) found that *Pontoporeia hoyi* totally dominated (99.6% by number) the amphipod group. The genera *Procladius*, was found to be the most abundant of the *Insecta* class in nearshore, western Lake Erie by Cole and Weigmann (1983). In 575 samples from nearshore Lake Erie, *Procladius* was found most frequently (in 402 samples) of any benthic taxa (Barton, 1988). *Pontoporeia hoyi* was found in 107 samples. In 1980, Barton and Griffiths (1984) sampled benthic invertebrates at 86 stations in Lake Huron, Georgian Bay and North Channel. They found both *Pontoporeia hoyi* and *Procladius* to be among the most frequently occurring taxa, depending on depth, exposure to wave action, and local geology (substrate). In Lake Ontario, *Pontoporeia hoyi* and *Procladius* were among the most frequently occurring taxa at a depth of 20 metres in a study of nearshore benthic invertebrates (Barton, 1986).

The Great Lakes Water Quality Agreement (International Joint Commission, 1987) includes a quantitative objective for *Pontoporeia hoyi* in Lake Superior to be used as an indicator of ecosystem health:

> "the abundance of the crustacean, *Pontoporeia hoyi*, maintained throughout the entire lake at present levels of

220-320/(metres)^2 (depths less than 100 metres) and 30-160/(metres)^2 (depths greater than 100 metres)."

Quigley (1988) identified the opportunistic, intermittent feeding strategy of *Pontoporeia hoyi* and suggested that it may explain the exclusive and widespread abundance of this organism in profundal areas of the upper Great Lakes, compared to other amphipods. *Pontoporeia hoyi* constitute approximately 65% of the benthic macroinvertebrate biomass at depths greater than 30 metres in Lake Michigan (Nalepa, 1989). In Lake Huron, *Pontoporeia hoyi* was reported by Shrivastava (1974) to make up over 80% of the total macroinvertebrate biomass in waters over 90 metres in depth. Poulton *et al.* (1988) found *Pontoporeia hoyi* abundance to be negatively correlated with metal content in western Lake Ontario sediments.

Bousfield (1989) revised the amphipod genera *Pontoporeia* by dividing it into three genera, including *Diporeia*, a new genus. All organisms previously identified as *Pontoporeia hoyi* have since been referred to as *Diporeia hoyi*. The sensitivity of *Diporeia* to cadmium, salinity and temperature has been examined in laboratory tests by Gossiaux *et al.* (1992). *Diporeia* was one of two dominant taxa in the diet of the slimy sculpin, an important link in Lake Ontario's food chain (Owens and Weber, 1995). Fratt *et al.* (1997) found that *Pontoporeia hoyi* made up 17% of the volume of food in burbot less than 350 mm in length

from western Lake Michigan and Green Bay.

2.3.3 Predictions of Organism Presence and Abundance

Poole (1978) reviewed statistical techniques that have been used to predict population fluctuations over time. Most of the applications he discusses are the use of forecasting techniques with classical ARIMA and transfer function models. He illustrates the forecasts with an example data set of sheep blowfly egg abundance.

Armitage *et al.* (1983) appraised a presence-absence scoring system for running-water sites that assigned weights to families of benthic macroinvertebrates based on their pollution tolerance. Multiple linear regressions were then developed to predict site scores based on physical and chemical variables. The score predictions were tested by calculating the sum of squares of differences between observed and predicted scores. Moss *et al.* (1987) used multiple discriminant analysis to estimate the probability of occurrence of individual macroinvertebrate species from 28 environmental variables. Armitage *et al.* (1987) reduced the number of predictor variables to five and used them to estimate the probability of occurrence and expected relative abundance of 37 commonly occurring macroinvertebrate families. The technique was applied to

the problem of assessing the effects of flow regulation on a set of upland reservoirs in Great Britain.

The approach of Reynoldson *et al.* (1997) was to predict membership in groups of macroinvertebrates based on physical and chemical variables from the Great Lakes reference site data base of Bailey *et al.* (1995). Data on organism abundance from impacted sites could then be compared to the appropriate reference group based on the physical and chemical data. Alternatively, the results of physical and chemical sampling could be used to predict species composition at sites where sediment remediation was proposed.

2.3.4 Approaches to Dependence in Environmental Data

Millard *et al.* (1985) examined the effect of spatial and temporal correlations on statistical models used to detect ecological change. They used the results of Monte Carlo simulations to show that either spatially or temporally correlated errors can seriously affect the outcome of statistical tests such as ANOVA. The example of the effect of a nuclear power plant in Lake Michigan on total phytoplankton abundance at sites within 3 km. of the discharge is used to illustrate their findings.

Elliott and El-Shaarawi (1995) analyzed sediment lead loads in the Niagara River. Their approach was to consider weekly data from two points on the river, Fort Erie (upstream) and Niagara-on-the-Lake (downstream). Each point was subjected to time series analysis separately and then the ratio of loads between the two points was modelled. They found that autoregressive models fit the data well after adjustment for seasonality and trend. AR(1) models were adequate for the Fort Erie data and the ratios of loads and an AR(4) model was needed for the Niagara-on-the-Lake data.

2.3.5 Environmental Applications of Spatial Statistics

Kitanidis (1983) used a set of spatially correlated rainfall data to illustrate estimation of spatial covariance methods discussed above. This example is of particular interest because the experimental semivariogram presented appears very similar to that found for the benthic data used in this study. The main common feature is many zero or very small squared differences at all distances.

Upton and Fingleton (1985, Ch. 5) used example data from Johnson *et al.* (1968) on the distribution of plant species in California coastal localities and the Galapagos Islands to illustrate a variety of regression based approaches for

modelling spatial dependence. These methods included maximum likelihood with spatially correlated errors and Poisson responses with a logarithmic link. These approaches are similar to the quasi-likelihood approach with time dependent errors used by Zeger (1988).

Brus and de Gruijter (1993) compared model-based prediction of spatial means to design-based estimation. They concluded that even if spatial autocorrelation is accounted for, model-based predictions are not, in general, unbiased. To illustrate the two approaches, they applied them to a case study of areal fraction of soil saturated with phosphorus in the Netherlands. Spatial structure was modeled using the spherical model as the covariance function as discussed above.

Gotway and Hartford (1996) used an extension of universal kriging known as "kriging with external drift" to predict areas of corn fields in Nebraska where variable rates of nitrogen fertilizer application could be used to reduce groundwater contamination. The problem was also addressed with cokriging and the results of the two analyses compared. In this application, the covariate was yield which can be measured easily and accurately with the aid of a global positioning system (GPS). Both methods were found to be viable ways to produce maps for variable-rate fertilizer application. The "kriging with external

drift" approach was easier to apply, but co-kriging utilized more of the information available from covariates.

Recently, Gotway and Stroup (1997) extended the methodology of generalized linear models to spatial statistics for discrete and categorical data. They show how the methods of quasi-likelihood can be applied to problems of estimation and prediction with spatially correlated variables. These methods were illustrated by three example applications: estimation of treatment effects in agronomic field trials, prediction of weed counts in a corn and soybean field, and indicator mapping of groundwater levels in a well. The weed count example is similar to the species count examples analyzed in subsequent chapters of this thesis. The results of this application were that generalized linear models using Poisson and negative binomial link functions gave better predictions than universal kriging.

Ecker and Gelfand (1997) adopted a Bayesian approach to variogram modelling (see Section 2.2.6). They applied it to example data of scallop catches (counts) in the Atlantic Ocean off the New Jersey and Long Island coastline. In this application the count data was log-transformed with the constant 1 being added to each value to address the problem of observed zero catches. Exploratory analysis of the transformed data indicated that the assumptions of

normality and second-order stationarity were justified. Isotropic variogram models were fit using both parametric and nonparametric methodologies. A utility-based model choice criterion based on the deviance associated with the Gamma distribution was introduced to differentiate between models. The best-fitting parametric model was the Cauchy (equation 2.7) while the best-fitting nonparametric model was a mixture of five Bessel functions.

Rathbun (1998) used restricted maximum likelihood (REML) and universal kriging to predict values for salinity and dissolved oxygen in Charleston Harbor, an estuary on the Atlantic Ocean in South Carolina. Of particular importance in this study was the selection of a distance metric. Use of Euclidian distance between sampling sites was found to give less realistic predictions, especially for salinity. An alternative metric, water distance, was defined as the shortest distance between two sites that could be traversed entirely by water. The use of this metric improved the salinity predictions. Dissolved oxygen was not as sensitive to the choice of the distance metric because the semivariogram for this parameter had a small range.

CHAPTER 3: QUASI-LIKELIHOOD APPLIED TO SPATIAL STATISTICS

3.1 Introduction

As noted above, one of the problems with modelling count data, with or without independence, is the occurrence of zeroes. This makes the Poisson distribution an attractive alternative, and quasi-likelihood methods even more so, when no distribution can be specified. In this chapter, the regression model of Zeger for time-dependent counts will be extended to the case of spatially dependent data. As an example application, the resulting model is applied to the data of Upton and Fingleton, using the same spatial covariance structure, but quasi-likelihood methods. This example, using California plant species data, originally assumed the errors were Gaussian.

Likelihood ratio tests are used to check the improvements to the model for plant species counts by accounting for spatial dependence in the case of least squares models and for a more realistic variance function in the case of quasi-likelihood models. Cross-validation of models with different distributional assumptions is used to assess goodness of fit. Quasi-likelihood residuals from two example applications are examined.

3.2 Parameter Estimation

3.2.1 Introduction

As discussed in Section 2.1.4, Zeger (1988) generalized the quasi-likelihood estimating equations to the time series case. This was done through the variance-covariance matrix:

$$\mathbf{V} = \text{var}(Z) = \mathbf{A} + \sigma^2 \mathbf{A} \boldsymbol{R}_\epsilon \mathbf{A} \qquad (3.1)$$

where: $\mathbf{A} = \text{diag}(\mu_1, \ldots, \mu_n)$

and $\boldsymbol{R}_\epsilon$ is an n x n matrix with each j,k element = $\rho_e(\tau)$

where: $\tau = |j - k|$.

The covariance function is then modeled through the $\boldsymbol{R}_\epsilon$ matrix. Zeger observed that this approach can be generalized to a more complex covariance structure. In this section, the generalization to spatial covariance will be accomplished. First, the independent case of the models to be used will be presented. Next, the estimating equations of Zeger will be applied to the example of polio counts without the approximation used in the 1988 paper. Finally, the equations will be applied to the species counts example of Upton and Fingleton (1985, Ch. 5) using an $\boldsymbol{R}_\epsilon$ matrix composed of spatial interactions specified by the authors.

3.2.2 Parameter Estimation with Independent Data

3.2.2.1 **Zeger's Polio Count Data**

Zeger (1988) illustrated the application of the quasi-likelihood model to data on the monthly incidence of poliomyelitis in the U.S. for the years 1970 to 1983 (Table 3.1). For comparison to the times series model, he included the case where independence among the polio counts was assumed. An iterative weighted least-squares procedure was used to obtain the parameter estimates $\hat{\beta}$. The resultant variance-covariance matrix for this case is equation 3.1 with $R_\epsilon = I$. The estimating equations reduce to those resulting from the score function (equation 2:2) of McCullagh (1983) and the parameter estimates $\hat{\beta}$ are found by iterating the following equation until convergence:

$$\hat{\beta}^{(j+1)} = (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\ \partial\mu/\partial\beta)^{-1}\ (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\ \mathbf{Z}) \qquad (3.2)$$

where $\hat{\beta}^{(j)}$ is used to evaluate β on the right-hand side of the equation

$\mathbf{Z} = (\partial\mu/\partial\beta)\beta + (\mathbf{z} - \mu)$

$\mu = \exp(\eta)$, the inverse of the log link

and $\eta = \mathbf{X}\beta$, the linear predictor.

The standard errors of the $\hat{\beta}$s are obtained from the variance-covariance matrix for the estimates:

$$V_{\hat{\beta}} = (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\ \partial\mu/\partial\beta)^{-1}. \qquad (3.3)$$

TABLE 3.1

Monthly Number of U.S. Cases of Poliomyelitis for 1970 to 1983 (Reported to the U.S. Centers for Disease Control and published in Morbidity and Mortality Weekly Report Annual Summary (1970-1983))

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1970	0	1	0	0	1	3	9	2	3	5	3	5
1971	2	2	0	1	0	1	3	3	2	1	1	5
1972	0	3	1	0	1	4	0	0	1	6	14	4
1973	1	0	0	1	1	1	1	0	1	0	1	0
1974	1	0	1	0	1	0	1	0	1	0	0	2
1975	0	1	0	1	0	0	1	2	0	0	1	2
1976	0	3	1	1	0	2	0	4	0	2	1	1
1977	1	1	0	1	1	0	2	1	3	1	2	4
1978	0	0	0	1	0	1	0	2	2	4	2	3
1979	3	0	0	2	7	8	2	4	1	1	2	4
1980	0	1	1	1	3	0	0	0	0	1	0	1
1981	1	0	0	0	0	0	1	2	0	2	0	0
1982	0	1	0	1	0	1	0	2	0	0	1	2
1983	0	1	0	0	0	1	2	1	0	1	3	6

The parameter σ^2 is estimated by the method of moments. Since $var(Z_t) = \mu_t + \sigma^2\mu_t^2$, σ^2 can be estimated by:

$$\hat{\sigma}^2 = \sum_{t=1}^{n} \{(z_t - \hat{\mu}_t)^2 - \hat{\mu}_t\} / \sum_{t=1}^{n} \hat{\mu}_t^2 . \qquad (3.4)$$

The parameter estimates and standard errors from Zeger (1988) are compared to those obtained by implementing the above equations on SAS/IML (SAS Institute Inc, 1990) for the independent case (Table 3.2). The estimates are very similar.

3.2.2.2 **California Plant Species Count Data**

Upton and Fingleton (1985, Ch. 5) use several example data sets to illustrate issues that arise when spatial data are analyzed with regression techniques. The example described in Section 2.3.3 for counts of plant species at various locations on the California coast was analyzed using a multiple regression model with spatially autocorrelated errors. Preliminary analysis of the data (Table 3.3) was done using a multiple regression model with independent, Gaussian errors (i.e., ordinary least squares (OLS)):

$$\mathbf{z} = \mathbf{X}\beta + \epsilon$$

where: $\mathbf{z}$ is a 26×1 vector of the number of plant species found in 26 California locations

TABLE 3.2

Parameter Estimates for Zeger Example, Independent Case

	Zeger's Results		SAS/IML Results	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	0.15	0.10	0.16	0.10
Trend x 10^{-3}	-4.28	2.06	-4.47	2.02
$\cos(2\pi t/12)$	-0.14	0.14	-0.13	0.14
$\sin(2\pi t/12)$	-0.49	0.15	-0.51	0.15
$\cos(2\pi t/6)$	0.18	0.14	0.18	0.14
$\sin(2\pi t/6)$	-0.42	0.14	-0.43	0.14
$\hat{\sigma}^2$	0.80		0.75	

TABLE 3.3

The Number of Plant Species and Other Characteristics of 26 California Locations

Site	Number of Species (z)	Site Area (x_1)	Maximum Elevation (x_2)	Latitude (x_3)
1	205	134	3950	28.2
2	163	98	4600	29.0
3	420	96	2470	34.0
4	340	84	1560	34.0
5	392	75	2125	33.3
6	235	56	1965	32.9
7	120	22	910	33.2
8	190	14	830	34.0
9	42	2.8	490	27.9
10	40	1.0	635	33.4
11	62	0.9	470	30.5
12	4	0.2	130	29.8
13	12	0.1	360	37.7
14	40	0.02	60	37.1
15	39	2.5	660	28.3
16	70	1.1	930	34.0
17	83	1.0	670	32.4
18	72	0.5	315	31.8
19	1450	4260	6535	33.0
20	1400	3324	5860	36.2
21	1060	529	2610	38.1
22	1200	1386	3810	37.3
23	640	320	3110	34.1
24	680	110	3985	34.4
25	640	45	930	37.8
26	370	5.9	750	37.9

X is a 26 × 4 matrix of explanatory variables composed of four vectors, $\mathbf{x}_0$, a column of ones, $\mathbf{x}_1$, the area of the site, $\mathbf{x}_2$, the maximum elevation at the site, and $\mathbf{x}_3$, the latitude of the site.

and ϵ is a 26 × 1 vector of independent, identically distributed Gaussian errors.

Upton and Fingleton (1985, p. 306) observe that while the assumption of normality of errors may not be unreasonable, it can lead to negative estimates for count variables. They suggest as an alternative, the Poisson distribution with site parameters μ_i, functions of the explanatory variables. This results in Poisson regression (Diggle *et al.*, 1994, p. 242) which is similar to the quasi-likelihood approach used by Zeger with a simpler variance function (i.e. $\text{var}(Z_i) = \mu_i$). Zeger's approach utilizes a negative binomial variance function (McCullagh and Nelder, 1983, p. 170).

Equation 3.2 can be used to obtain parameter estimates for any of the above independent errors models by changing the variance function and, in the cases of Poisson regression and quasi-likelihood, taking the natural log of the explanatory vectors, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$. The logs of the variables are used to allow comparison to the OLS case. Tables 3.4 and 3.5 present the results of parameter estimation for each model assuming independent errors. Figure 3.1

TABLE 3.4

Parameter Estimates for California Plant Species Example Using Ordinary Least Squares (OLS) (Upton and Fingleton, 1985, p. 275)

	OLS	
	$\hat{\beta}$	Std err.
Intercept	-1668.2	370.0
Area	0.1642	0.0502
Max. Alt	0.1165	0.0281
Latitude	52.5103	10.8296
$\hat{\sigma}^2 \times 10^{-6}$	0.0275	

TABLE 3.5

Parameter Estimates for Upton and Fingleton Example Using Quasi-Likelihood with $var(Z_i) = \mu_i$ (Poisson Regression) and $var(Z_i) = \mu_i + \sigma^2\mu_i^2$.

	$var(Z_i) = \mu_i$		$var(Z_i) = \mu_i + \sigma^2\mu_i^2$	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-12.839	0.790	-13.335	2.332
log(Area)	0.267	0.014	0.338	0.046
log(Max. Alt.)	0.190	0.042	0.058	0.128
log(Latitude)	4.581	0.173	4.914	0.557
$\hat{\sigma}^2$	-		0.050	

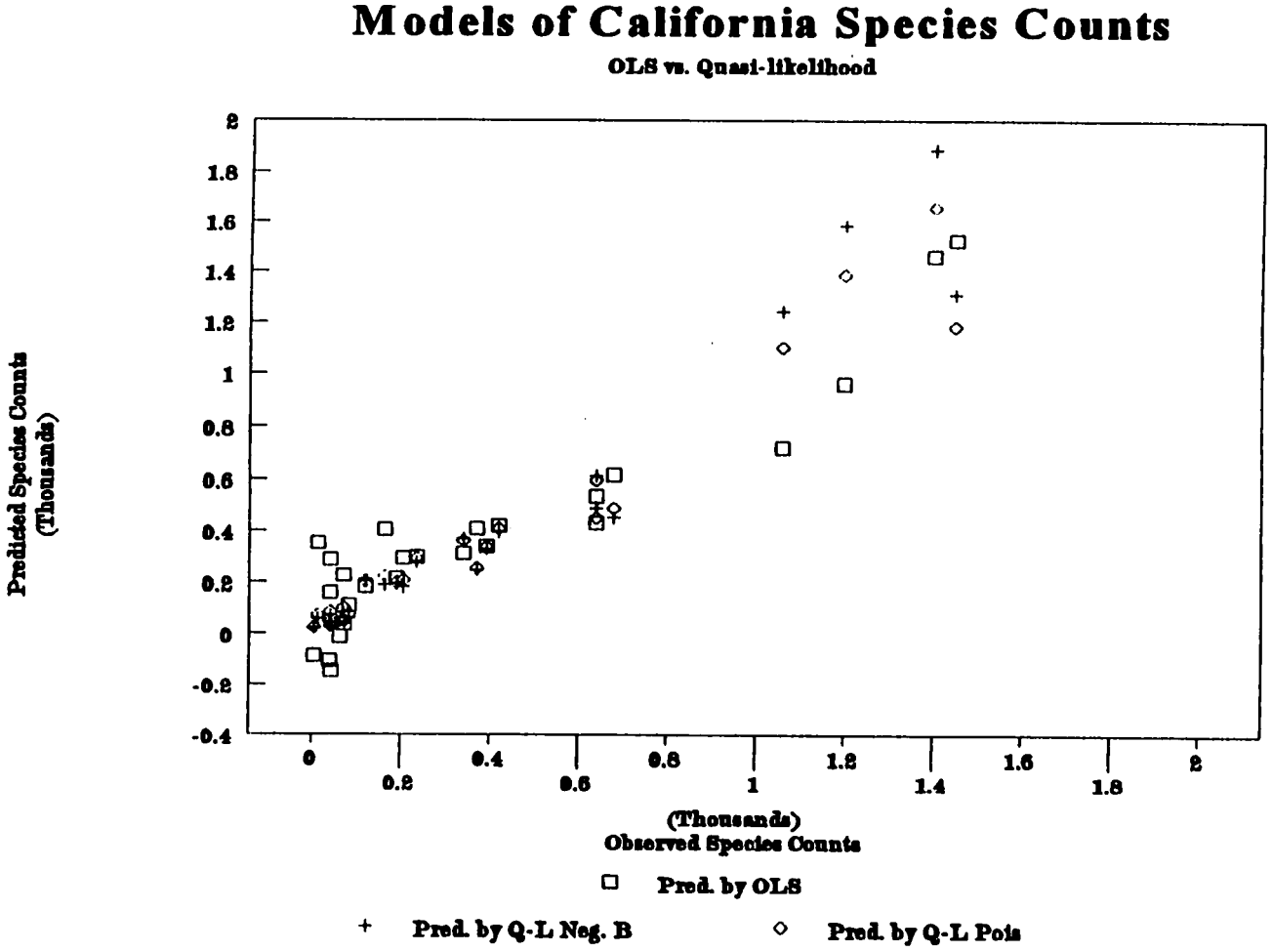

FIGURE 3.1 Comparison of Models of California Species Counts

compares the predicted counts for each model to the observations. Note the negative predictions for the OLS model for observations close to zero.

Since some of the OLS predictions are negative, the application of this model results in inadmissible values. Although the Poisson regression model results in admissible predictions for all observations, the consequences of overdispersion in the count data should not be ignored. These include underestimation of standard errors of regression coefficients and, consequently, over-statement of significance in hypothesis testing (Diggle *et al.*, 1994, p. 245). The standard error estimates for Poisson regression in Table 3.5 are considerably lower than for the quasi-likelihood with negative binomial variance function case . Also, the estimated coefficient for log maximum altitude would not be significantly different from zero when the negative binomial variance function is assumed. These models are compared further in Section 3.3.

3.2.3 Parameter Estimation with Time Dependent Data

The iterative weighted least-squares procedure used by Zeger for the time dependent case is the same as that used for quasi-likelihood with independent data (equation 3.2). Autocorrelation is introduced through the R_ϵ matrix in equation 3.1. Zeger assumed a first-order autoregressive process,

so the j,k element of $\boldsymbol{R}_\epsilon$ was equal to $\hat{\rho}_e^{\tau}(1)$, the estimated first lag autocorrelation raised to the $\tau = |j - k|$ th power. As in the independent case, σ^2 in the variance-covariance matrix can be estimated by a method of moments (equation 3.4). Similarly, the first lag autocorrelation, $\rho_e(1)$, can be estimated by:

$$\hat{\rho}_e(1) = \hat{\sigma}^{-2} \sum_{t=2}^{n} \{(z_t - \hat{\mu}_t)(z_{t-1} - \hat{\mu}_{t-1})\} / \sum_{t=2}^{n} \hat{\mu}_t \hat{\mu}_{t-1} \qquad (3.5)$$

The parameter estimates and standard errors from Zeger (1988) are compared to those obtained by implementing the above equations on SAS/IML (SAS Institute Inc., 1990) for the time dependent case (Table 3.6). The estimates are very similar except for the first lag autocorrelation. The conclusions drawn by Zeger are supported by the SAS/IML estimates: the evidence for a linear decrease in the number of monthly polio cases is weaker when autocorrelation is taken into account. The slight differences in estimates in Table 3.6 are due to the approximation that Zeger used to avoid the inversion of the large (168 × 168) variance-covariance matrix. The SAS/IML (SAS Institute Inc., 1990) estimates were done on an IBM mainframe for which inverting a 168 by 168 matrix presented little difficulty.

3.2.4 Parameter Estimation with Spatially Dependent Data

Following the analogy of time series analysis, spatial covariance is

TABLE 3.6

Parameter Estimates For Zeger Example, Time Dependent Case

	Zeger's Results		SAS/IML Results	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	0.17	0.13	0.17	0.15
Trend x 10^{-3}	-4.35	2.68	-3.80	2.94
$\cos(2\pi t/12)$	-0.11	0.16	-0.12	0.17
$\sin(2\pi t/12)$	-0.48	0.17	-0.49	0.18
$\cos(2\pi t/6)$	0.20	0.14	0.19	0.14
$\sin(2\pi t/6)$	-0.41	0.14	-0.43	0.14
σ^2	0.77		0.77	
$\hat{\rho}_e(1)$	0.77		0.55	

incorporated into the quasi-likelihood estimation process through the $\boldsymbol{R}_\epsilon$ matrix. In Section 3.2.3, this matrix was n x n with each element equal to $\hat{\rho}_\epsilon^\tau(1)$, where $\tau = |j - k|$. Since this corresponded to an AR(1) process, the SAR model (equation 2.3) is now introduced to account for spatial dependence and $\boldsymbol{R}_\epsilon$ has the form (Haining, 1990, p. 81),

$$\boldsymbol{R}_\epsilon = ((\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W}))^{-1}$$

where **W** is the spatial interaction matrix as discussed below.

In Section 3.2.2.2, the example of Upton and Fingleton was modelled using the quasi-likelihood approach assuming independence of plant species counts. Now the **W** matrix from Upton and Fingleton (1985, pp. 291-292) will be used to illustrate the application of quasi-likelihood to the spatially dependent case. The spatial interactions were accounted for by a 'two-part distance-based' weights matrix:

$$w^*_{ij} = d^{-2}_{ij}, \qquad d_{ij} \leq k, \tag{3.6}$$

$$w^*_{ij} = 0, \qquad d_{ij} > k, \tag{3.7}$$

$$w_{ij} = w^*_{ij} / \sum_j w^*_{ij}, \tag{3.8}$$

where k is the 'cut-off' distance and d_{ij} is the distance between sites i and j. Upton and Fingleton assumed that the weight that the error value at site i carries with respect to the value at site j decreased non-linearly with increasing intersite distance until the distance k where the autocorrelation mechanism ceases to

function. Part of the **W** matrix is illustrated in Table 3.7.

As for the time dependent case, the maximum quasi-likelihood equations are written:

$$(\partial \boldsymbol{\mu}/\partial \beta)\, \mathbf{V}^{-1}\{\beta, \theta(\beta)\}\, (\mathbf{z} - \boldsymbol{\mu}) = 0$$

where θ are now the parameters σ^2 and ρ which are treated separately and not estimated with weighted least squares

and **V** is the variance-covariance matrix, equal to

$$\mathbf{A} + \sigma^2 \mathbf{A}((\mathbf{I} - \rho \mathbf{W})'(\mathbf{I} - \rho \mathbf{W}))^{-1}\mathbf{A}.$$

Once again, $\mu = \exp(\mathbf{X}\beta)$ where the vectors of **X** are the logs of the three explanatory variables plus a vector of 1's for the intercept.

Equations 3.1-3.3 are used to obtain the parameter estimates and standard errors, with the exception of $\hat{\rho}$, which was assumed to be equal to -0.72 as found by Upton and Fingleton (1985, p. 292). The results of parameter estimation, with and without accounting for spatial dependence are presented in Table 3.8. Except for an almost threefold decrease in $\hat{\beta}_2$, the coefficient for maximum altitude, the effect of accounting for spatial dependence on the parameter estimates is slight. These results are similar to the those of Upton and

TABLE 3.7

W Matrix (Proximity Values) Used for California Sites

ROW / COL	1	2	3	4	...	25	26
1	0	0.0067	0	0	...	0	0
2	0.1928	0	0	0	...	0	0
3	0	0	0	0.2200	...	0	0
4	0	0	0.2406	0	...	0	0
.	.	.	.	.	...	.	.
.	.	.	.	.	...	.	.
.	.	.	.	.	...	.	.
25	0	0	0	0	...	0	0.4953
26	0	0	0	0	...	0.5096	0

TABLE 3.8

Parameter Estimates for Upton and Fingleton Example Using Quasi-Likelihood With and Without Spatial Dependence

	Independent Case		Spatial Dependence Case	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-13.335	2.332	-13.389	2.145
log(Area)	0.338	0.046	0.361	0.056
log(Max. Alt.)	0.058	0.128	0.020	0.146
log(Latitude)	4.914	0.557	4.988	0.452
σ^2	0.050		0.067	

Fingleton for the multiple linear regression case (Table 3.9). In this case, the main difference caused by the assumption of spatial dependence is to lower $\hat{\sigma}^2$, which, in turn, lowers the standard errors of the parameter estimates. For the quasi-likelihood case, $\hat{\sigma}^2$ increases and the changes in the standard errors of the parameter estimates are not as large. The similarity in the effect on the values of $\hat{\beta}$ is to be expected, because the estimating equations used for $\hat{\beta}$ given by Upton and Fingleton are identical to the iterative weighted least squares used by Zeger (Equation 3.2) for the case of the log linear model. Upton and Fingletons'equations are:

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}).$$

These are the same as Zeger's equations, since $\partial\mu/\partial\beta = \mathbf{X}$. The differences in the effects on the standard errors for the two models (least squares vs. quasi-likelihood) are due to the estimation procedure for determination of $\hat{\sigma}^2$ and the a priori specification of $\hat{\rho}$ for the quasi-likelihood model which may not be optimal. The topic of the best spatial covariance function for quasi-likelihood models is the subject of Chapter 4.

TABLE 3.9

Parameter Estimates for California Plant Species Example Using Multiple Linear Regression With and Without (OLS) Spatial Dependence (Upton and Fingleton, 1985, pp. 275 and 292)

	OLS		Autocorrelated Errors	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-1668.2	370.0	-1618.9	173.5
Area	0.1642	0.0502	0.1628	0.0361
Max. Alt	0.1165	0.0281	0.1040	0.0224
Latitude	52.5103	10.8296	51.8470	4.8765
$\sigma^2 \times 10^{-6}$	0.0275		0.0153	

3.3 Goodness of Fit

3.3.1 Introduction

Cressie (1993, p. 498) points out that sensible model-selection criteria are needed for the case of spatially dependent data. The likelihood functions for Gaussian models are readily available and so, likelihood ratio tests are a possibility. For quasi-likelihood models, no actual likelihood function is specified. However, McCullagh and Nelder (1983, p. 170) list quasi-likelihood functions associated with some simple variance functions. The quasi-likelihood corresponding to the variance function ($\mu + \sigma^2\mu^2$) from Zeger's model is from the negative binomial likelihood. Cressie (1993 pp. 432-433) defines the auto negative binomial spatial model but notes that it is "almost never possible to construct a closed-form likelihood" for this type of model because of the intractability of the normalizing constant. In order to compare spatial models with different distributional assumptions a different type of procedure is needed.

In the original example with no spatial dependence, Upton and Fingleton used ordinary least squares to fit the number of species to the three explanatory variables. They then used maximum likelihood with spatially correlated errors to estimate the parameters for the dependent case. The main

effect of accounting for spatial interactions was to change the standard errors of the estimates, thus more realistically describing the uncertainties with the fit of the model. Upton and Fingleton (1985, p. 306) then suggest the Poisson distribution with site parameters μ_i, functions of the explanatory variables, as an alternative to guarantee non-negative counts. As noted previously, this is similar to the quasi-likelihood approach of Zeger, with a simpler variance function.

Accounting for spatial dependence in the least squares model amounts to changing the variance function from $\mathbf{V} = \sigma^2\mathbf{I}$ in the independent case, to $\mathbf{V} = \sigma^2((\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W}))^{-1}$. The significance of this change can be assessed by the likelihood-ratio test as suggested by Upton and Fingleton (1985, p. 303). Accounting for overdispersion in the quasi-likelihood model amounts to changing the variance function from $\mathbf{V} = \mathbf{A}$ in the case of Poisson regression, to $\mathbf{V} = \mathbf{A} + \sigma^2\mathbf{A}^2$, where $\mathbf{A} = \text{diag}(\mu_1,\ldots,\mu_n)$ as in equation 3.1 (the negative binomial variance function). A likelihood-ratio test for comparing two variance functions with discrete observations is described by McCullagh and Nelder (1983, pp. 212-214). These tests are applied to the models for the Upton and Fingleton example in this section.

It is reasonable to ask if the overall fit has improved by proceeding from least squares to quasi-likelihood. Because different models are involved, a

measure of goodness of fit that will allow for different modelling assumptions as well as assess the fit is needed. Cross-validation as developed by Stone (1974) and Geisser (1975) has been suggested by Cressie (1993, p. 101) as a tool to employ in this situation. In Chapter 5, this procedure will be used to compare different variogram models. In this section, cross-validation is used to compare models with different distributional assumptions.

3.3.2 Likelihood-Ratio Tests

3.3.2.1 **Least Squares, Dependent vs. Independent**

Upton and Fingleton (1985, p.293) report that the explained proportion of the variation in observed California plant species counts increases from 0.876 to 0.918 when spatial dependence is accounted for in the multiple linear regression model. The significance of this increase can be tested with a likelihood-ratio test (Upton and Fingleton, 1985, p. 304). The test compares a pair of nested models which are identical except for k restrictions. In this case, the single restriction is that $\rho = 0$ (i.e., the errors are uncorrelated). The null hypothesis is that the restrictions are true. Since it is more convenient to work with log-likelihoods, the test statistic is $2\{\ln(L_U) - \ln(L_R)\}$ which is asymptotically distributed as χ^2_k, where L_U is the likelihood of the unrestricted model and L_R is

the likelihood of the restricted model.

The likelihood function for ordinary least squares assuming the residuals, ϵ, are MVN($\mathbf{0}, \sigma^2\mathbf{I}$), is:

$$L = (1/(\sigma^n(2\pi)^{n/2})\exp((-\epsilon'\epsilon))/2\sigma^2)$$

where $\epsilon = \mathbf{z} - \mathbf{X}\beta$.

The log maximum likelihood for the restricted model is thus:

$$\ln(L_R) = -(n/2)\ln(2\pi\hat{\sigma}^2) - (1/2\hat{\sigma}^2)(\hat{\epsilon}'\hat{\epsilon})$$

where $\hat{\epsilon} = \mathbf{z} - \mathbf{X}\hat{\beta}$.

Cressie (1993, p. 92) gives the negative log-likelihood for least squares with spatially correlated errors as:

$$-l = (n/2)\ln(2\pi) + (1/2)\ln(|\mathbf{V}|) + (1/2)(\epsilon'\boldsymbol{V}^{-1}\epsilon)$$

where $\boldsymbol{V} = \sigma^2((\mathbf{I} - \rho\mathbf{W}')(\mathbf{I} - \rho\mathbf{W}))^{-1}$

and $\mathbf{W}$ and ρ are as defined in Section 3.2.3.

The log maximum likelihood for the unrestricted model is thus:

$$\ln(L_U) = -(n/2)\ln(2\pi) - (1/2)\ln(|\hat{\boldsymbol{V}}|) - (1/2)(\hat{\epsilon}'\hat{\boldsymbol{V}}^{-1}\hat{\epsilon}).$$

Using the results from the example in the above equations, $\ln(L_U) = -161.26$, $\ln(L_R) = -167.62$, and $2\{\ln(L_U) - \ln(L_R)\} = 12.73$ which is greater than $\chi^2_{.001,1} = 10.83$. Therefore, the null hypothesis is rejected and the improvement due to accounting for spatially dependent errors is highly significant.

3.3.2.2 **Quasi-likelihood, Poisson vs. Negative Binomial**

Although the Poisson assumption that Var(Z) = E(Z) is traditionally used with count data, it is often inconsistent with empirical evidence (Diggle *et al.*, 1994, p.164). With biological data, over-dispersion (Var(Z) > E(Z)) is more often the case. Application of the quasi-likelihood approach requires the specification of the relationship between the mean and the variance. If the Poisson assumption is inadequate, some way to test the significance of changing variance functions is needed. The likelihood-ratio test for comparing two variance functions with discrete observations described by McCullagh and Nelder (1983, pp. 212-214) is applied here to the Upton and Fingleton example.

McCullagh and Nelder define the log-likelihood ratio statistic, Λ, for two variance functions, $V_1(\mu)$ and $V_2(\mu)$ with deviances D_1 and D_2, given the same link and linear predictor (see equation 3.2), as:

$$\Lambda = n \ln(D_1/D_2) + \sum_{i=1}^{n} \ln(V_1(z_i)/V_2(z_i))$$

where $D(\mathbf{z};\hat{\boldsymbol{\mu}}) = -2\{l(\hat{\boldsymbol{\mu}};\mathbf{z}) - l(\mathbf{z};\mathbf{z})\}$

$l(\hat{\boldsymbol{\mu}};\mathbf{z})$ is the maximized log-likelihood at $\hat{\boldsymbol{\mu}}$

$l(\mathbf{z};\mathbf{z})$ is the maximum log-likelihood achievable for an exact fit

(i.e., $\hat{\boldsymbol{\mu}} = \mathbf{z}$).

The null hypothesis is that the second variance function is as good as the first.

For discrete distributions, McCullagh and Nelder (1983. p.214) recommend replacing the variance function by $V'(z_i)$,

where $V'(z_i) = z_i + c$ for Poisson

$V'(z_i) = (z_i + c) + \sigma^2(z_i + c)^2$ for negative binomial

and $c = 1/6$.

The Poisson and negative binomial both have the same link functions ($\eta = \ln \mu$) and linear predictor ($\eta = \mathbf{X}\beta$). McCullagh and Nelder (1983, p.25) give the deviance of the Poisson distribution as:

$$D(\mathbf{z};\hat{\mu}) = 2\{\sum_{i=1}^{n} [z_i \ln(z_i/\mu_i) - (z_i - \mu_i)]\}.$$

For the negative binomial distribution, the log quasi-likelihood (McCullagh and Nelder (1983, p.170) is:

$$l(\hat{\mu};\mathbf{z}) = \sum_{i=1}^{n} z_i \ln(\mu_i/(\mu_i + k)) + k \ln(k/(\mu_i + k))$$

where $k = 1/\sigma^2$.

so:

$$\begin{aligned} D(\mathbf{z};\hat{\mu}) &= -2\{l(\hat{\mu};\mathbf{z}) - l(\mathbf{z};\mathbf{z})\} \\ &= -2\{\sum_{i=1}^{n} [z_i \ln(\mu_i/(\mu_i + k)) + k \ln(k/(\mu_i + k))] \\ &\quad - [z_i \ln(z_i/(z_i + k)) + k \ln(k/(z_i + k))]\} \\ &= 2\{\sum_{i=1}^{n} [z_i \ln(z_i(\mu_i + k)/(\mu_i(z_i + k))) \\ &\quad + k \ln((\mu_i + k)/(z_i + k))]\}. \end{aligned}$$

Using the results of the Upton and Fingleton example in the above equations, D_1 = 556.10 (negative binomial), D_2 = 65.69 (Poisson), $\sum_{i=1}^{n} \ln(V_1(z_i))$ = 133.50, $\sum_{i=1}^{n} \ln(V_2(z_i))$ = 195.20, and Λ = 6.16 which is greater than $\chi^2_{.05,1}$ = 3.84 but less than $\chi^2_{.01,1}$ = 6.63. Therefore the null hypothesis is rejected and the improvement due to the use of the negative binomial variance function is significant.

3.3.3 Cross-Validation to Assess Fit

Cross-validation is used to assess the fit of the two quasi-likelihood models (with the Poisson and the negative binomial variance functions) vs. the OLS model. The procedure used follows that suggested by Cressie (1993, p. 102) and applied by Carroll and Cressie (1996). The basic idea is to iteratively delete one observation at a time from the data set and then estimate $\hat{\beta}_{-}$, which are the parameters based on one less observation (Stone, 1974; Geisser, 1975). $\hat{\beta}_{-}$ is used to predict z_{-}, the missing observation. The prediction residuals, $(z_{-}-\hat{z}_{-})$, are standardized by dividing by the square root of σ^2_{k-}, the root mean-square prediction error for one less observation. The standardized prediction residuals (R_s) are averaged and compared for each model. R_s should be close to zero. The root mean-square standardized prediction residuals ($RMSR_s$) are also compared and should be close to one. The prediction sum of squares (PRESS)

statistic (Draper and Smith, 1981, p. 325) is the same as $RMSR_s$ without standardization by σ^2_{k-} and should be as small as possible.

In Chapter 5, the mean-square prediction error, σ^2_{k-}, is determined from the kriging equations. For independent models,

$$\sigma^2_{k-} = \mathrm{Var}(z_- - \hat{z}_-) = \mathrm{Var}(z_- - f(\boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-))$$

where $f(\boldsymbol{x}_-(\hat{\boldsymbol{\beta}}_-)) = \boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-$ for OLS

$f(\boldsymbol{x}_-(\hat{\boldsymbol{\beta}}_-)) = \exp(\boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-)$ for quasi-likelihood

and $\boldsymbol{x}_-$ is the explanatory vector for the deleted observation.

Therefore, the root mean-square prediction errors used to standardize the prediction residual, $(z_- - \hat{z}_-)$, for each model are:

$$\sigma_{k-} = \sqrt{(\hat{\sigma}^2\,(1 + \boldsymbol{x}_-'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_-))} \quad \text{for OLS}$$

$$\sigma_{k-} = \sqrt{(\exp(\boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-)\,(1 + \exp(\boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-)\boldsymbol{x}_-'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_-))} \quad \text{for QL (Poisson)}$$

$$\sigma_{k-} = \sqrt{(\exp(\boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-)\,(1 + \exp(\boldsymbol{x}_-\hat{\boldsymbol{\beta}}_-)(\hat{\sigma}^2 + \boldsymbol{x}_-'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_-))} \quad \text{for QL (neg. bin.)}$$

where $\mathbf{V} = \mathbf{A} + \hat{\sigma}^2\mathbf{A}^2$ for the negative binomial variance function and $\mathbf{V} = \mathbf{A}$ for the Poisson variance function

and $\mathbf{X}$, $\mathbf{A}$ and $\hat{\sigma}^2$ are as defined previously except with one less observation.

The quasi-likelihood root mean-square prediction errors are approximate because they are obtained from a Taylor Series expansion.

Table 3.10 contains the results of the cross-validation for the independent models. The OLS model has the best R_s and $RMSR_s$ but the PRESS is the worst due to the poor fit for low values (see Figure 3.1). The quasi-likelihood model with Poisson variance also has good R_s, but the $RMSR_s$ is very large (indicative of the underestimation of the variance). The quasi-likelihood with negative binomial variance had the worst R_s, but values around 0.1 are not indicative of a severe problem. The other cross-validation statistics are intermediate to the results of the other two models. These model will be compared to the dependent models developed in Chapter 5.

3.4 Quasi-likelihood Residuals

3.4.1 Zeger's Polio Time Series

The residuals resulting from Zeger's model for the independent and dependent cases are presented as histograms in Figures 3.2 and 3.3, respectively. In both cases, the distribution of residuals appears non-Gaussian. This is due to the heavy right tails which are caused by a few observations which are probably outliers. Zeger recognizes the observation responsible for the residual at +11.5 as an outlier, but chose to conduct the analysis with the complete data set. The effect on the residuals of accounting for time

TABLE 3.10

Cross-Validation Results for Independent Models of California Plant Species Data.

Model	R_s	$RMSR_s$	PRESS
OLS	-.020	1.095	187.8
QL (Poisson)	.048	5.428	157.2
QL (negative binomial)	.107	1.572	178.2

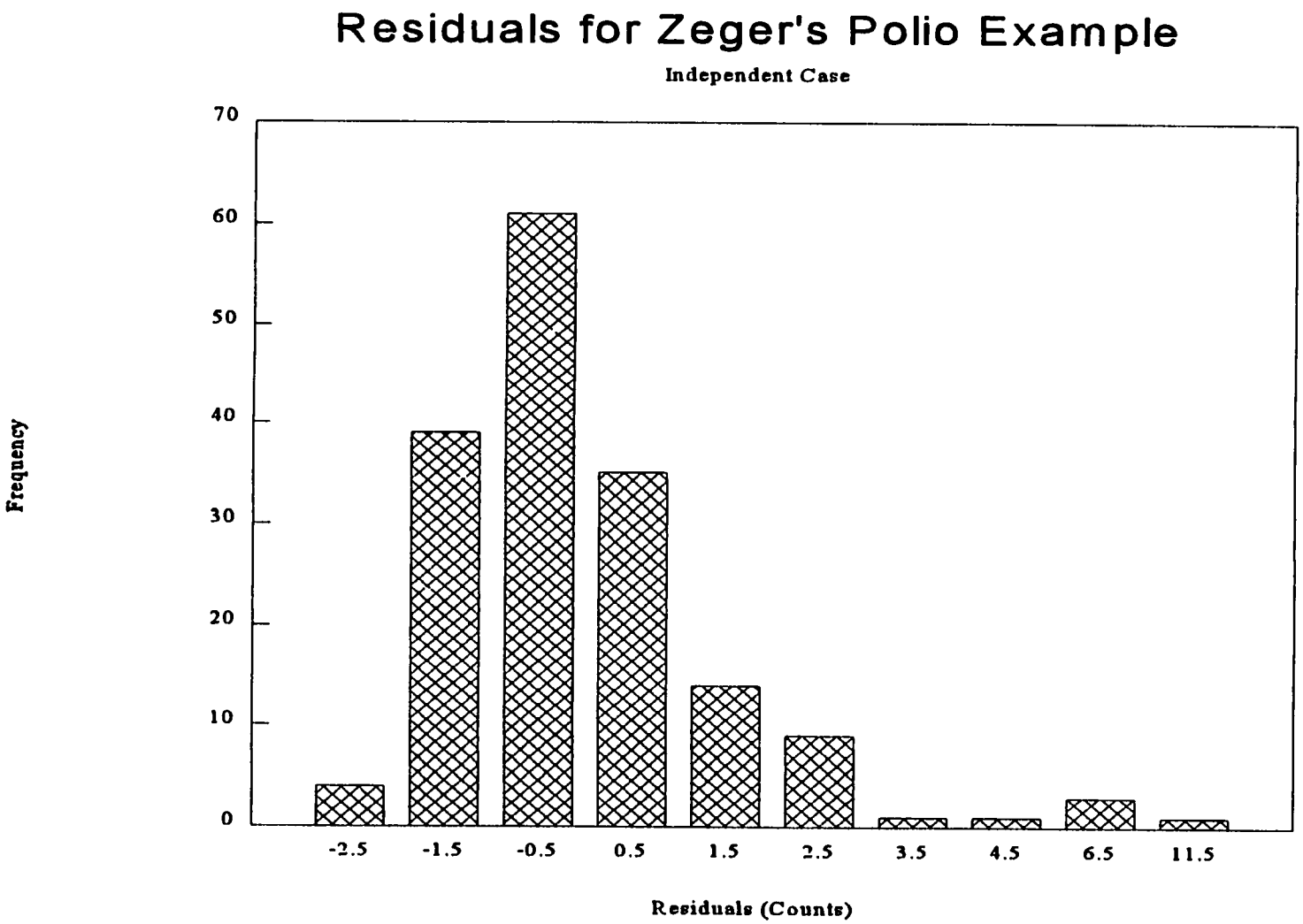

FIGURE 3.2 Residuals for Zeger's Polio Example, Independent Case

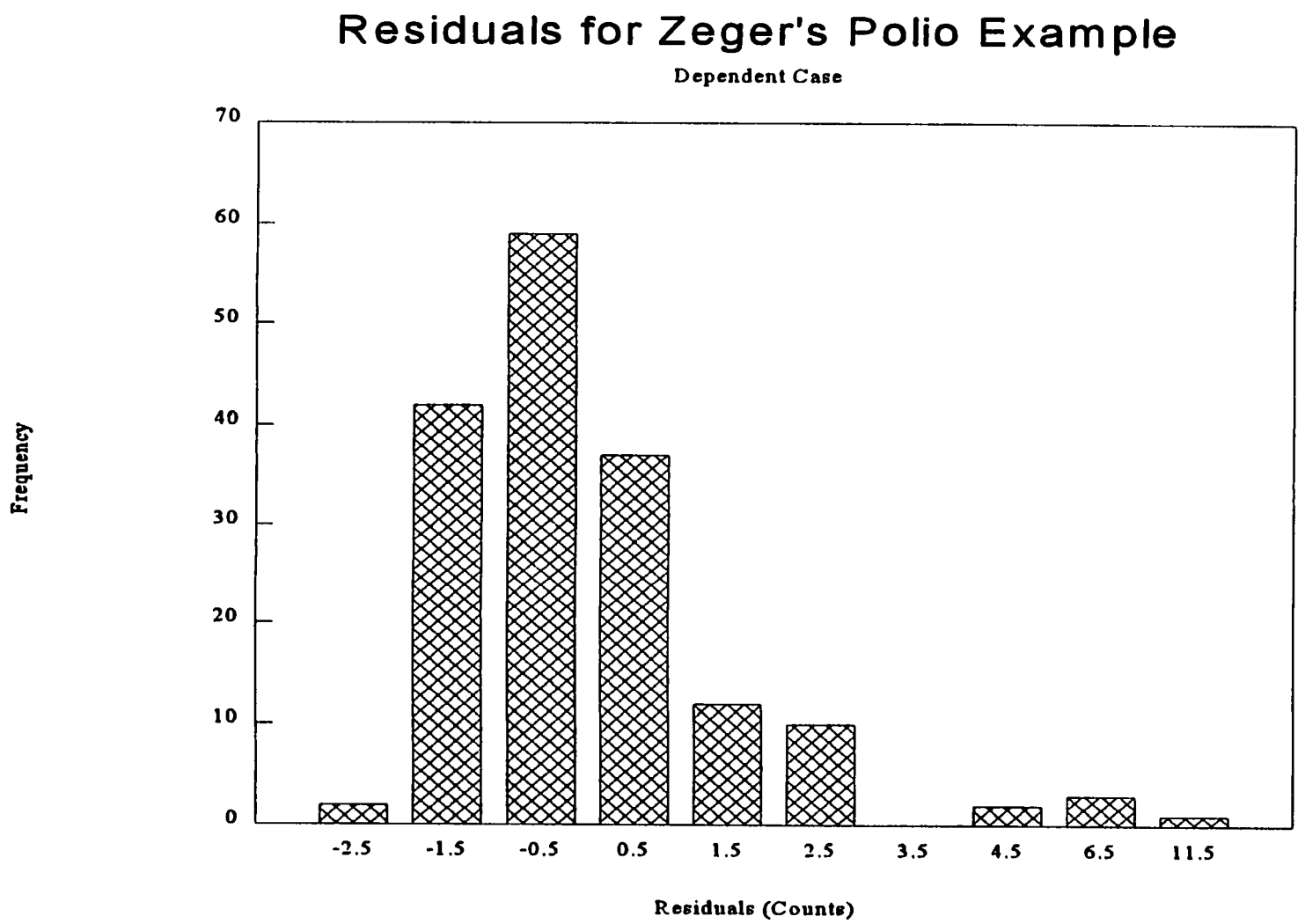

FIGURE 3.3 Residuals for Zeger's Polio Example, Dependent Case

dependence is a very slight increase in the skewness (from 2.701 to 2.747) and kurtosis (from 11.979 to 12.299) of the residuals.

3.4.2 California Plant Species

The quasi-likelihood residuals for the independent case are presented as a stem and leaf plot in Figure 3.4. These plots are used instead of histograms since there are not enough observations to form groups. Because the observations in this example span almost three orders of magnitude (Table 3.3), the residuals appear to have several outliers. This will have implications for the estimation of the variogram in the next chapter because two or three high values will dominate the observed mean-squared differences. If the quasi-likelihood residuals are normalized by division by $\hat{\mu}$, the stem and leaf plot (Figure 3.5) is much more Gaussian. In fact, when the Shapiro-Wilk W statistic is calculated for these 26 residuals (Shapiro and Wilk, 1965), the null hypothesis of normality cannot be rejected. The W statistic is the ratio of the best estimator of the variance to the sum of squares estimator of the variance and for these residuals its value is 0.967.

Stem	Leaf	Number
2	3	1
1	245	3
0	012222236	9
-0	8543332110	10
-1	8	1
-2		0
-3	9	1
-4	9	1

Multiply Stem.Leaf by 200

FIGURE 3.4 Stem and Leaf Plot For Quasi-Likelihood Residuals, California Plant Species Example, Independent Case

Stem	Leaf	Number
6	7	1
4	2200	4
2	03	2
0	145129	6
-0	6654183	7
-2	64	2
-4	51	2
-6	9	1
-8	2	1

Multiply Stem.Leaf by 0.1

FIGURE 3.5 Stem and Leaf Plot For Normalized Quasi-Likelihood Residuals, California Plant Species Example, Independent Case

CHAPTER 4: MODELLING SPATIAL AUTOCORRELATIONS

4.1 Introduction

The autocorrelation structure in the models of Chapter 3 was specified a priori. In Zeger's model, the first lag autocorrelation of the residuals was found by a method of moments and then an AR(1) structure was assumed for the rest of the lags, ie. $\rho(\tau) = \rho(1)^{\tau}$. In Upton and Fingleton's California plant species model, the autocorrelation of the residuals was a parameter to be estimated by maximum likelihood and the structure of the variance-covariance matrix was determined by the interaction matrix, **W**, which had elements w_{ij} based on the inverse squared distance between sites. In spatial statistics, the usual way to determine spatial autocorrelation structure is by fitting a semivariogram, as reviewed in Chapter 2.

In this Chapter, the variogram is used to identify autocorrelation structure for use with the quasi-likelihood model with dependent errors. It is first applied to Zeger's polio example, as a further illustration of the similarities between time series analysis and spatial statistics. It is then applied to Upton and Fingleton's example to continue the idea of replacing the regression with

dependent errors model with the quasi-likelihood model. Finally, it is applied to Reynoldson's benthic data. In all cases, the semivariogram of residuals is used because of non-stationarity.

4.2 Time Series Models With Variograms

4.2.1 Introduction

As discussed in Chapter 2, there is no reason that variograms cannot be used to identify autocorrelation structure in time series data, instead of the usual autocorrelation function (ACF) and partial autocorrelation function (PACF) (Cox, 1981; Diggle, *et al*, 1994, pp. 51-54). Figure 4.1 is the semivariogram for Zeger's polio data for 39 lags. It exhibits all of the characteristics described in Chapter 2: the steep ascent to a "sill" at about lag 5, after which it is roughly constant. Because the polio data is non-stationary, the semivariogram of residuals (Figure 4.2) is used because it reflects more accurately the features in the data. Also, since Zeger used iterative weighted least squares to solve for the parameter estimates, $\hat{\beta}$, updating the value of $\hat{\rho}$ in each iteration based on the current values of $\hat{\mu}$, an analogous procedure using the variogram should be applied, if the comparison of approaches is to be maintained. This is developed in the next section.

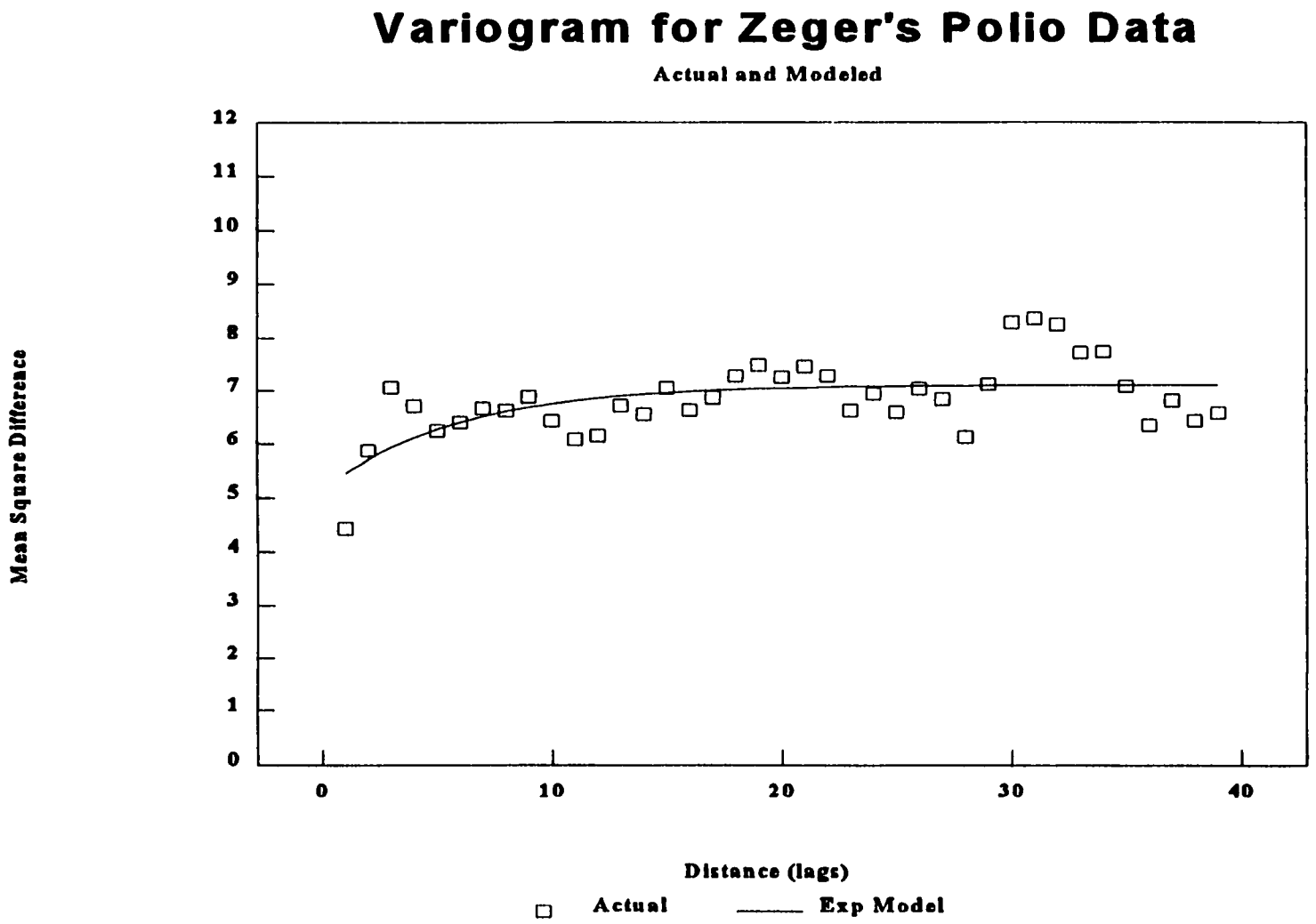

FIGURE 4.1 Variogram for Zeger's Polio Example, Actual Data

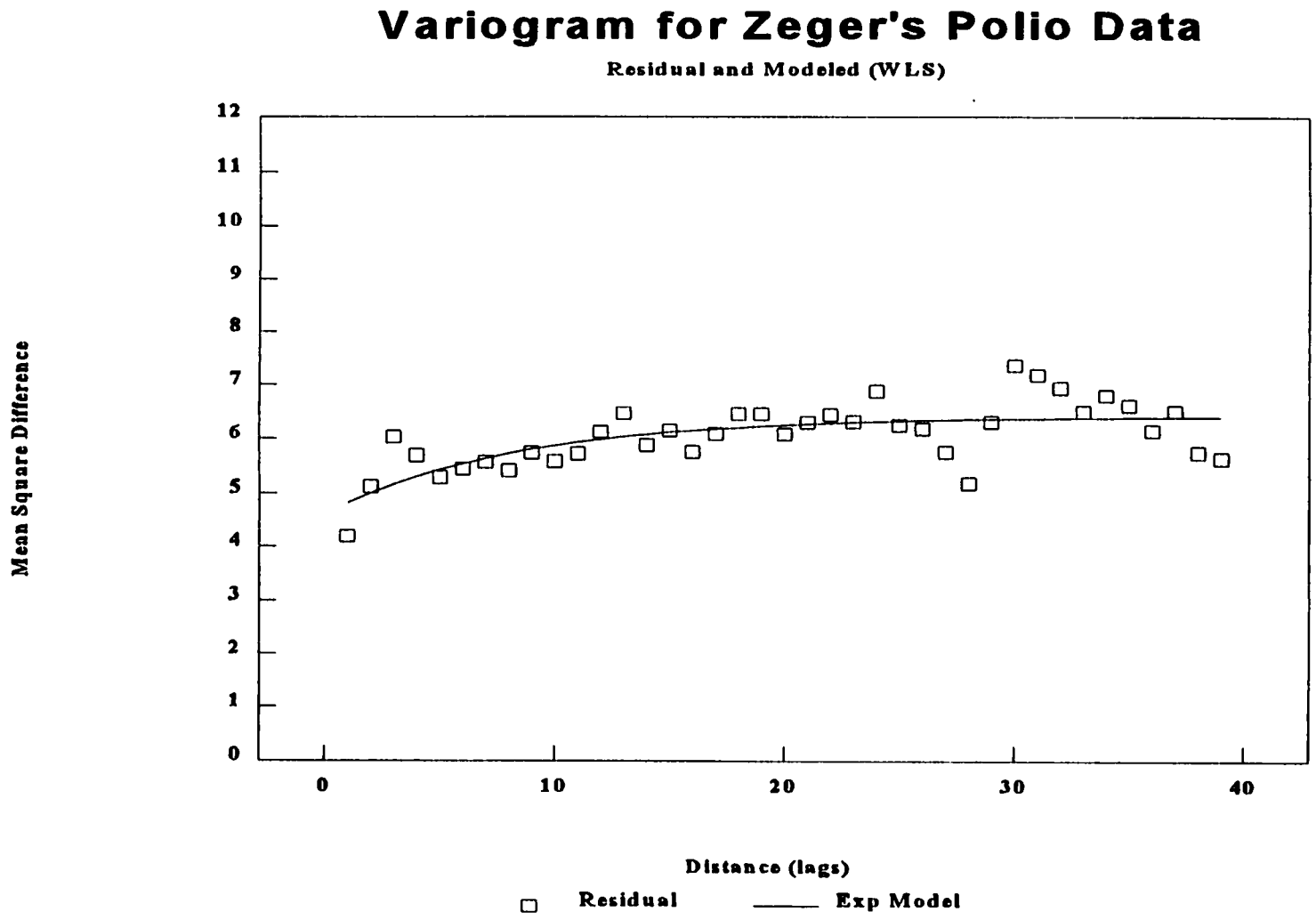

FIGURE 4.2 Variogram for Zeger's Polio Example, Residuals

4.2.2 Fitting the Semivariogram

As reviewed in Chapter 2, there is a variety of semivariogram fitting procedures in the literature. The weighted least squares procedure of Cressie (1985) uses an iterative process to minimize the quantity:

$$\sum_{j=1}^{k} |N(\mathbf{h}(j))| \left((\hat{\gamma}(\mathbf{h}(j))/\gamma(\mathbf{h}(j),\theta) - 1\right)^2 \qquad (4.1)$$

over θ

where k is the number of lags to be used

$|N(\mathbf{h}(j))|$ is the number of distinct pairs of observations at lag $\mathbf{h}(j)$

$\hat{\gamma}(\mathbf{h}(j))$ is the observed semivariogram at lag $\mathbf{h}(j)$

$\gamma(\mathbf{h}(j),\theta)$ is the assumed theoretical (fitted) semivariogram at lag $\mathbf{h}(j)$

and θ is a vector of fitting parameters.

The classical estimate of the semivariogram, $\hat{\gamma}(\mathbf{h}(j))$ (equation 2.5), can be replaced in the above equation, by the robust estimate:

$$\bar{\gamma}(\mathbf{h}(j)) = \{(1/|N(\mathbf{h}(j))|) \sum_{N(\mathbf{h}(j))} |z_t - z_{t-h(j)}|^{.5}\}^4 /2(0.457 + 0.494/|N(\mathbf{h}(j))|),$$

which is more efficient.

Zimmerman and Zimmerman (1991) report that either estimate for the

semivariogram, when fitted by minimizing the above quantity (equation 4.1), sacrifices little to more complex methods when used in conjunction with ordinary kriging. Since the semivariogram estimators, $\hat{\gamma}$ and $\bar{\gamma}$, are moments estimators and since the weighted least squares procedure avoids distributional assumptions (Rathbun, 1998), the above process was implemented. Cressie (1993, p. 99) suggests that the above equation (4.1) can be implemented using the SAS procedure NLIN with the Gauss-Newton method (Hartley, 1961). The equations from NLIN (SAS Institute Inc., 1982, pp. 22-23) were written in SAS/IML (SAS Institute Inc., 1990) for use with the quasi-likelihood model.

4.2.3 Quasi-likelihood Model with Variogram

In the polio count example from Zeger, $\hat{\sigma}^2$ and $\hat{\rho}_e(1)$ were estimated by the method of moments (equations 3.4 and 3.5) at each iteration of the least squares procedure. These equations can now be replaced by the method for fitting the semivariogram of Cressie (equation 4.1) and estimates of $\hat{\sigma}^2$ and $\hat{\rho}_e(1)$ from this fit can be used instead. The theoretical variogram model

$$\gamma(\tau,\theta) = \theta_1(1 - \theta_2^{\tau}),$$

was assumed. $\hat{\theta}_2$ corresponds to $\hat{\rho}_e(1)$, and $\hat{\theta}_1$, the sill of the variogram, is an estimate of the variance of the counts, var(Z_t). Since var(Z_t) = $\hat{\mu}_t + \hat{\sigma}^2\hat{\mu}_t^2$ in Zeger's model, an estimate of σ^2 is

$$\hat{\sigma}^2 = (\hat{\theta}_1 - \bar{\hat{\mu}})/\overline{\hat{\mu}^2}$$

where $\bar{\hat{\mu}} = \sum_{t=1}^{n} \hat{\mu}_t/n$

and $\overline{\hat{\mu}^2} = \sum_{t=1}^{n} \hat{\mu}_t^2/n.$

In the notation of the previous section, $\hat{\theta} = (\hat{\theta}_1, \hat{\rho}_e(1))$ and the optimal $\hat{\theta}$ is used for each iteration of the weighted least squares procedure for parameter estimation.

Residual semivariogram plots for both the classical estimator, $\hat{\gamma}$, and the robust estimator, $\bar{\gamma}$, were examined and the classical estimator appeared to be smoother and easier to fit. This may be due to the equal spacing of the polio data series and the relatively large number of observations for each lag (n=167 for τ=1, n=166 for τ=2, etc.). Table 4.1 gives the results of the parameter estimates of Zeger's model using the fitted semivariogram to obtain $\hat{\rho}_e(1)$ and $\hat{\sigma}^2$ for each iteration. The results of the method of moments estimates are repeated here for comparison. Table 4.2 gives the correlations between the parameter estimates from the asymptotic variance-covariance matrix for $\hat{\beta}$ for both cases.

The results are very similar, with slight changes mainly in the standard error of the estimates. The main difference is in the estimates of $\hat{\rho}_e(1)$ and $\hat{\sigma}^2$ which is due to the use of more lags in the variogram case. The maximum number of lags could also be a fitting parameter which would probably further

TABLE 4.1

Parameter Estimates for Zeger Example, Time Dependent Case, Semivariogram (k=9) vs. Method of Moments

	Fitted Semivariogram		Method of Moments	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	0.17	0.14	0.17	0.15
Trend x 10^{-3}	-3.89	2.72	-3.80	2.94
$\cos(2\pi t/12)$	-0.12	0.11	-0.12	0.17
$\sin(2\pi t/12)$	-0.49	0.17	-0.49	0.18
$\cos(2\pi t/6)$	0.18	0.13	0.19	0.14
$\sin(2\pi t/6)$	-0.43	0.10	-0.43	0.14
$\hat{\sigma}^2$	0.67		0.77	
$\hat{\rho}_e(1)$	0.52		0.55	

TABLE 4.2

Correlation Matrices for Parameter Estimates for Zeger Example, Time Dependent Case, Semivariogram (k=9) vs. Method of Moments

a) Semivariogram

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\hat{\beta}_0$	1					
$\hat{\beta}_1$	0.0636	1				
$\hat{\beta}_2$	0.0413	0.0181	1			
$\hat{\beta}_3$	0.1318	0.0641	0.0919	1		
$\hat{\beta}_4$	-0.0717	0.0071	0.0020	-0.1157	1	
$\hat{\beta}_5$	0.1286	0.0364	0.1148	0.0640	-0.0123	1

b) Method of Moments

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\hat{\beta}_0$	1					
$\hat{\beta}_1$	0.0551	1				
$\hat{\beta}_2$	0.0340	0.0177	1			
$\hat{\beta}_3$	0.1165	0.0625	0.0841	1		
$\hat{\beta}_4$	-0.0657	0.0074	0.0016	-0.1070	1	
$\hat{\beta}_5$	0.1163	0.0354	0.1070	0.0604	-0.0112	1

improve the estimates. The number of lags used for this example was nine.

The correlation between most of the estimates is small, except for between $\hat{\beta}_0$ and $\hat{\beta}_3$ and $\hat{\beta}_5$ and between $\hat{\beta}_2$ and $\hat{\beta}_3$ and $\hat{\beta}_5$. There is a moderately large negative correlation between $\hat{\beta}_3$ and $\hat{\beta}_4$. The effect of using the semivariogram is to increase the correlation between the estimates slightly in most cases.

4.3 Variograms for Spatial Residuals

4.3.1 Introduction

In the transition from time series analysis to spatial data analysis, several new issues become important. The most obvious is the addition of one or more dimensions that needed to be accounted for in the analysis. With each dimension comes the possibility of anisotropy, i.e. the variogram depends on distance and direction, which was not a problem with time series. Further, unless the spatial data is on a uniform grid, the idea of a "lag" loses meaning and must be replaced with a differently defined geographic distance. The definition and measurement of distance may not be straightforward as reviewed in Chapter 2 (i.e. Rathbun, 1998).

In this section, two examples of spatial count data will be used to illustrate some of these issues as they affect the variogram for quasi-likelihood residuals. Alternate spatial covariance models are used to account for spatial autocorrelation in the examples of a) Upton & Fingleton's plant species data and b) Reynoldson's benthic data.

4.3.2 California Plant Species Counts

This example, originally from Johnson, *et al.* (1968), was reviewed in Chapter 2 and the data were modelled using quasi-likelihood in Chapter 3. The use of quasi-likelihood was shown to improve the prediction residual sum of squares (PRESS) compared to ordinary least squares (for cross-validation), but the goodness-of-fit was also improved (for a likelihood-ratio test) when spatial dependence was accounted for using a somewhat arbitrary spatial autocorrelation structure. It may be possible to further improve the model for this example data by fitting a semivariogram to the quasi-likelihood residuals as was done in Section 4.2.

As illustrated in Table 3.7, Upton and Fingleton use a 'two-part distance-based' weights matrix or **W** matrix to account for spatial interactions. The weights are based on the inverse square of the geographic distance

between two sites (equations 3.6-3.8). Knowing w_{ij} from Upton and Fingleton and selected d_{ij} from Johnson *et al.* (1968), it is possible to calculate all of the intersite distances, d_{ij}, from:

$$d_{ij} = \sqrt{(w_{ij} \sum_{j=1}^{n} w^{*}_{ij})^{-1}}$$

where the sum, $\sum_{j}^{*} w^{*}_{ij}$ ($\epsilon\ d_{ij} \leq k$) is calculated using any known distance from a site i to another site j and the corresponding w_{ij} (ie., $\sum_{j}^{*} w^{*}_{ij} = 1/w_{ij}d^{2}_{ij}$). The calculated d_{ij} may then be used to construct the complete intersite distance matrix, **D**, and then used with the residuals from the quasi-likelihood model for California plant species in Chapter 3 to estimate the variogram as in Section 4.2.

Figure 4.3 is the semivariogram for the quasi-likelihood residuals for the California plant species counts. The semivariogram was fit using the weighted least squares (equation 4.1) and a special case of the exponential model (from combining equations 2.4 and 2.6):

$$\gamma(\mathbf{h},\theta) = \theta_1 + \theta_2(1 - \exp(-\mathbf{h}\theta_3)).$$

Due to the variability of the species counts, the residuals are not stationary. As mentioned in Chapter 3, the residuals are not Gaussian either. In the maximum likelihood case, the autocorrelation was estimated to be negative. In this situation, Cressie (1993, p. 62) recommends that the following wave

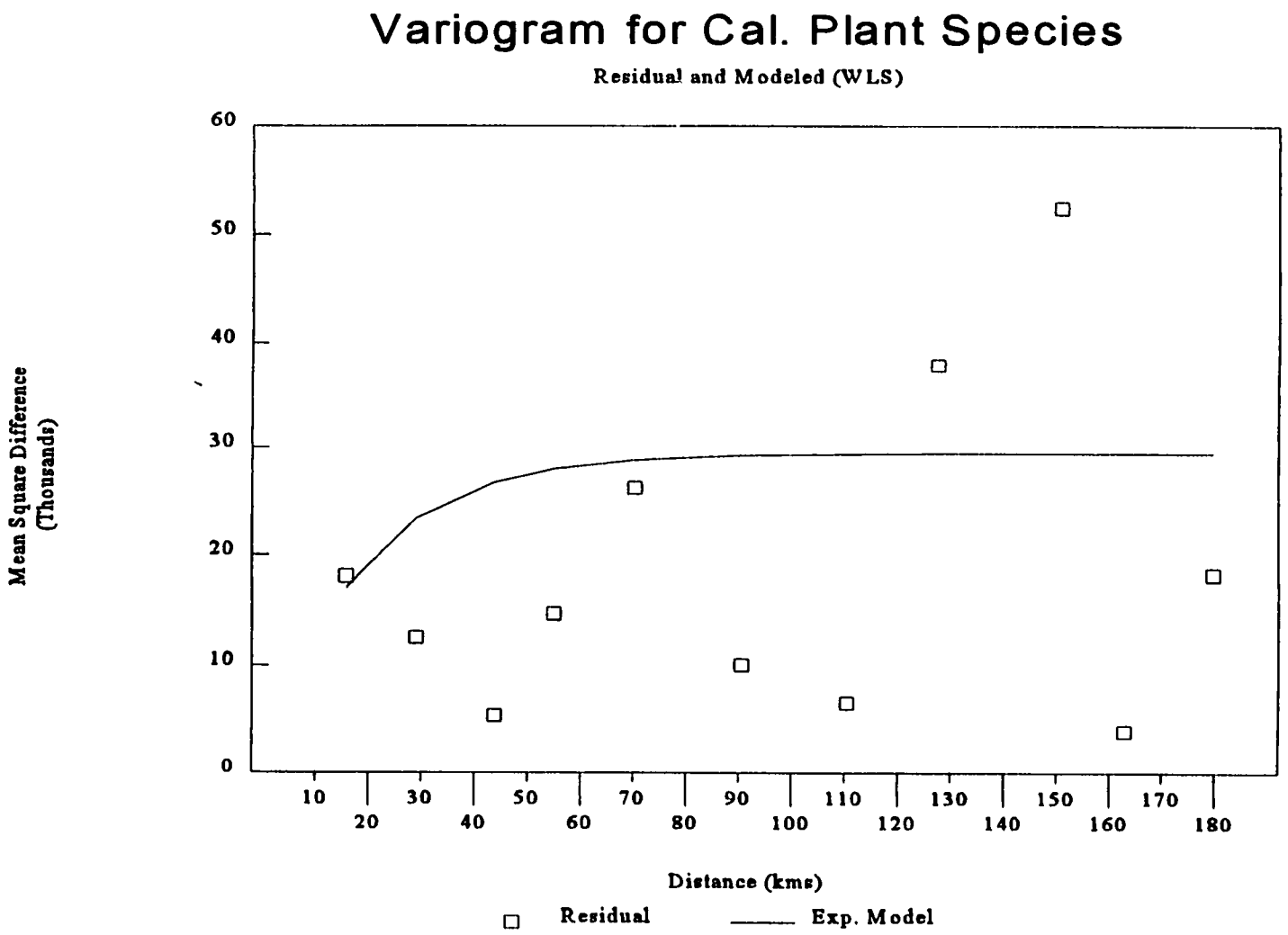

FIGURE 4.3 Variogram for California Plant Species, Exponential Model

model be fit to the semivariogram:

$$\gamma(\mathbf{h},\theta) = \theta_1 + \theta_2(1 - \theta_3 \sin(\mathbf{h}/\theta_3)/\mathbf{h}).$$

This results in the semivariogram shown in Figure 4.4. Using the wave model to fit the semivariogram allows the quasi-likelihood parameters, $\hat{\beta}$, to be estimated more precisely (Table 4.3). Comparing these results to the independent case from Chapter 3 (Table 3.8), there is very little difference among the $\hat{\beta}$s. For the exponential model, some of the standard errors seem to be quite large. Table 4.4 shows the correlation matrix for the $\hat{\beta}$s for the exponential and wave models. There are extremely large negative correlations between $\hat{\beta}_0$ and $\hat{\beta}_3$ and between $\hat{\beta}_1$ and $\hat{\beta}_2$ for both models. This is the same type of correlation structure observed by Upton and Fingleton for their OLS model and is due to colinearity in the data and confounding of the explanatory variables. From a model selection viewpoint, it may be advisable to drop the maximum altitude variable, $\mathbf{x}_2$. However, the variables from the original example have been retained for comparison purposes.

Cressie (1993, pp. 64-66) discusses the relative variogram in which a spatial process is only stationary within limited subregions. He notes that if a simple relationship of mean to variogram exists, it may be possible to combine estimates. One such relationship has been described by Journel and Huijbregts

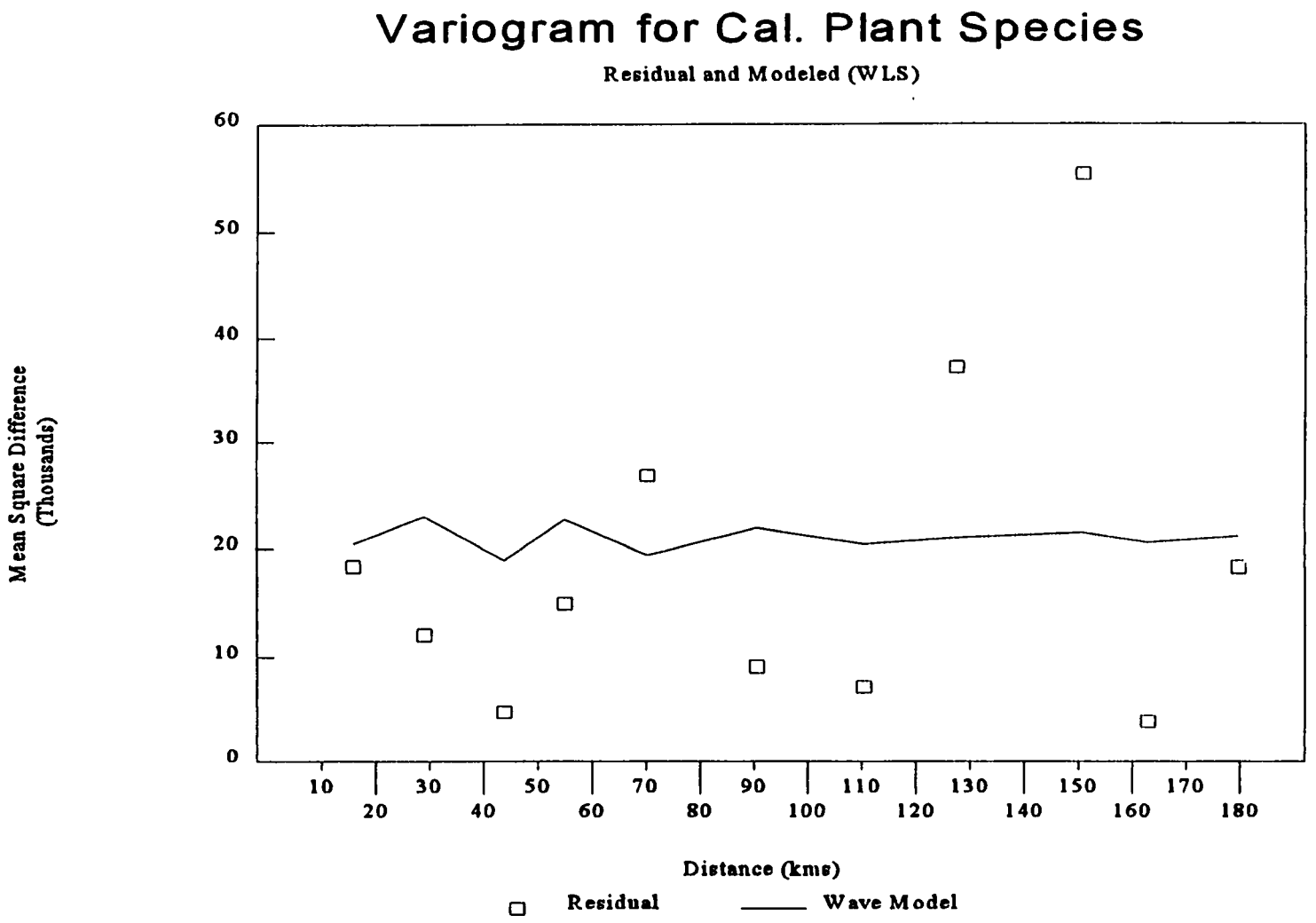

FIGURE 4.4 Variogram for California Plant Species, Wave Model

TABLE 4.3

Parameter Estimates for Upton and Fingleton Example Using Quasi-Likelihood with Spatial Dependence and Exponential and Wave Models for Semivariogram

	Exponential Case		Wave Case	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-12.744	2.951	-13.221	2.267
log(Area)	0.344	0.050	0.330	0.046
log(Max. Alt.)	0.031	0.142	0.075	0.129
log(Latitude)	4.794	0.754	4.854	0.275
θ_1	0		0	
$\delta^2=(\hat{\theta}_2-\bar{\hat{\mu}})/\overline{\hat{\mu}^2}$	0.072		0.050	
$\hat{\theta}_3$	0.054		5.119	

NOTE: $\bar{\hat{\mu}} = \sum_{i=1}^{n} \hat{\mu}_i/n$

$\overline{\hat{\mu}^2} = \sum_{i=1}^{n} \hat{\mu}_i^2/n$

TABLE 4.4

Correlation Matrices for Parameter Estimates for Upton and Fingleton Example Using Quasi-Likelihood with Spatial Dependence and Exponential and Wave Models for Semivariogram

a) Exponential Model

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\hat{\beta}_0$	1			
$\hat{\beta}_1$	0.5156	1		
$\hat{\beta}_2$	-0.4964	-0.9446	1	
$\hat{\beta}_3$	-0.9569	-0.2710	0.2245	1

b) Wave Model

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\hat{\beta}_0$	1			
$\hat{\beta}_1$	0.5472	1		
$\hat{\beta}_2$	-0.5392	-0.9434	1	
$\hat{\beta}_3$	-0.9483	-0.2830	0.2456	1

(1978, pp. 187-190) and is called the proportional effect. They use this model to define the relative variogram:

$$2\gamma_Z^{(j)}(\mathbf{h})/\mu_j^2$$

where Z is the spatial process of interest

j indexes the subregions that are to be combined

μ_j is the mean of the process within subregion j.

If the true mean is replaced by the estimated mean in the above definition of the relative variogram, then a variogram of normalized residuals can be estimated from the residuals defined in Section 3.4.2 that is analogous to Journel and Huijbregts' definition. So the variogram of the stationary process, Y, combined over various subregions where the nonstationary process, Z, is in effect is:

$$2\gamma_Y(\mathbf{h}) \approx \text{var}(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s}))/(\hat{E}(Z(\mathbf{s})))^2$$

where $\hat{E}(Z(\mathbf{s})) = \hat{\mu} = \exp(\mathbf{X}\hat{\beta})$.

These estimated means depend on **s** through **X**, the vector of explanatory variables.

Figures 4.5 and 4.6 are "relative" semivariograms of the normalized quasi-likelihood residuals using the exponential and wave models, respectively.

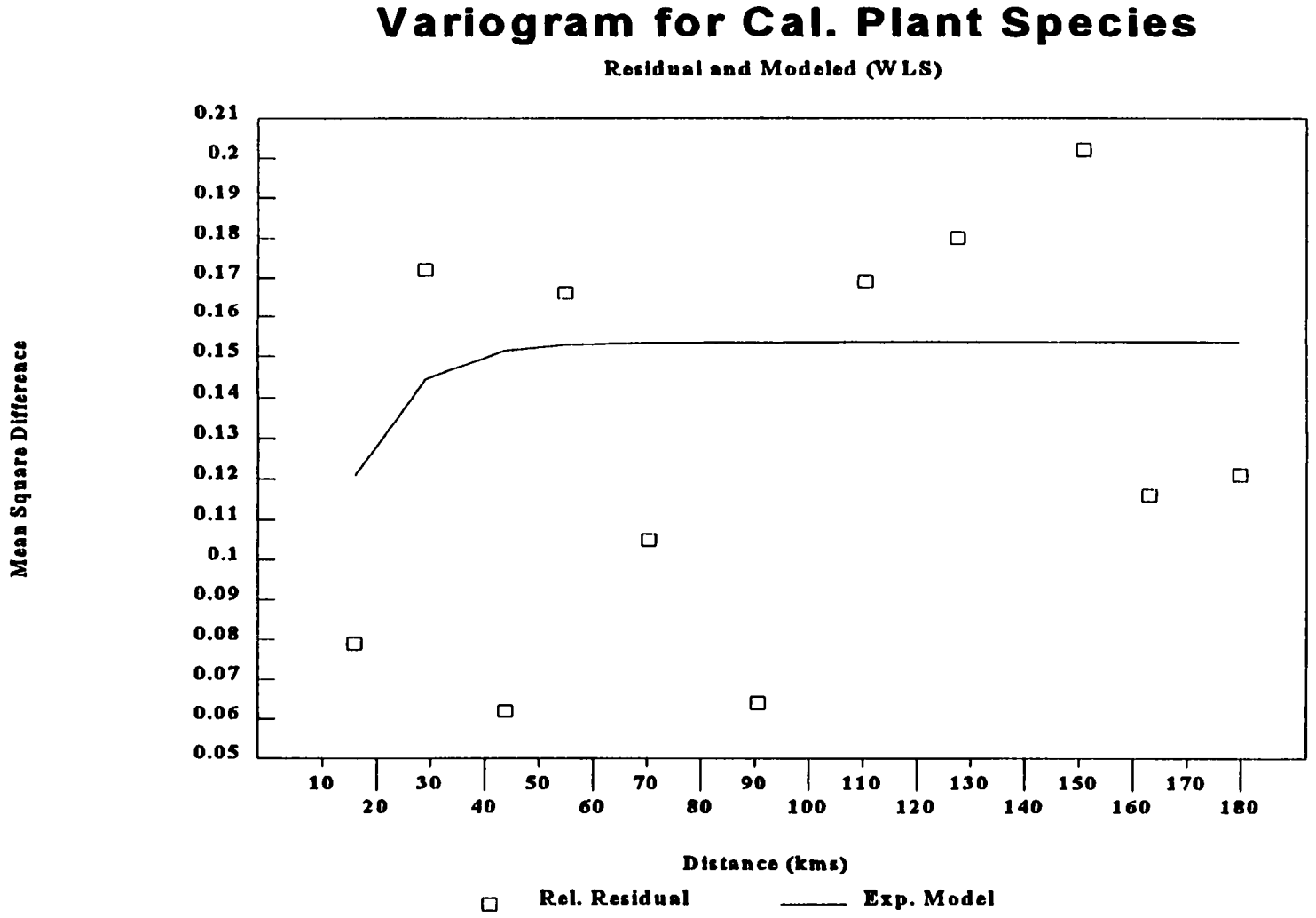

FIGURE 4.5 Variogram for California Plant Species, Relative Residuals, Exponential Model

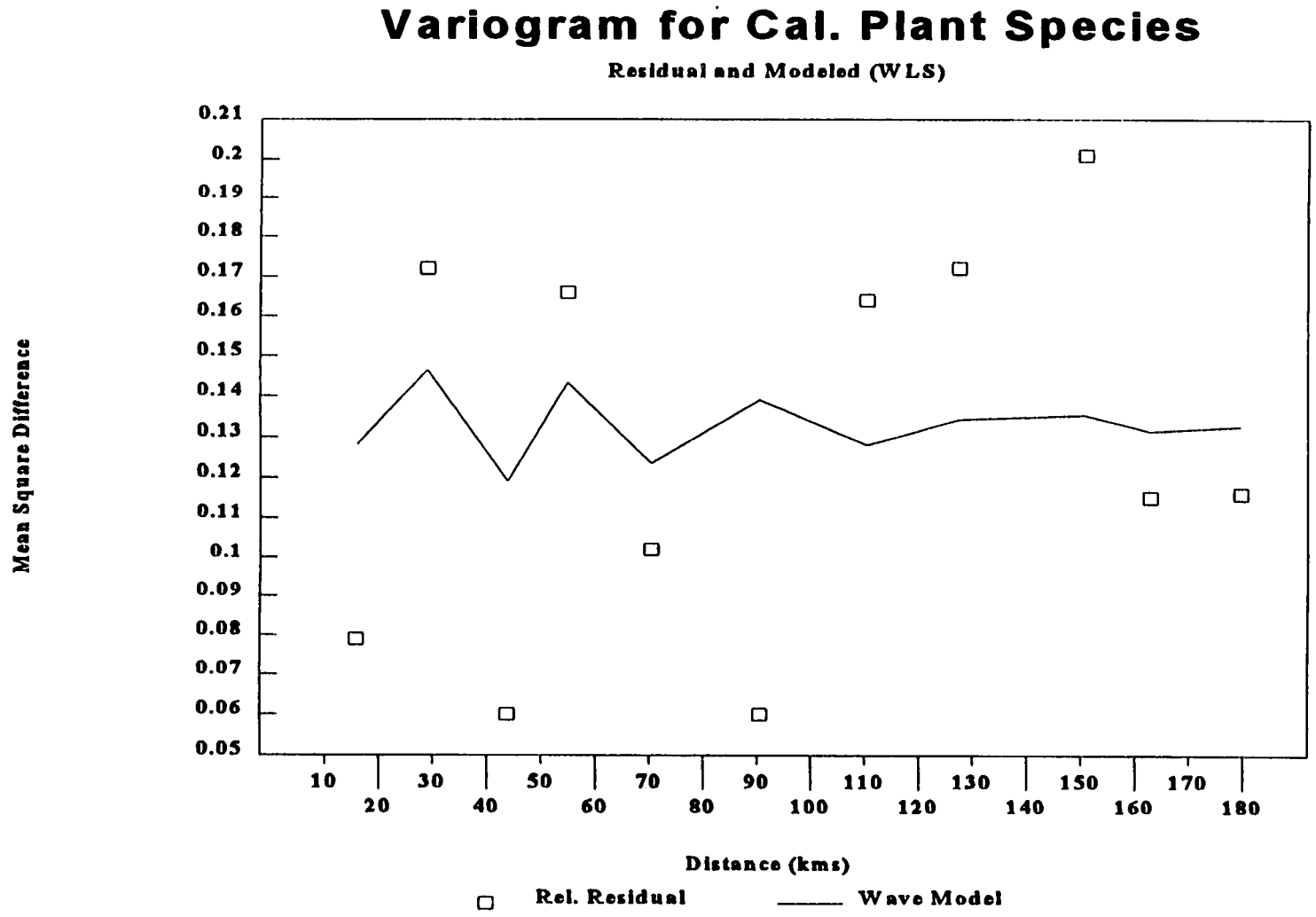

FIGURE 4.6 Variogram for California Plant Species, Relative Residuals, Wave Model

The quasi-likelihood parameter estimates are shown in Table 4.5 and the correlation matrices for these estimates are shown in Table 4.6. The estimates are highly correlated as in Table 4.4. The wave model for the semivariogram (Table 4.2) seems to give a slightly better fit, but the models can be best evaluated by cross-validation (Chapter 5).

4.3.3 Benthic Invertebrate Counts

4.3.3.1 **Introduction**

The example data set described in Sections 1.4 and 2.3.4 provides many interesting case studies of spatial dependence of counts of environmentally significant benthic species. In this section, the quasi-likelihood model of Chapter 3 and the variogram models of the previous sections will be applied to two sets of benthic species counts to illustrate some of the issues involved in spatial analysis of environmental counts data. Reynoldson, *et al.* (1995) have identified explanatory variables that affect benthic community structure at reference sites in the Great Lakes. Of the over 300 reference sites established in this project, 79 are located in the Georgian Bay and North Channel regions of Lake Huron (Figures 4.7 and 4.8). Since the site density of

TABLE 4.5

Parameter Estimates for Upton and Fingleton Example Using Quasi-Likelihood with Spatial Dependence and Exponential and Wave Models for "Relative" Semivariogram

	Exponential Case		Wave Case	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-13.209	3.845	-13.373	3.235
log(Area)	0.338	0.073	0.327	0.068
log(Max. Alt.)	0.054	0.204	0.084	0.187
log(Latitude)	4.884	0.956	4.881	0.754
θ_1	0		0	
$\sigma^2=\hat{\theta}_2-\overline{1/\hat{\mu}}$	0.151		0.121	
$\hat{\theta}_3$	0.098		5.171	

NOTE: $\overline{1/\hat{\mu}} = 1/n \sum_{i=1}^{n} 1/\hat{\mu}_i$

TABLE 4.6

Correlation Matrices for Parameter Estimates for Upton and Fingleton Example Using Quasi-Likelihood with Spatial Dependence and Exponential and Wave Models for "Relative" Semivariogram

a) Exponential Model

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\hat{\beta}_0$	1			
$\hat{\beta}_1$	0.5472	1		
$\hat{\beta}_2$	-0.5392	-0.9434	1	
$\hat{\beta}_3$	-0.9483	-0.2830	0.2456	1

b) Wave Model

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\hat{\beta}_0$	1			
$\hat{\beta}_1$	0.6343	1		
$\hat{\beta}_2$	-0.6457	-0.9407	1	
$\hat{\beta}_3$	-0.9431	-0.3765	0.3564	1

GREAT LAKES REFERENCE SITES
Georgian Bay

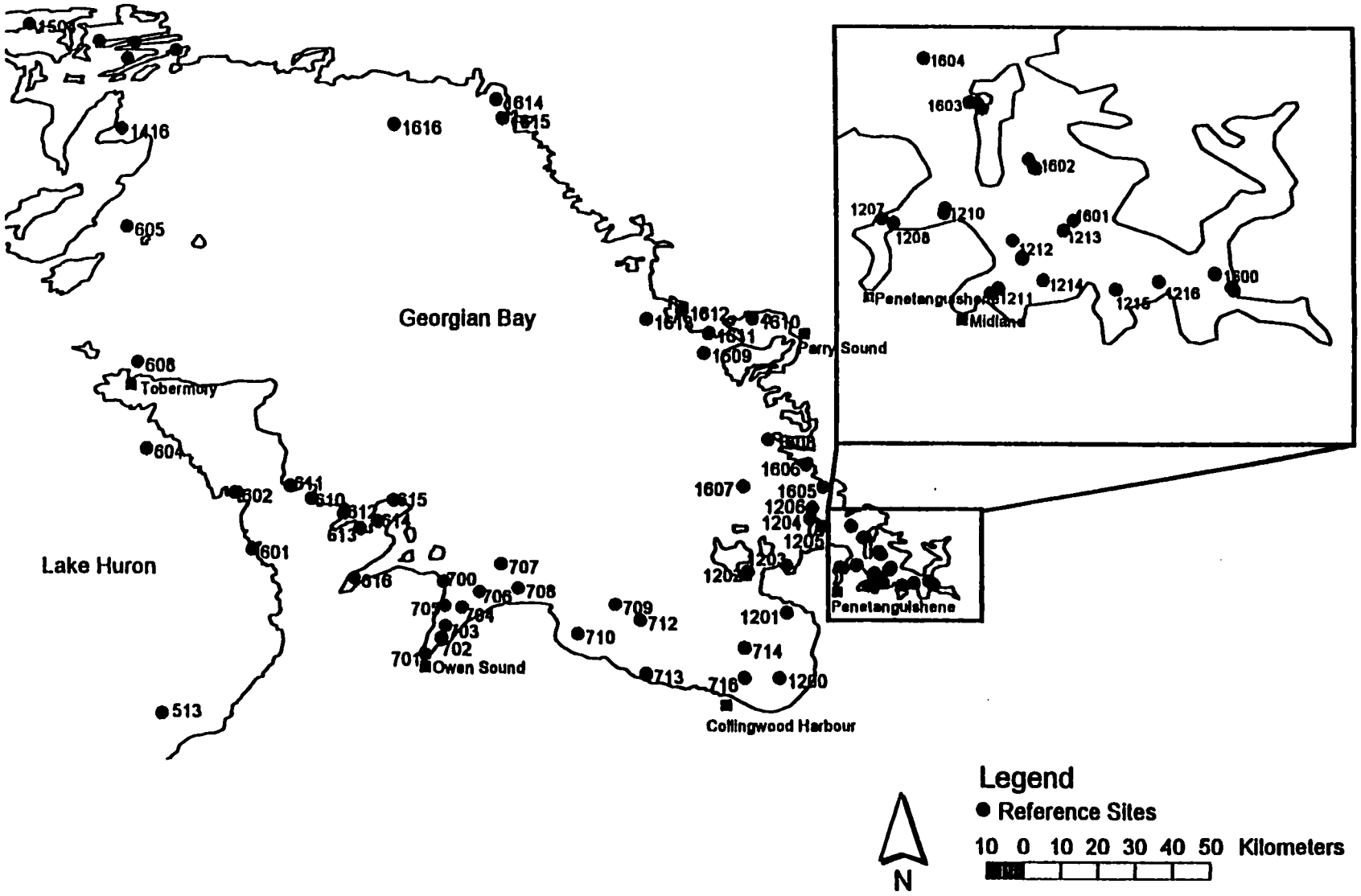

FIGURE 4.7 Georgian Bay Reference Sites

GREAT LAKES REFERENCE SITES

North Channel

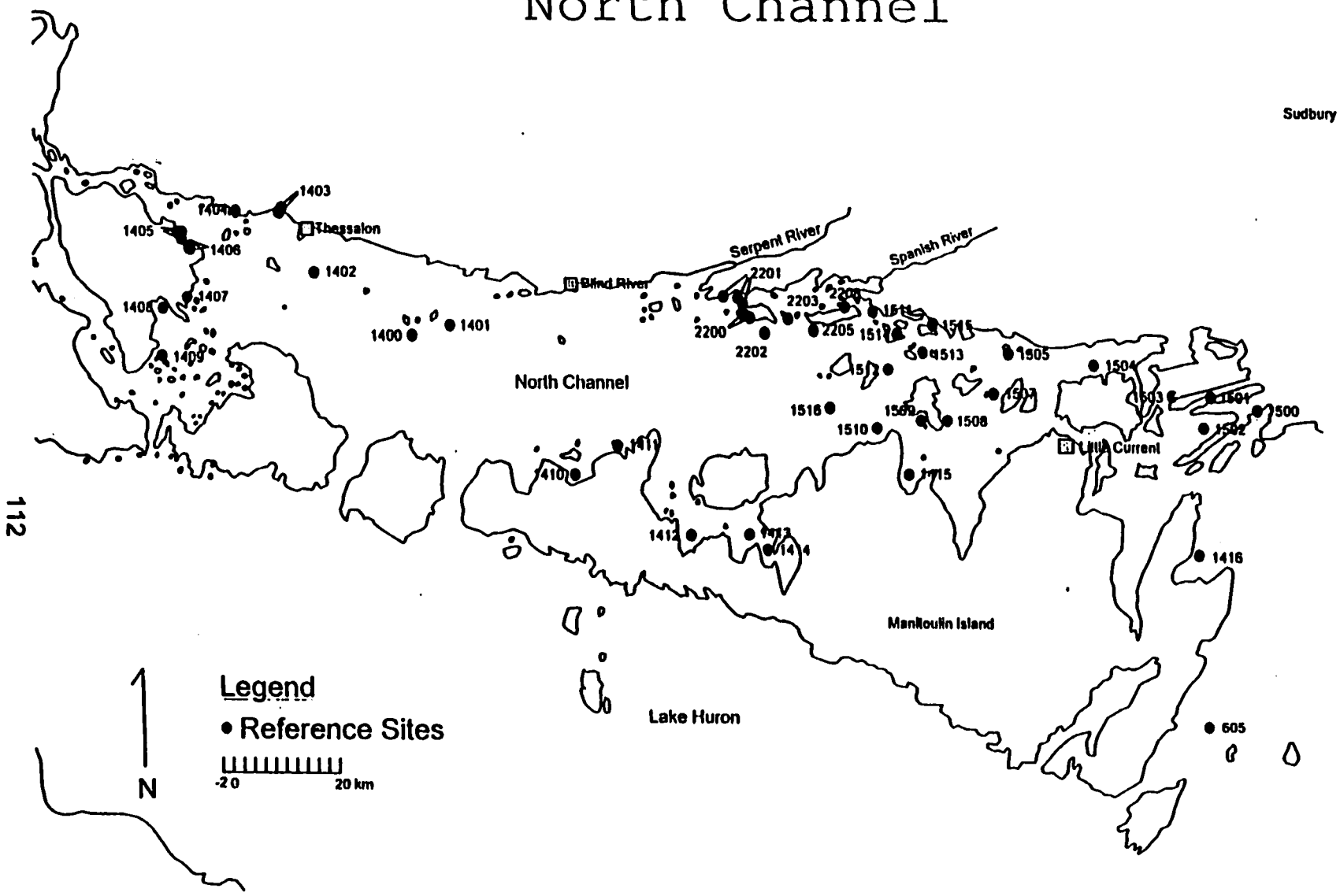

FIGURE 4.8 North Channel Reference Sites

some areas of these regions is high, spatial dependence of observed counts among sites was likely.

4.3.3.2 **Independent Case**

Two species of considerable ecological importance, *Diporeia hoyi* and *Procladius* were modelled separately to check the appropriateness of the procedures being considered. Histograms of mean counts (n=5, see Section 1.4) for sites in Georgian Bay and the North Channel are presented in Figures 4.9 and 4.10. First, the quasi-likelihood model from Chapter 3 was applied assuming spatial independence. Explanatory variables were selected using a stepwise procedure from among several physical and chemical measurements. The results are reported in Table 4.7. The resultant models are intuitively reasonable. Of the explanatory variables in Table 4.7, depth is probably the most directly related to cause and effect, while the rest are surrogates for geochemical and aquatic conditions (Reynoldson, personal communication). *Diporeia hoyi* prefer deeper, cooler waters, while *Procladius* tend to be found at shallower, warmer sites. The residuals for these models are presented in Figures 4.11 and 4.12. Although the residuals appear symmetric, they are not Gaussian. This is due to the heavy tails on both sides of the center of the histogram. These tails are caused by anomalous values which are outliers in the sense that they do not fit

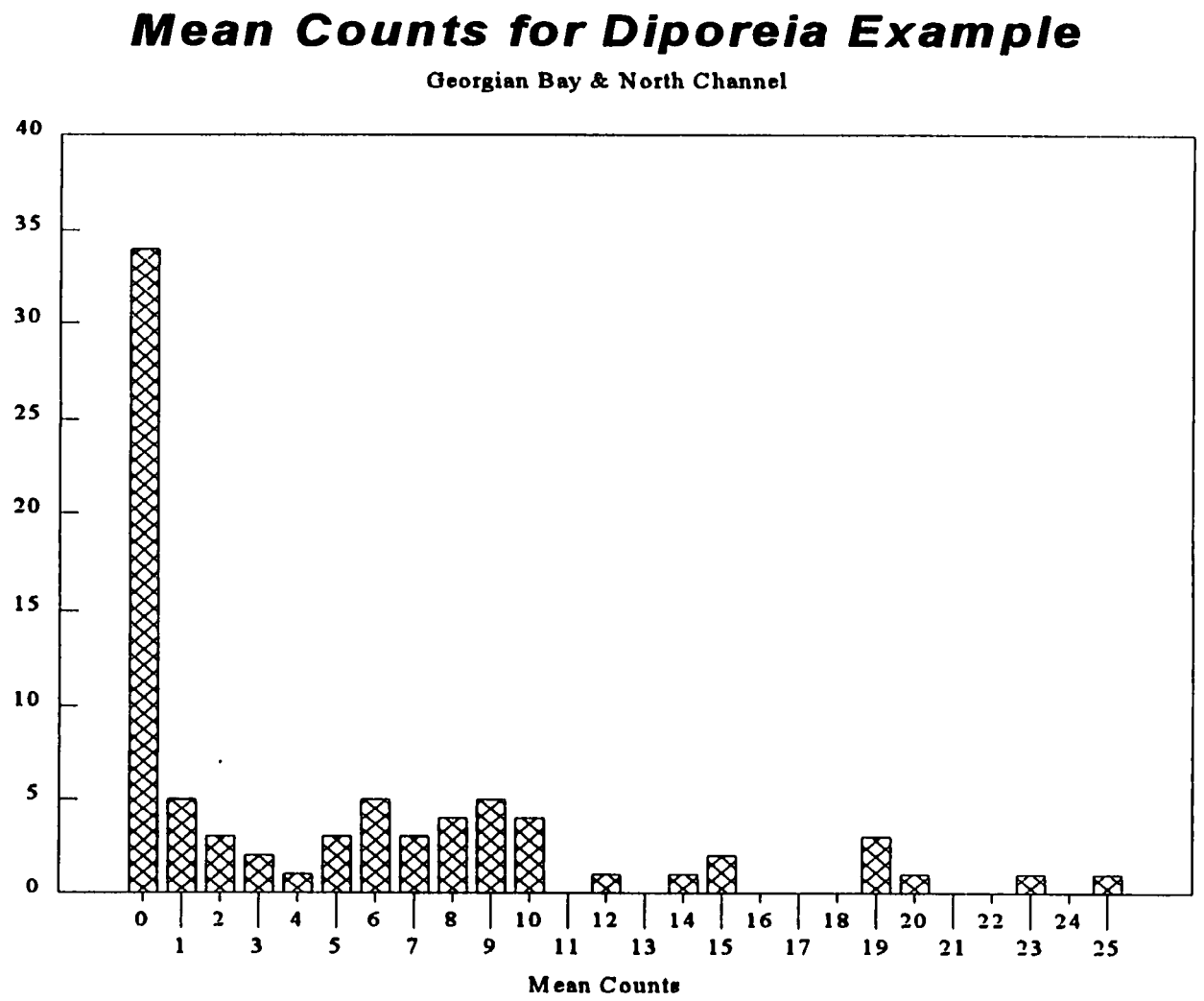

FIGURE 4.9 Histogram of Mean Counts for *Diporeia* Example

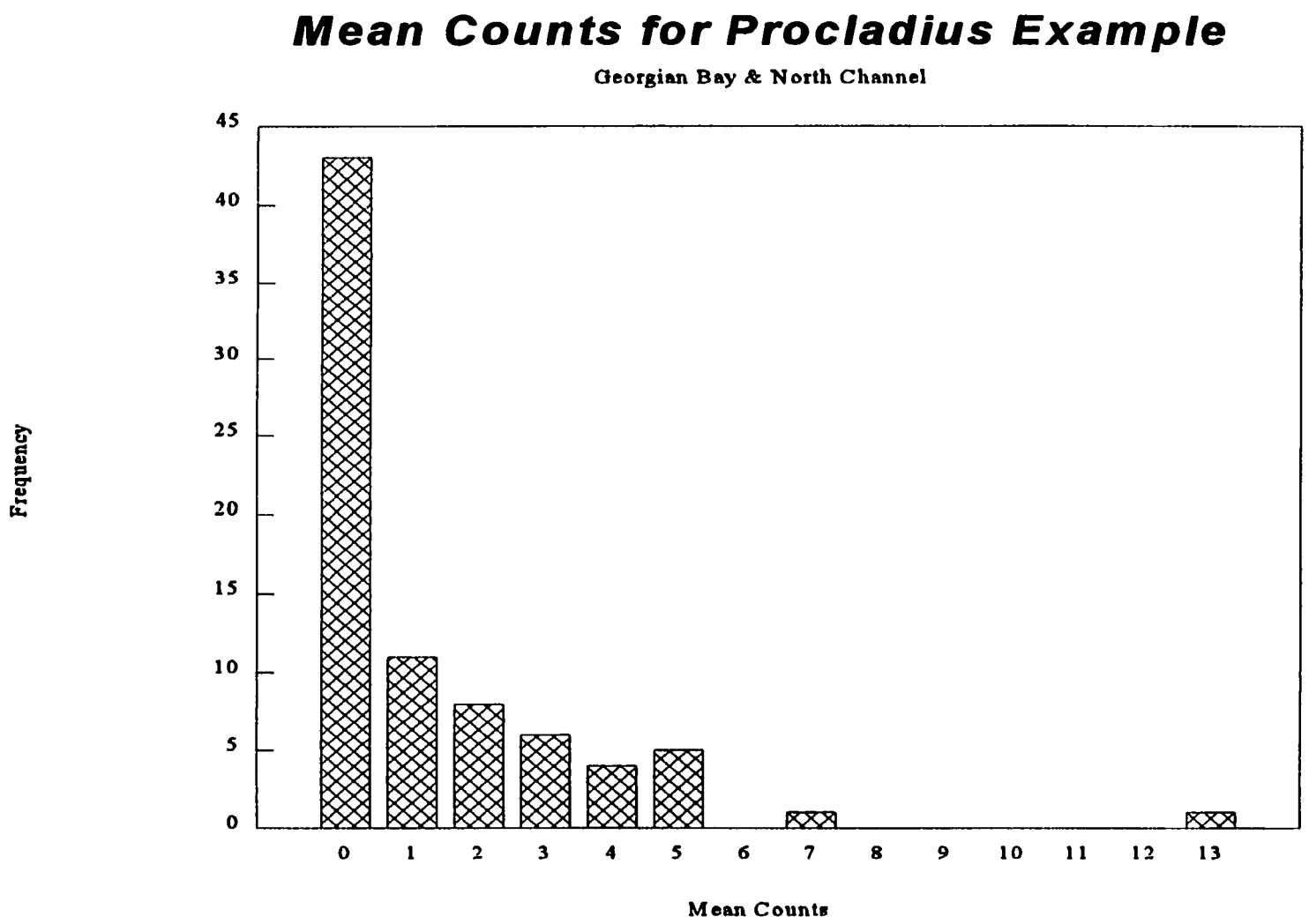

FIGURE 4.10 Histogram of Mean Counts for *Procladius* Example

TABLE 4.7

Parameter Estimates for *Diporeia hoyi* and *Procladius* Using Quasi-Likelihood and Assuming Spatial Independence

	Diporeia hoyi		*Procladius*	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	10.39	3.91	-3.53	2.05
log(Depth)	0.866	0.19	-1.36	0.20
$\log(Al_2O_3)$ (Aluminum Oxide)	-2.50	0.67	3.03	0.87
$\log(K_2O)$ (Potassium Oxide)	3.01	0.74	-	-
log(alkalinity)	-1.98	0.89	-	-
$\hat{\sigma}^2$	0.688		0.40	

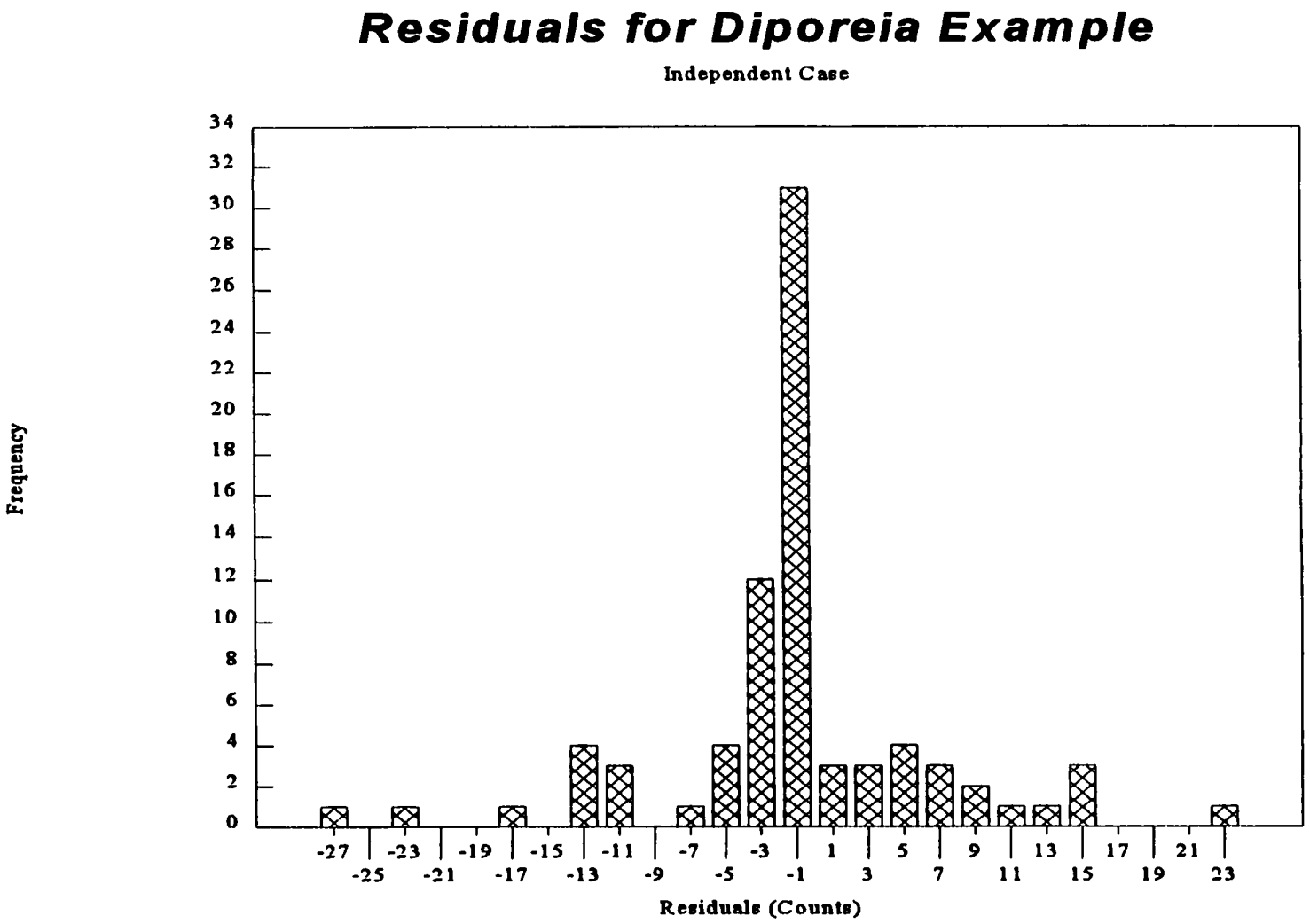

FIGURE 4.11 Histogram of Residuals for *Diporeia* Example

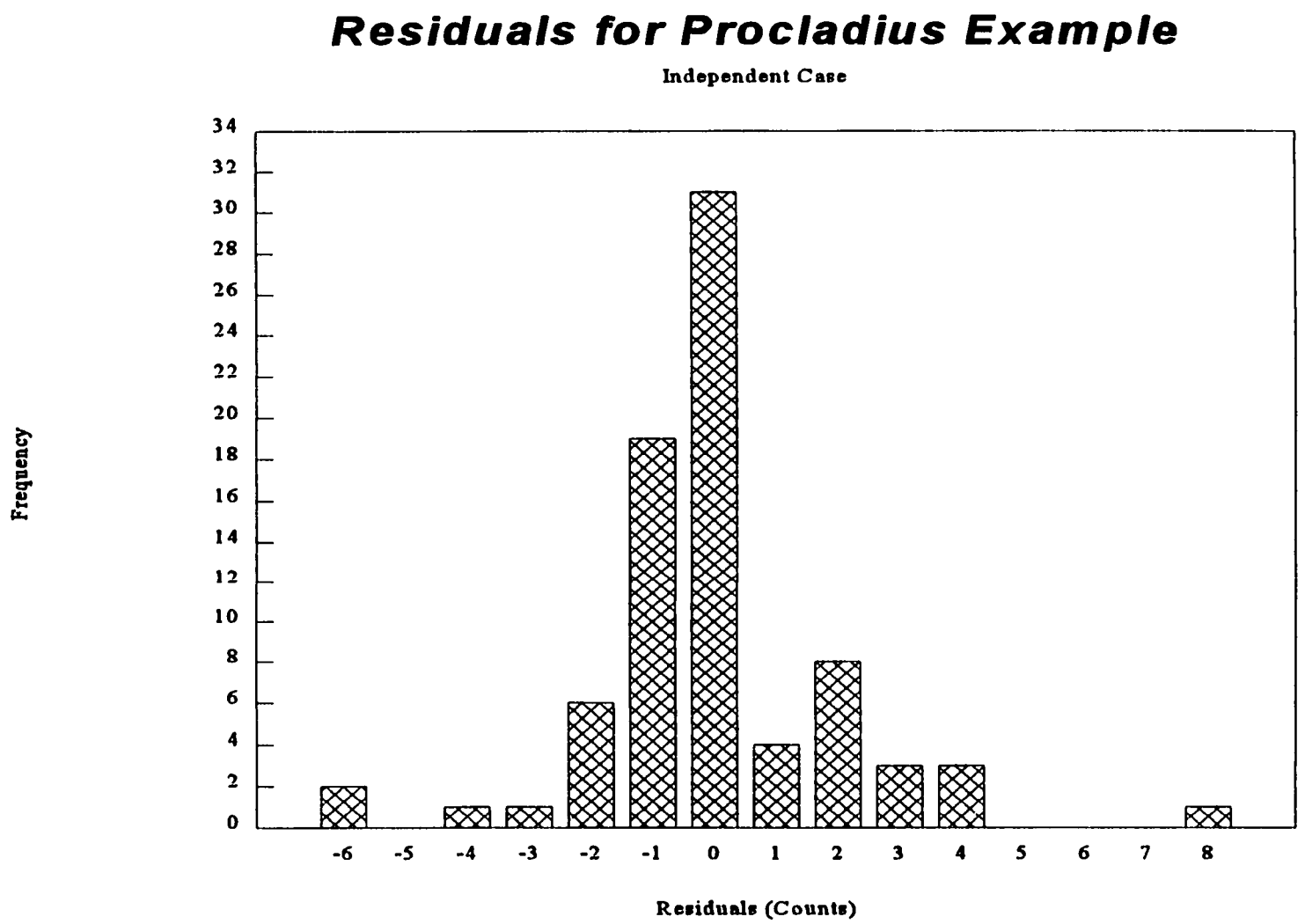

FIGURE 4.12 Histogram of Residuals for *Procladius* Example

the model (Figures 4.13 and 4.14). However, they are real observations for which the abundance of the benthic species does not agree with that predicted by the model. Robust methods may be required to handle these observations in a satisfactory manner.

Table 4.8 shows the correlations between the parameter estimates for the independent case. As in the California plant species example, some of the parameter estimates are highly correlated. For *Diporeia hoyi*, $\hat{\beta}_0$ and $\hat{\beta}_4$ and $\hat{\beta}_2$ and $\hat{\beta}_3$ are highly correlated. For *Procladius*, $\hat{\beta}_0$ and $\hat{\beta}_2$ are highly correlated. Once again, the explanatory variables are probably confounded and the data are colinear. The high correlations with the intercept parameter may indicate that it is not needed in the model. Note the relatively low correlations with $\hat{\beta}_1$, the depth coefficient, which intuitively would be independent of geochemical or water quality effects for reference sites. If the focus of this research was on variable selection, it may be necessary to find explanatory variables that are more orthogonal.

4.3.3.3 **Spatially Dependent Benthic Counts - Direction**

Besides the explanatory variables for the 79 sites in Figures 4.7 and 4.8, accurately determined latitude and longitude (via the global positioning

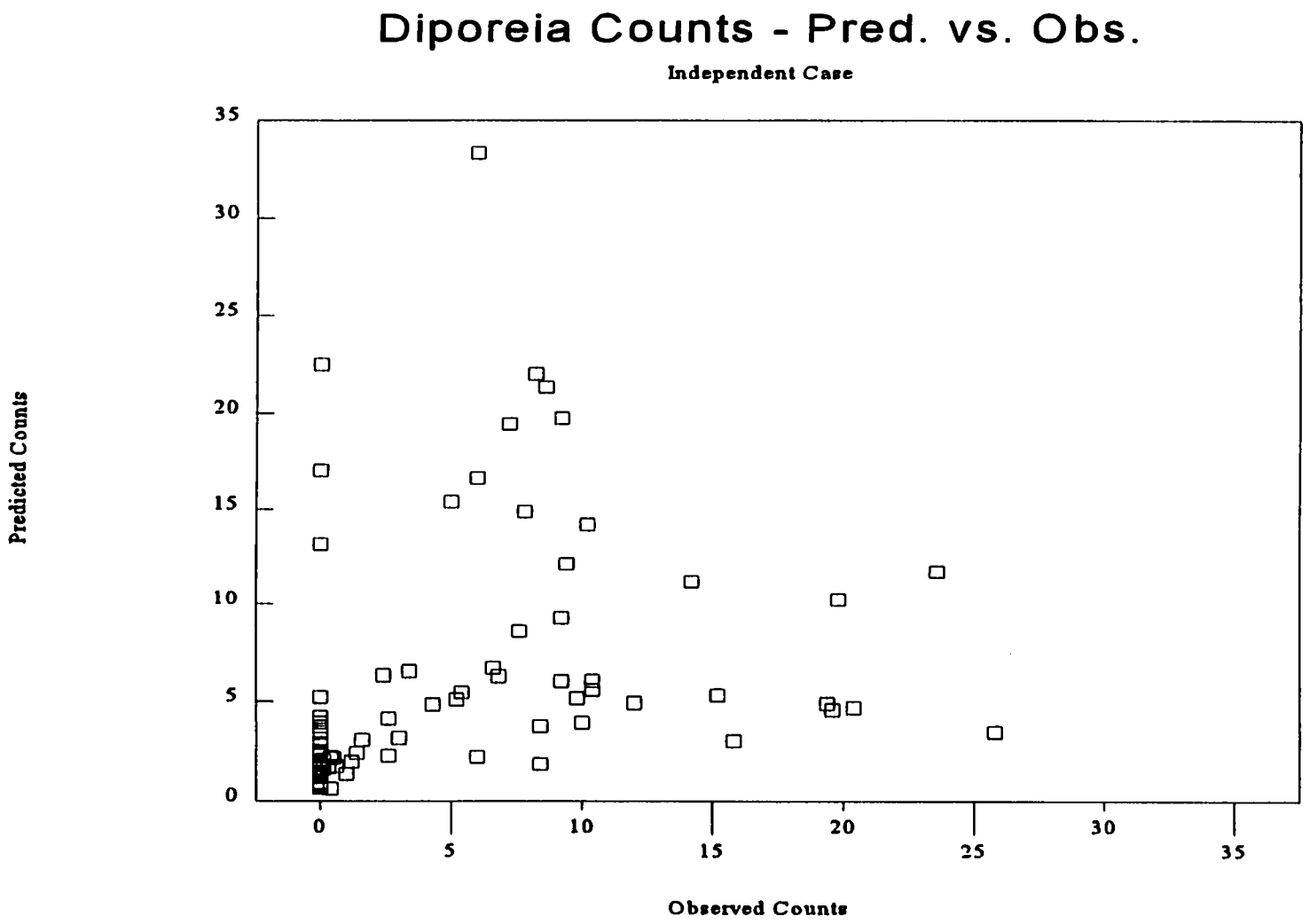

FIGURE 4.13 Scatterplot of Predicted vs. Observed Mean *Diporeia* Counts

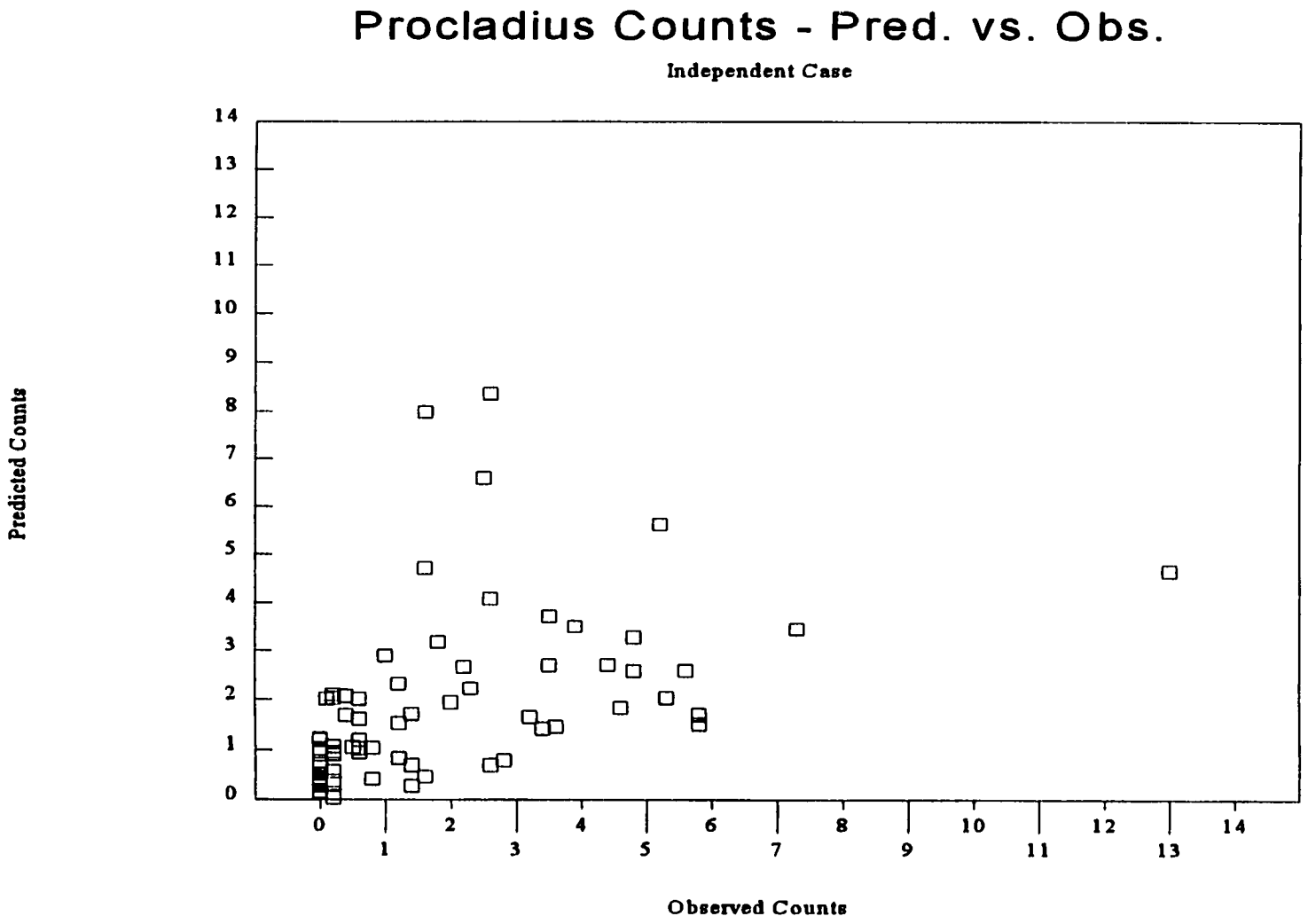

FIGURE 4.14 Scatterplot of Predicted vs. Observed Mean *Procladius* Counts

TABLE 4.8

Correlation Matrices for Parameter Estimates for *Diporeia hoyi* and *Procladius* Using Quasi-Likelihood and Assuming Spatial Independence

a) *Diporeia hoyi*

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_0$	1				
$\hat{\beta}_1$	0.1346	1			
$\hat{\beta}_2$	-0.4050	0.0583	1		
$\hat{\beta}_3$	0.2778	-0.4306	-0.6240	1	
$\hat{\beta}_4$	-0.9390	-0.2288	0.0991	-0.1533	1

b) *Procladius*

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
$\hat{\beta}_0$	1		
$\hat{\beta}_1$	0.1145	1	
$\hat{\beta}_2$	-0.9705	-0.3441	1

system - GPS) are available. This allows the construction of the intersite distance matrix, **D**, directly by converting the coordinates to distances:

$$d_{ij} = \sqrt{\Delta_{lat}^2 + \Delta_{long}^2}$$

where $\Delta_{lat} = (lat_i - lat_j) * 111.198$

and $\Delta_{long} = (long_i - long_j) * 78.867$.

The constants in the above equation convert degrees of latitude (lat) or longitude (long) to kilometres.

Further, this location information can be used to check for anisotropy of the variogram by sorting pairs of sites according to their compass orientation (ie. N-S vs. E-W). In this section, two pairs of orientations will be checked for anisotropy: N-S vs. E-W and NW-SE vs. NE-SW. Semivariograms are estimated using only site pairs that have the orientation of interest. This is determined as follows.

For N-S vs. E-W, the absolute value of the ratio, $\Delta_{lat}/\Delta_{long}$ is compared to 1.0. If the value is greater than 1.0, then the N-S distance between sites is greater than the E-W distance and the pair of sites is considered to have a N-S orientation. If the value is less than 1.0, the opposite is the case. For NW-SE vs. NE-SW, the sign of the ratio, $\Delta_{lat}/\Delta_{long}$ is examined. If the sign is negative, the pair of sites is considered to have a NW-SE orientation (ie. in the

second or fourth quadrants of the unit circle), while if the sign is positive, the opposite is true.

Figures 4.15-4.18 are the estimated semivariograms for *Diporeia hoyi* counts using the different orientation of site pairs. For the first three orientations (Figures 4.15-4.17), the estimated semivariograms and the fits using the exponential model are quite similar. This is an indication that anisotropy is not pronounced. For the NE-SW orientation (Figure 4.18), the exponential model could not be fit to the estimated semivariogram because the outliers have too much weight. This is a problem for checks of this nature, because dividing the observation pairs into two separate groups decreases the number of pairs at each distance and increases the importance of outliers. Table 4.9 gives the number of pairs for each distance group by orientation. Journel and Huijbregts (1978, p. 194) recommend that the number of pairs in each distance group be at least 30. As Table 4.9 shows, this is not usually achieved for these semivariograms. It was not possible to consistently fit the same exponential model to the estimated semivariograms for all of the above orientations.

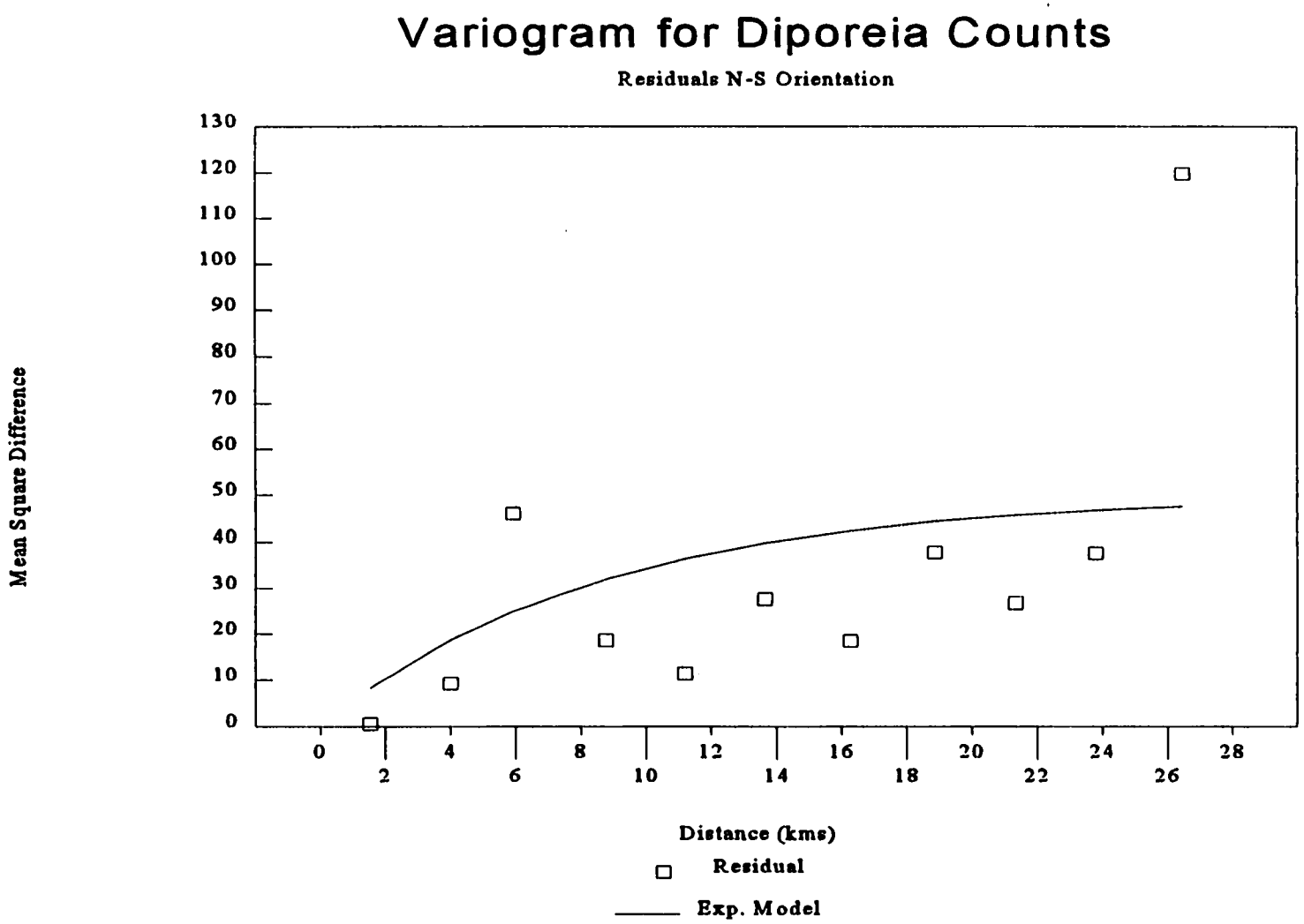

FIGURE 4.15 *Diporeia* Residuals Variogram (North-South Direction)

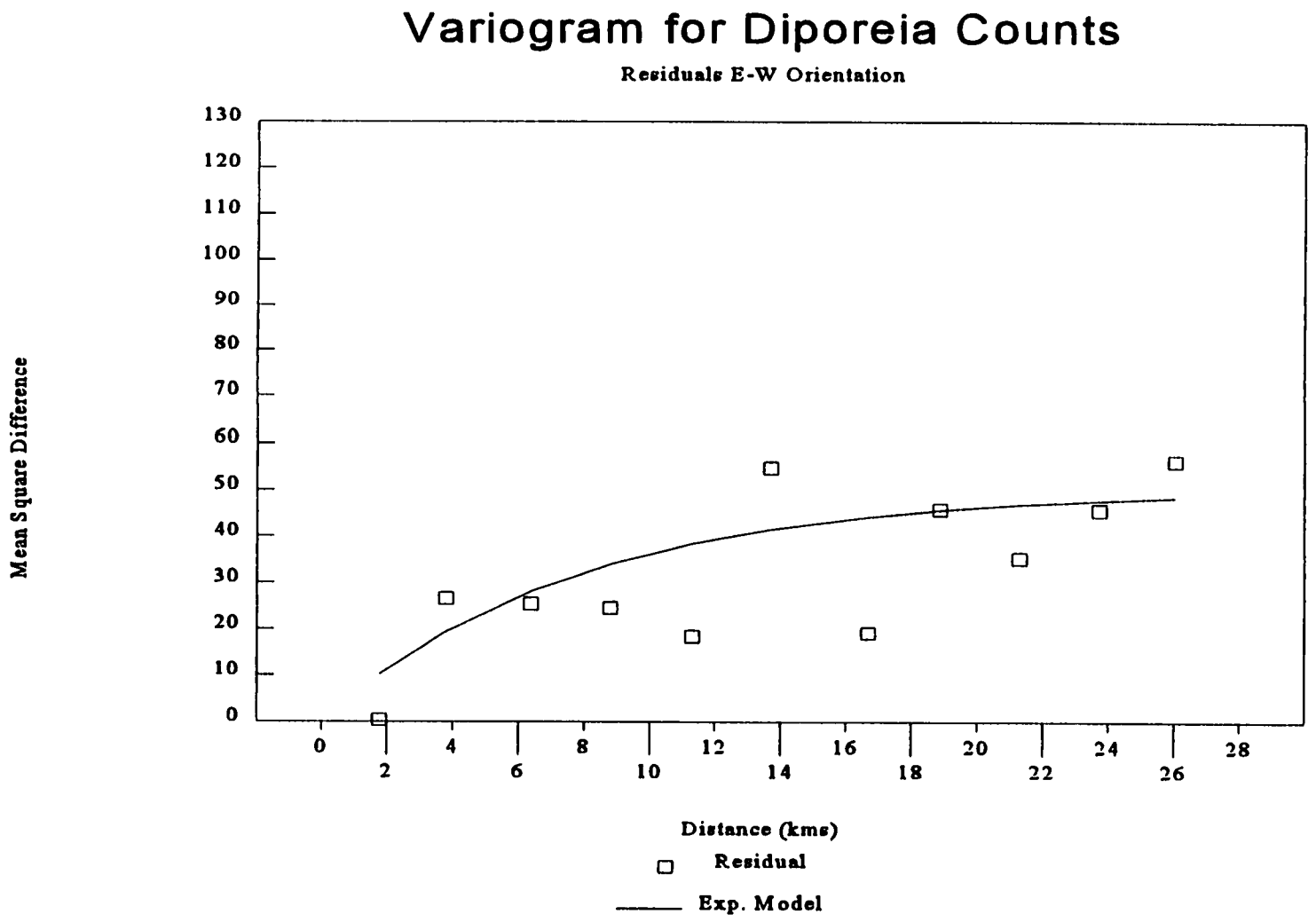

FIGURE 4.16 *Diporeia* Residuals Variogram (East-West Direction)

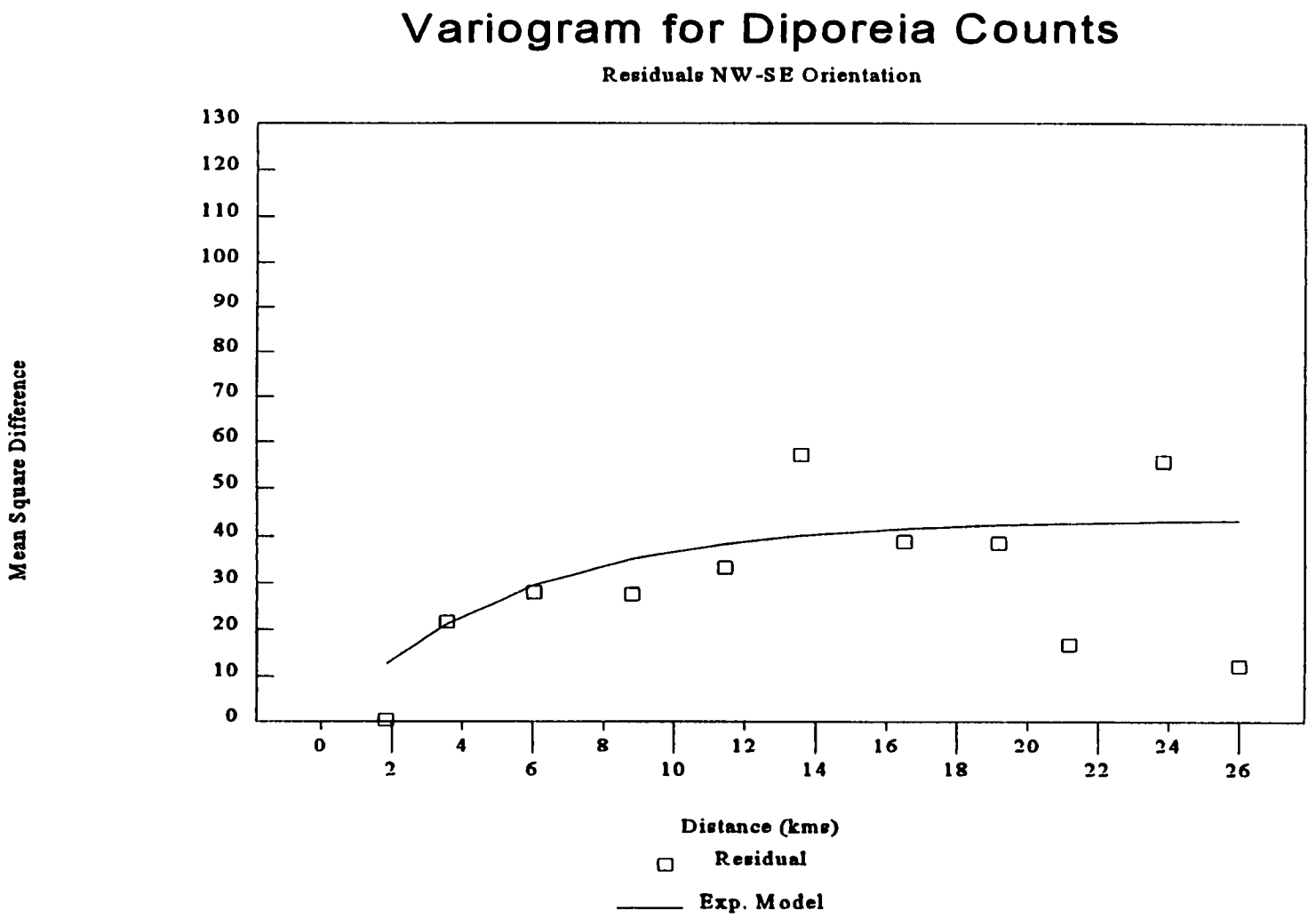

FIGURE 4.17 *Diporeia* Residuals Variogram (Northwest-Southeast Direction)

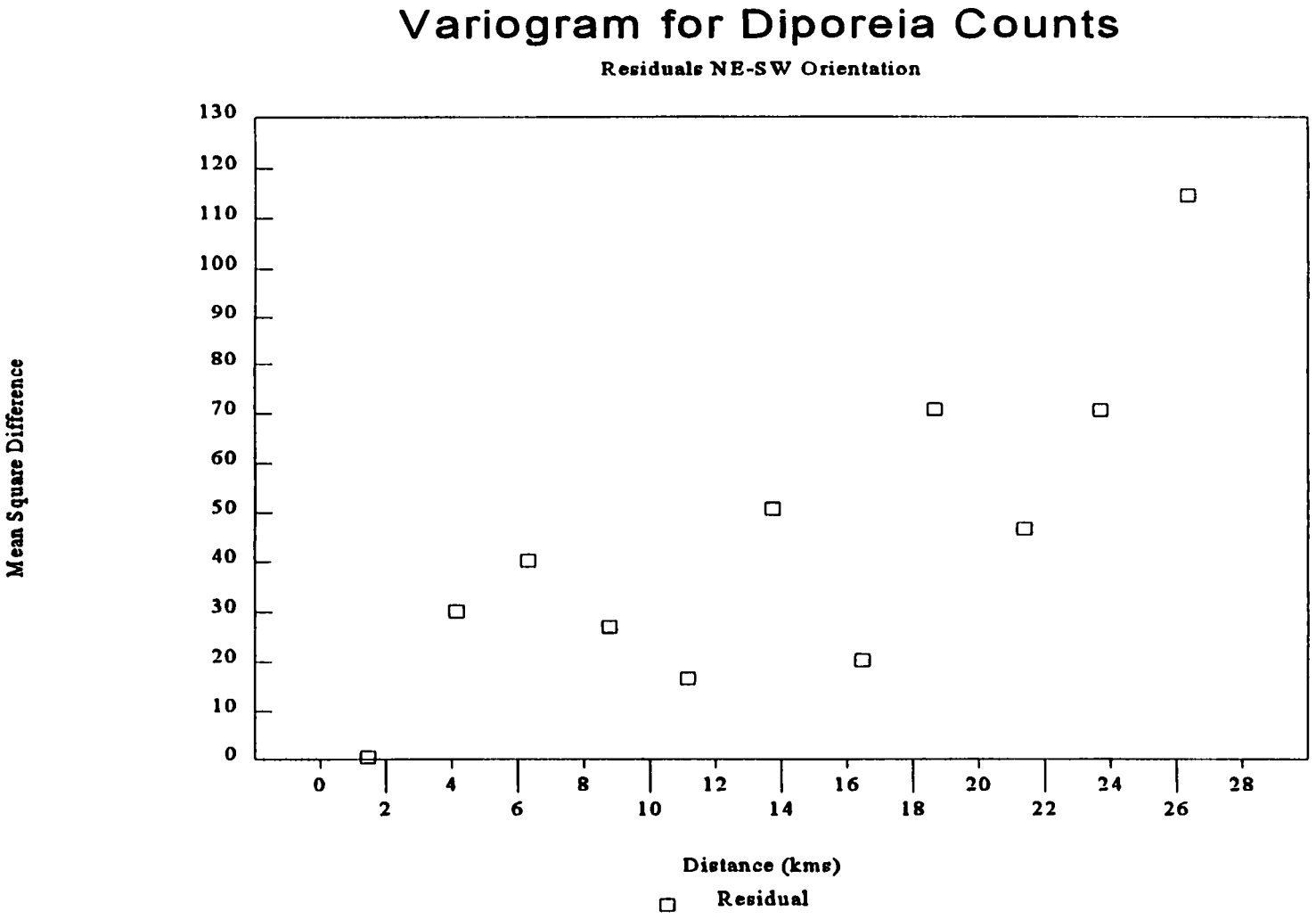

FIGURE 4.18 *Diporeia* Residuals Variogram (Northeast-Southwest Direction)

TABLE 4.9

Number of Site Pairs for Semivariogram Estimation by Orientation

Distance (km)	# N-S	# E-W	TOTAL	# NW-SE	# NE-SW	TOTAL
2	2	3	5	3	2	5
4	14	15	29	12	17	29
6	21	29	50	22	28	50
8	25	25	50	27	23	50
11	21	18	39	12	27	39
13	19	31	50	15	35	50
16	21	21	42	17	25	42
18	12	23	35	13	22	35
21	17	19	36	10	26	36
23	11	21	32	11	21	32
26	14	16	30	8	22	30

4.3.3.4 **Spatially Dependent Benthic Counts - Models**

Since anisotropy does not appear to be too great a problem, at least for *Diporeia hoyi*, all site pairs were used to estimate the semivariogram (Figure 4.19). The resulting model parameters are shown in Table 4.10, with the estimates for the independent case from Table 4.7 repeated for comparison. The main effect of including spatial dependence in the model is to reverse the sign for $\hat{\beta}_4$, the coefficient for log alkalinity which is highly correlated (negatively) with $\hat{\beta}_0$ which also reversed sign. The standard error of the parameter estimates also increase slightly and the need for an intercept is questionable.

Two additional improvements have been considered for the *Diporeia hoyi* model. Although the fit of the exponential model in Figure 4.19 is reasonable, there is an obvious influence of outliers. The robust semivariogram estimator, $\overline{\gamma}$, discussed in Section 4.2, often gives better estimates in the presence of outliers. Another possible improvement is the use of water distance as defined by Rathbun (1998). If the sites in Figures 4.7 and 4.8 are plotted in a GIS, such as SPANS, a utility can be used that will report the distance between any two points along any path. If the path selected is the shortest distance by water (ie. around islands and peninsulas), then the distance reported will be Rathbun's water distance. Using this distance will have the effect of increasing

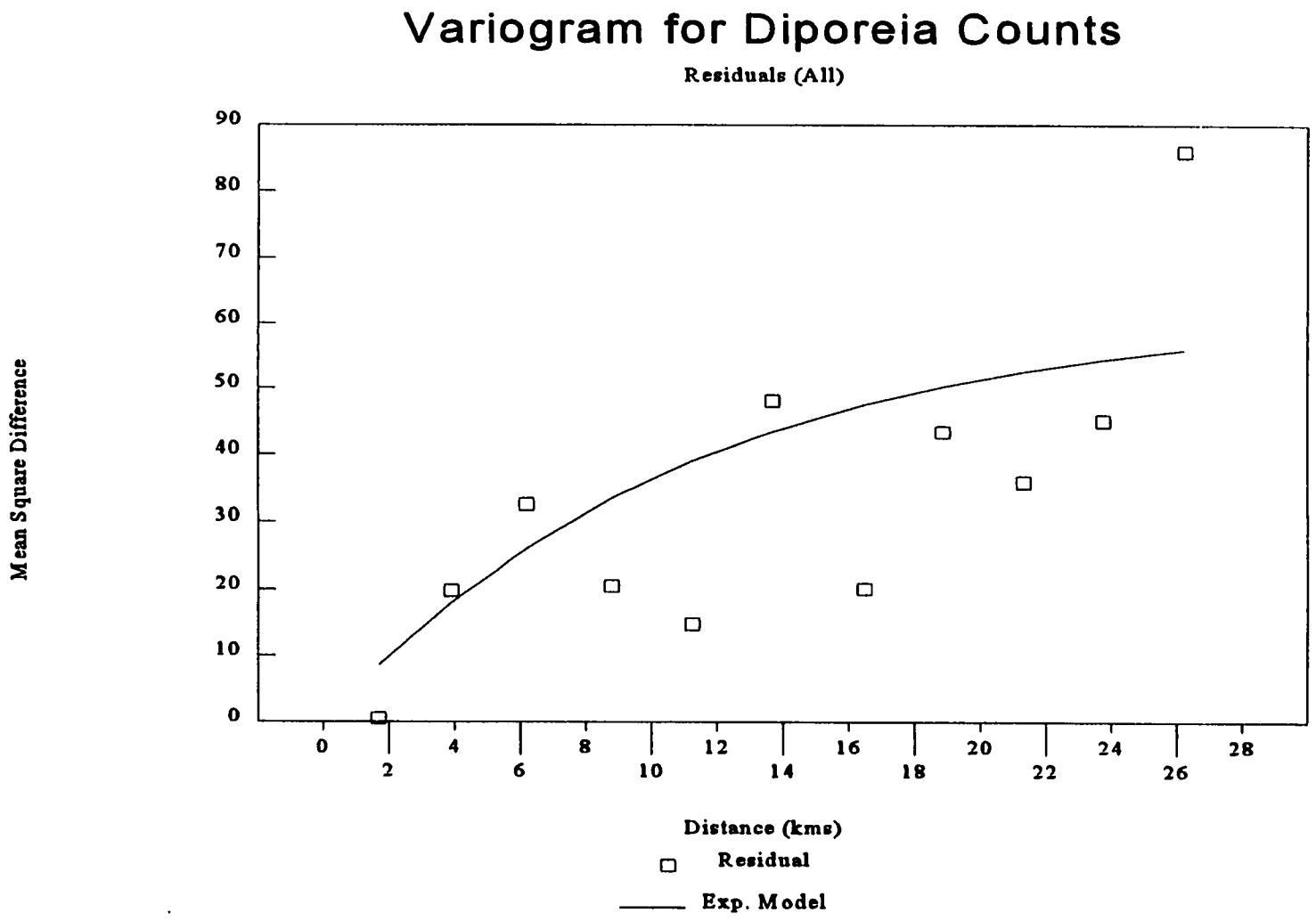

FIGURE 4.19 *Diporeia* Residuals Variogram (All Directions)

TABLE 4.10

Parameter Estimates for *Diporeia hoyi* Using Quasi-Likelihood for Cases With and Without Spatial Dependence

	Independent		Dependent	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	10.39	3.91	-6.14	4.74
log(Depth)	0.866	0.19	0.805	0.19
$\log(Al_2O_3)$	-2.50	0.67	-1.62	0.78
$\log(K_2O)$	3.01	0.74	1.89	0.75
log(alkalinity)	-1.98	0.89	1.76	1.08
σ^2	0.688		0.623	
$\hat{\theta}_3$	-		0.0885	

the interaction distance between many of the sites. Table 4.11 gives the number of pairs for each water distance group. The resulting semivariograms from these two improvements are shown in Figures 4.20 and 4.21. Figure 4.22 is the semivariogram that results when both improvements are combined. The parameter estimates are presented in Tables 4.12 (previous results repeated for comparison) and 4.13.

The use of the robust variogram estimator, $\bar{\gamma}$, seems to have the expected effect of lowering the variance estimates for all parameters of the quasi-likelihood model. The use of the water distance has the effect of lowering both parameter estimates and standard errors. Whether this will lead to better predictions is best decided by cross-validation (Chapter 5). Tables 4.14 and 4.15 show the correlation matrices for the parameter estimates. The different choices of distance and variogram model have little effect on this structure.

The effect of these changes in the model for a second species, *Procladius*, are examined here. Figures 4.23 - 4.26 are semivariograms for quasi-likelihood residuals of *Procladius* counts with and without the robust semivariogram estimator and with and without water distances. Common to all four figures is the fast approach to the sill of the semivariogram with a range of about 4 km. This is in contrast to *Diporeia hoyi*, which had a range of about

TABLE 4.11

Number of Site Pairs for Semivariogram Estimation with Water Distance

Lag Number	Average Water Distance (km)	Number of Site Pairs
1	1.33	3
2	3.78	25
3	6.01	39
4	8.73	45
5	11.17	42
6	13.83	45
7	16.04	31
8	18.78	26
9	20.99	32
10	23.44	19
11	26.11	15

Minimum Water Distance = 0.8 km

Maximum Water Distance = 65.0 km

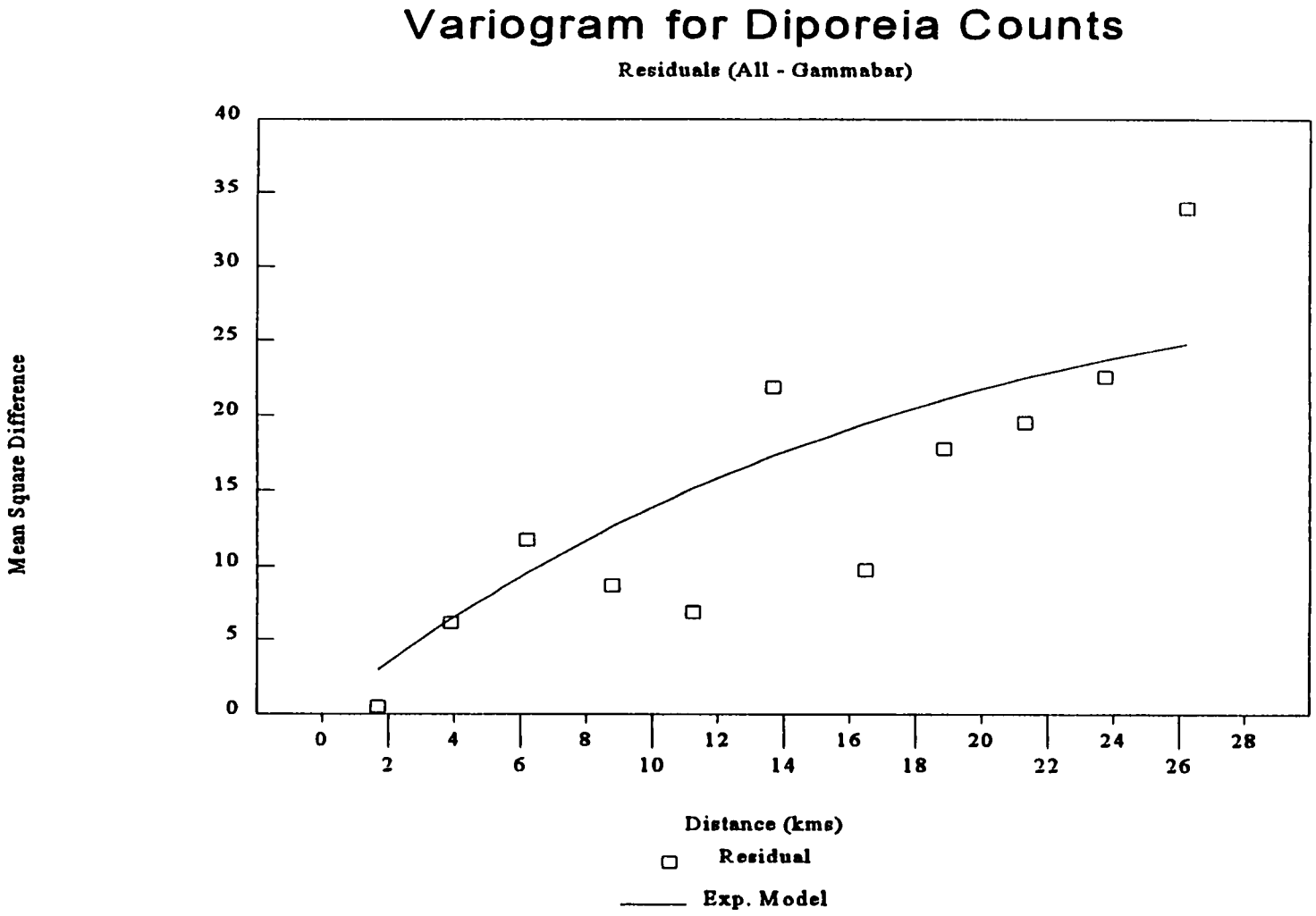

FIGURE 4.20 *Diporeia* Residuals Variogram (Robust Estimator)

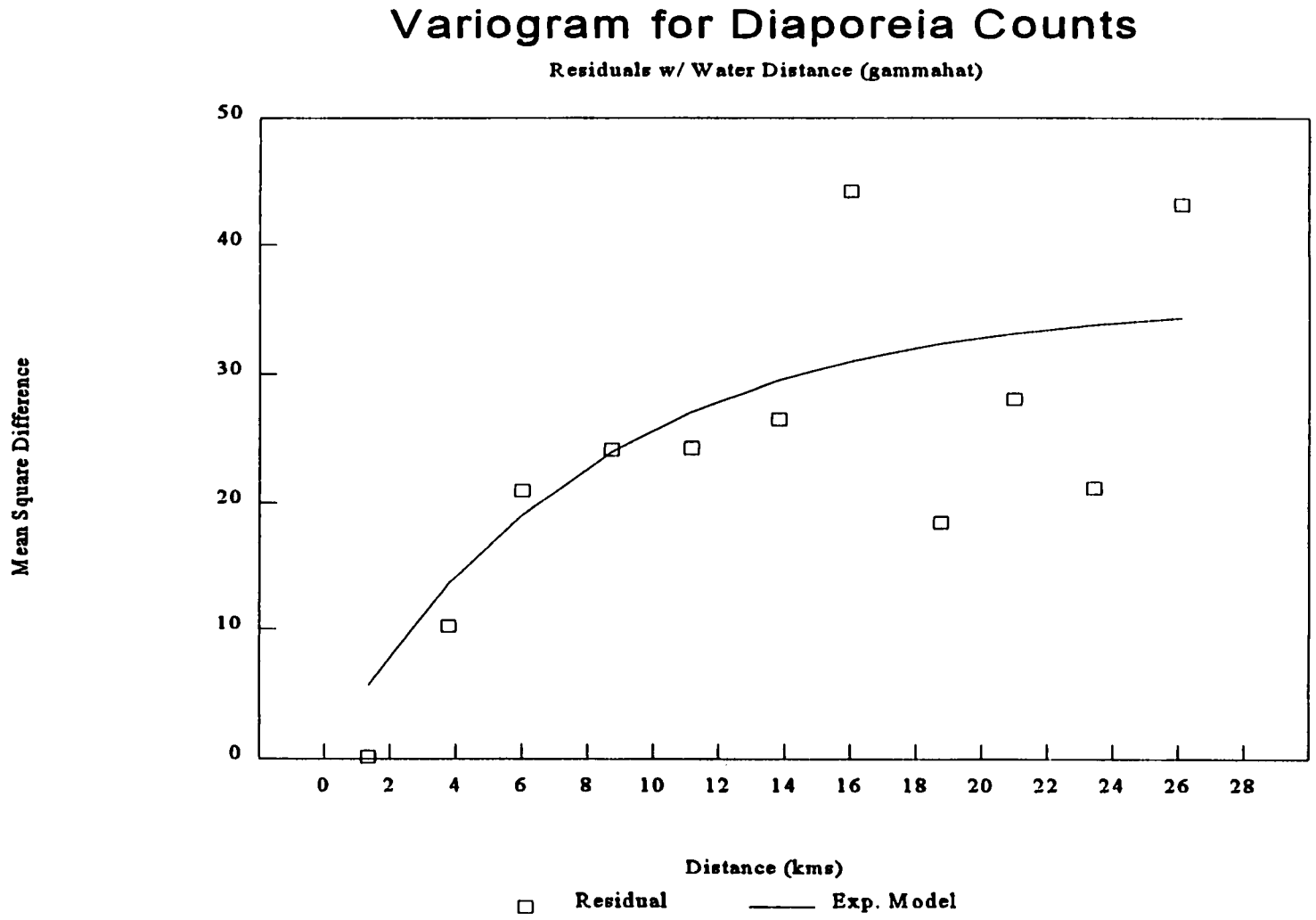

FIGURE 4.21 *Diporeia* Residuals Variogram (Classical Estimator) with Water Distance

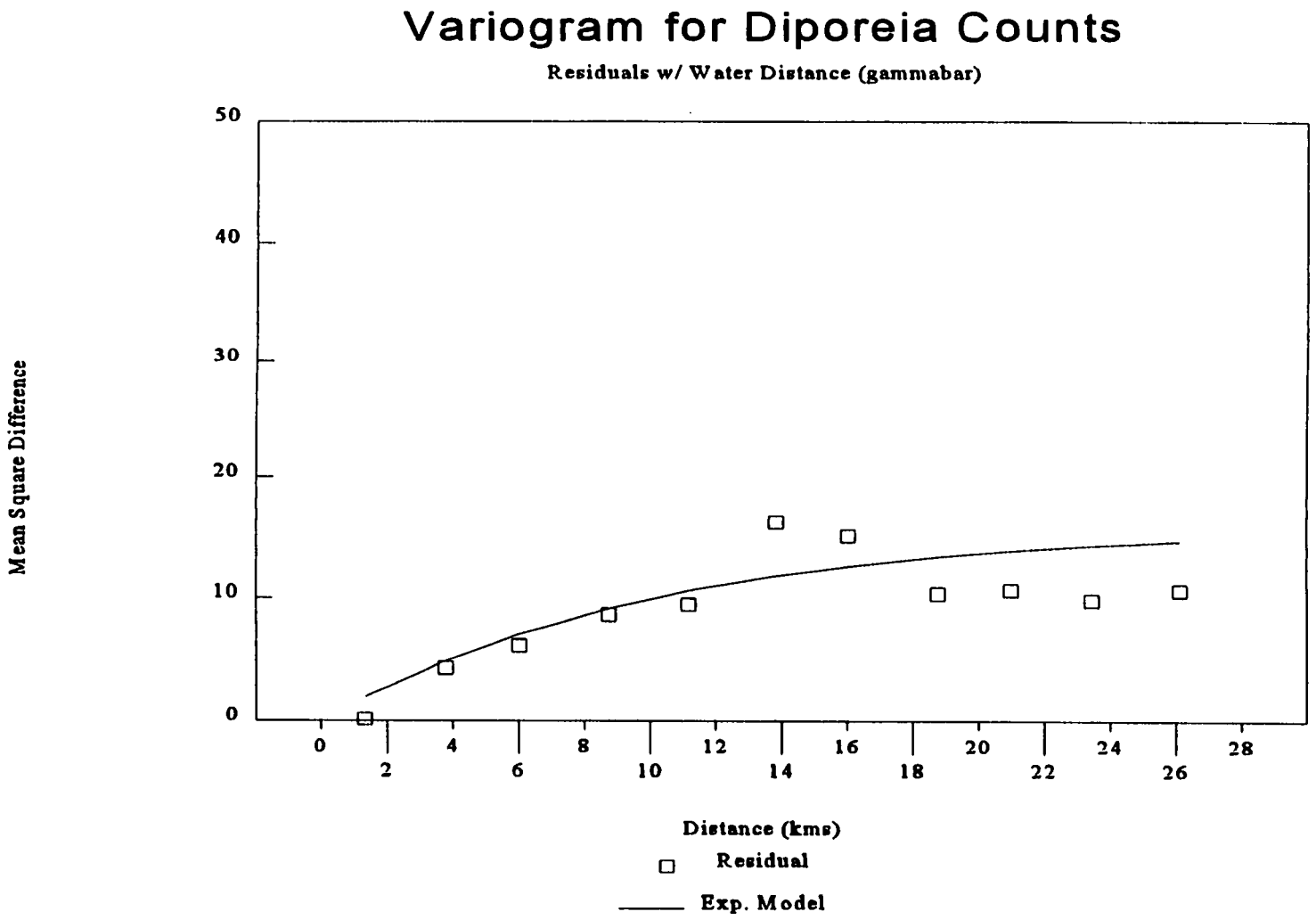

FIGURE 4.22 *Diporeia* Residuals Variogram (Robust Estimator) with Water Distance

TABLE 4.12

Parameter Estimates for *Diporeia hoyi* Using Quasi-Likelihood for Cases With and Without $\bar{\gamma}$ and Distance from Latitude and Longitude

	$\bar{\gamma}$		$\hat{\gamma}$	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-6.70	4.01	-6.14	4.74
log(Depth)	0.747	0.15	0.805	0.19
$\log(Al_2O_3)$	-1.51	0.61	-1.62	0.78
$\log(K_2O)$	1.78	0.74	1.89	0.75
log(alkalinity)	1.91	0.92	1.76	1.08
$\hat{\sigma}^2$	0.311		0.623	
$\hat{\theta}_3$	0.0577		0.0885	

TABLE 4.13

Parameter Estimates for *Diporeia hoyi* Using Quasi-Likelihood for Cases With and Without $\bar{\gamma}$ and Water Distance

	$\bar{\gamma}$		$\hat{\gamma}$	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
Intercept	-2.60	3.21	-2.45	4.02
log(Depth)	0.681	0.13	0.758	0.17
$\log(Al_2O_3)$	-1.23	0.51	-1.41	0.67
$\log(K_2O)$	1.55	0.52	1.75	0.68
log(alkalinity)	0.86	0.73	0.83	0.92
σ^2	0.185		0.456	
$\hat{\theta}_3$	0.0979		0.1271	

TABLE 4.14

Correlation Matrices for Parameter Estimates for *Diporeia hoyi* Using Quasi-Likelihood for Cases With and Without $\bar{\gamma}$ and Distance from Latitude and Longitude

a) With $\bar{\gamma}$

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_0$	1				
$\hat{\beta}_1$	0.0309	1			
$\hat{\beta}_2$	-0.3141	-0.1393	1		
$\hat{\beta}_3$	0.2981	-0.2058	-0.7234	1	
$\hat{\beta}_4$	-0.9612	-0.0617	0.0677	-0.1617	1

b) With $\hat{\gamma}$

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_0$	1				
$\hat{\beta}_1$	0.0363	1			
$\hat{\beta}_2$	-0.3324	-0.1067	1		
$\hat{\beta}_3$	0.3025	-0.2087	-0.7252	1	
$\hat{\beta}_4$	-0.9521	-0.0817	0.0574	-0.1496	1

TABLE 4.15

Correlation Matrices for Parameter Estimates for *Diporeia hoyi* Using Quasi-Likelihood for Cases With and Without $\bar{\gamma}$ and Water Distance

a) With $\bar{\gamma}$

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_0$	1				
$\hat{\beta}_1$	0.1119	1			
$\hat{\beta}_2$	-0.3678	-0.0472	1		
$\hat{\beta}_3$	0.2859	-0.3151	-0.6658	1	
$\hat{\beta}_4$	-0.9532	-0.1723	0.1003	-0.1619	1

b) With $\hat{\gamma}$

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_0$	1				
$\hat{\beta}_1$	0.1009	1			
$\hat{\beta}_2$	-0.3622	-0.0358	1		
$\hat{\beta}_3$	0.2848	-0.2976	-0.6853	1	
$\hat{\beta}_4$	-0.9481	-0.1694	0.0786	-0.1447	1

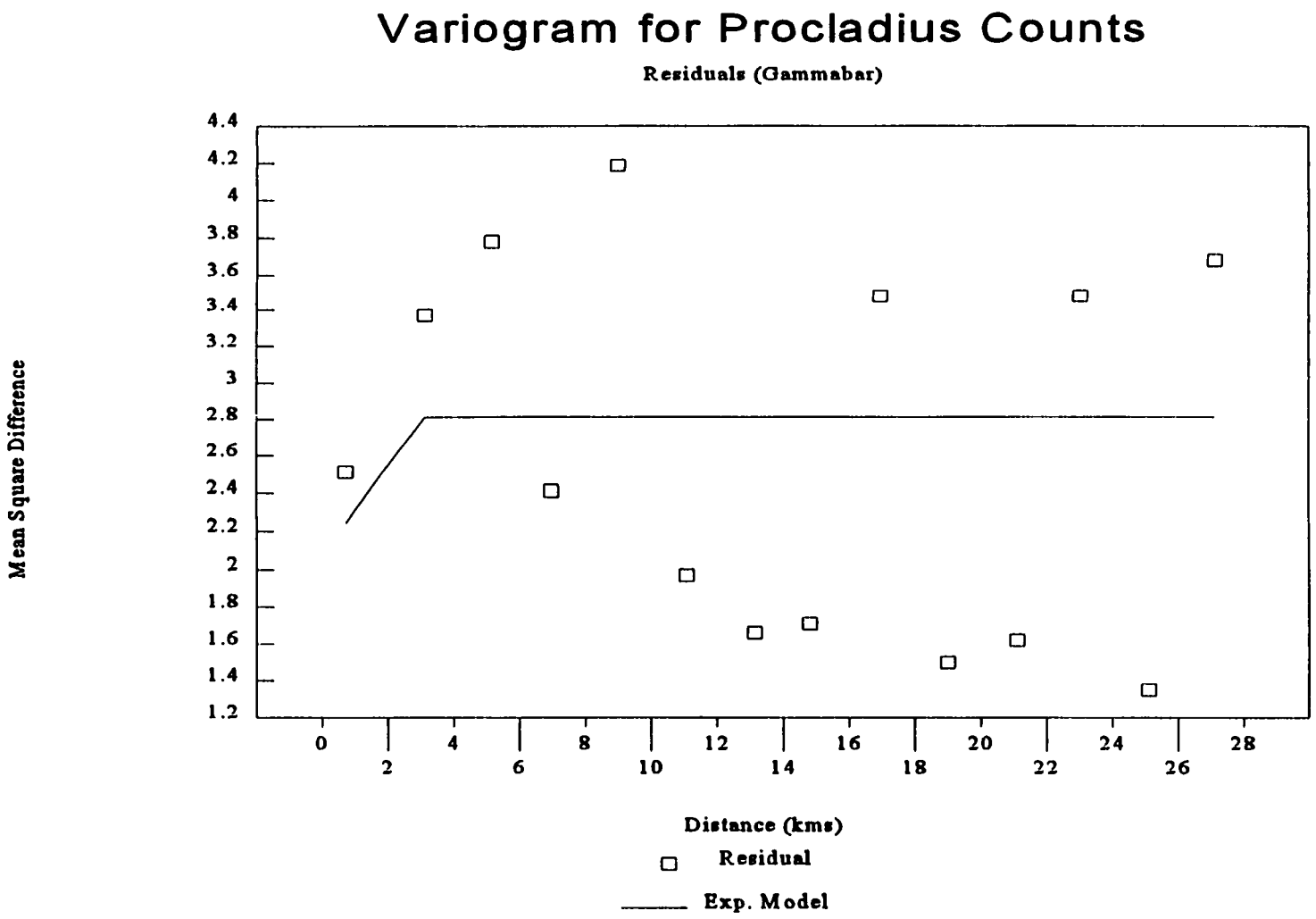

FIGURE 4.23 *Procladius* Residuals Variogram (Robust Estimator)

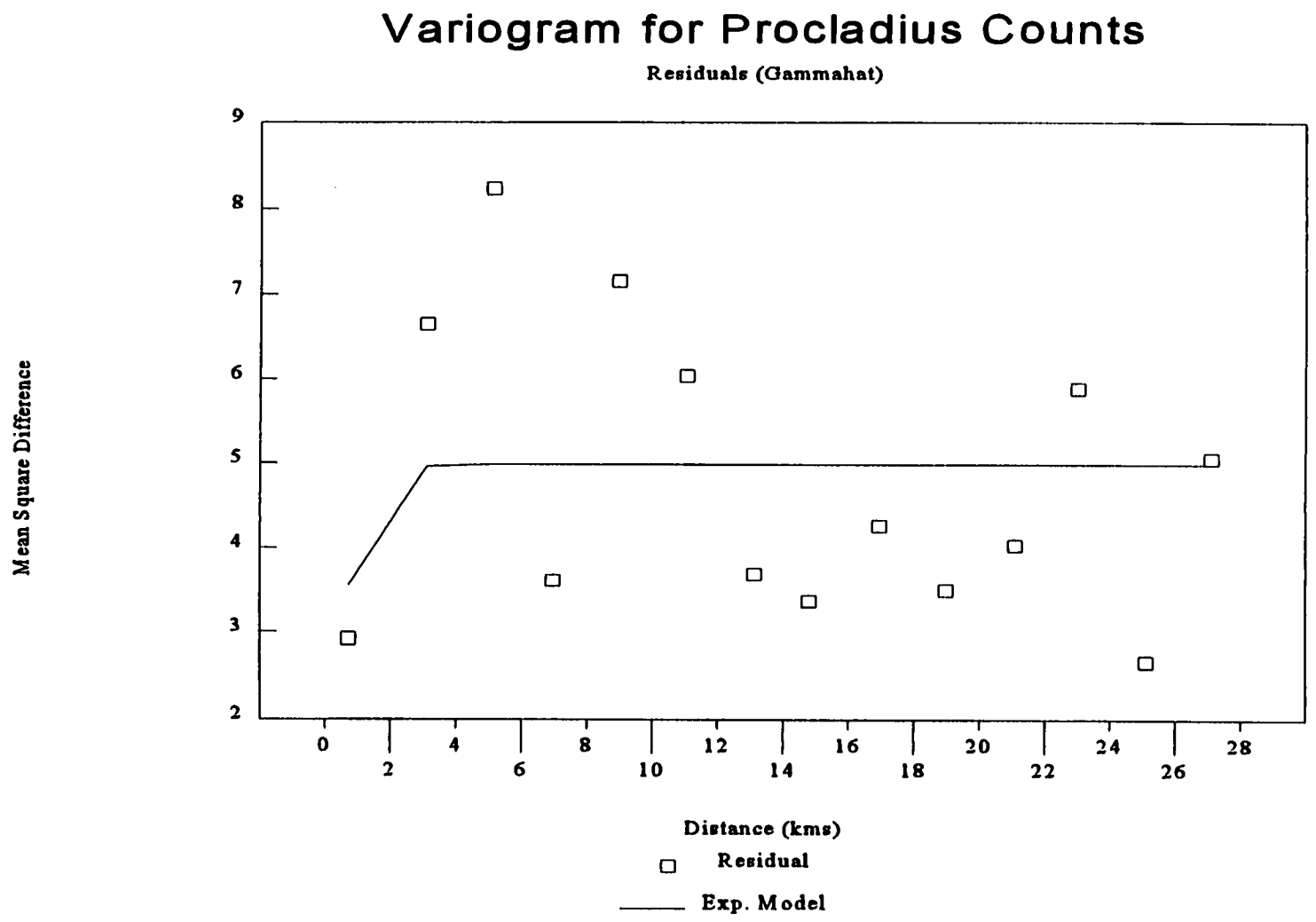

FIGURE 4.24 *Procladius* Residuals Variogram (Classical Estimator)

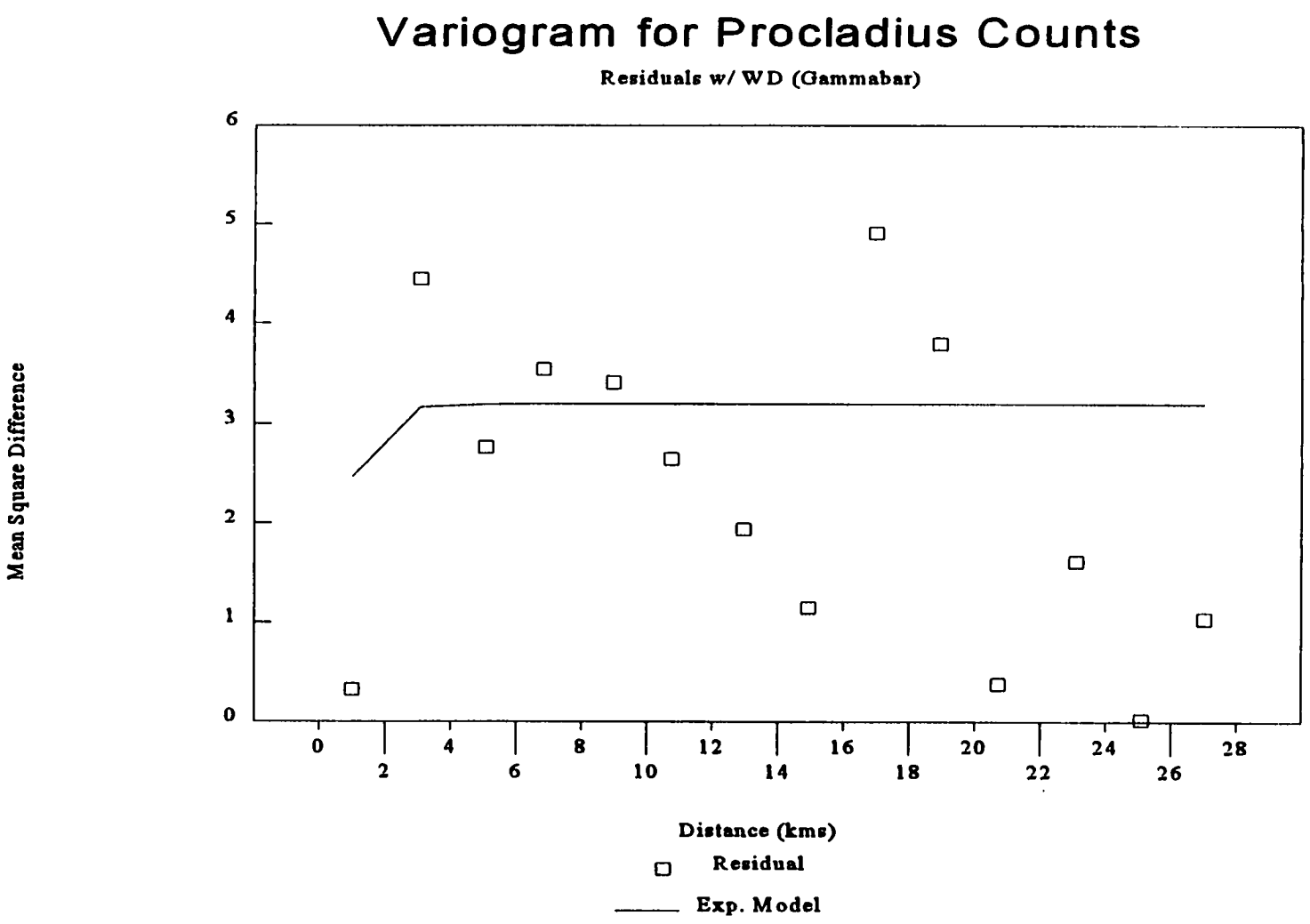

FIGURE 4.25 *Procladius* Residuals Variogram (Water Distance, Robust Estimator)

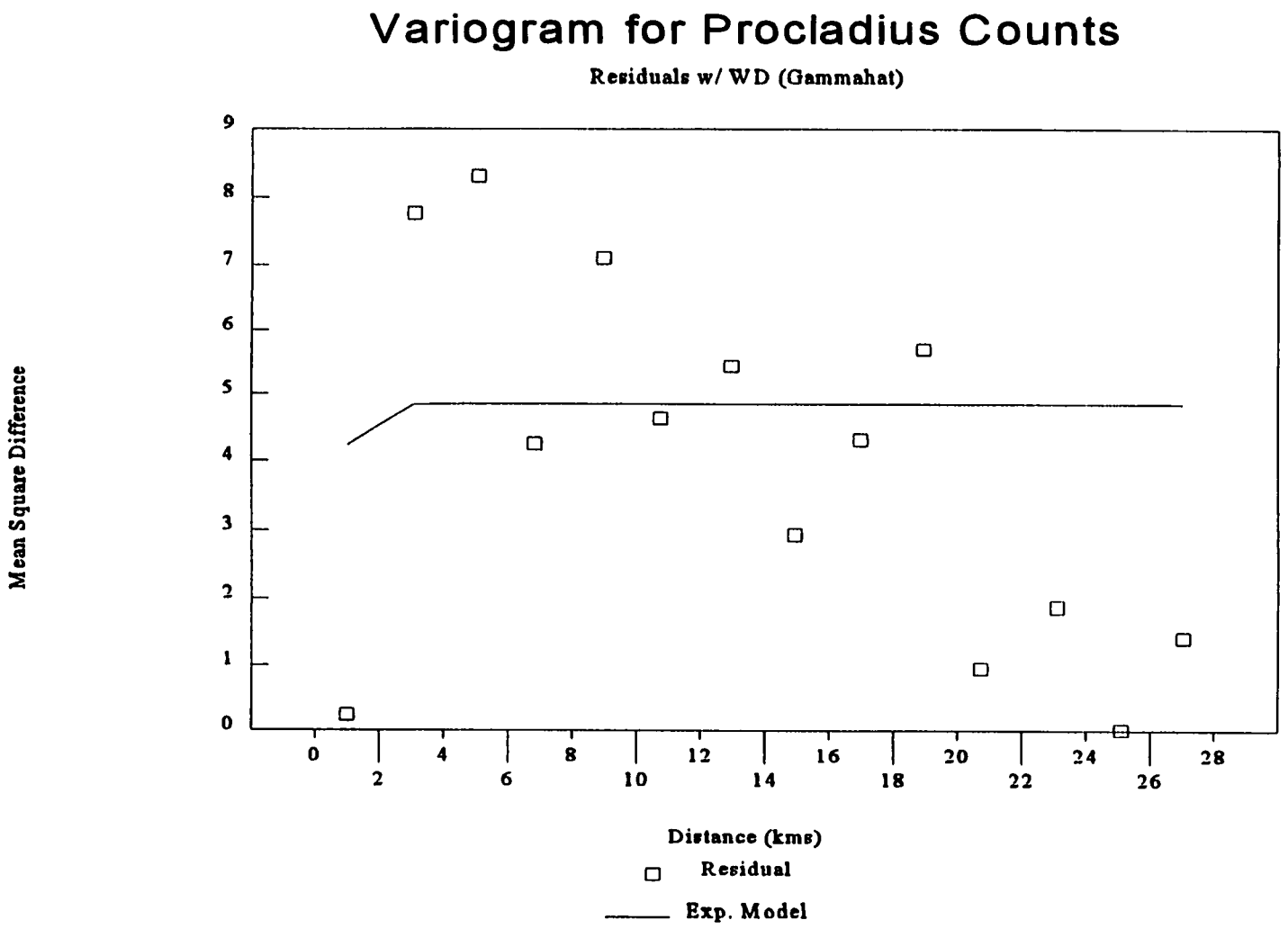

FIGURE 4.26 *Procladius* Residuals Variogram (Water Distance, Classical Estimator)

30 km. Examining the semivariograms, it may be that a more appropriate variogram model could be found. However, it is more likely that spatial autocorrelation for *Procladius*, if present, is occurring at a scale smaller than observable for this data set. Tables 4.16 and 4.17 present the parameter estimates. No intercept (β_0) was used in the log linear model because that parameter was found to be not statistically different from zero. There are very little differences among the estimates. There is a large negative correlation between the parameter estimates in all cases.

TABLE 4.16

Parameter Estimates for *Procladius* Using Quasi-Likelihood for Cases With and Without $\bar{\gamma}$ and Distance from Latitude and Longitude

	$\bar{\gamma}$		$\hat{\gamma}$	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
log(Depth)	-1.275	0.18	-1.402	0.21
$\log(Al_2O_3)$	1.534	0.18	1.662	0.22
$\hat{\sigma}^2$	0.201		0.514	
$\hat{\theta}_3$	2.1863		1.6946	
$Corr(\hat{\beta}_1, \hat{\beta}_2)$	-0.9705		-0.9719	

TABLE 4.17

Parameter Estimates for *Procladius* Using Quasi-Likelihood for Cases With and Without $\bar{y}$ and Water Distance

	$\bar{y}$		$\hat{y}$	
	$\hat{\beta}$	Std err.	$\hat{\beta}$	Std err.
log(Depth)	-1.314	0.19	-1.408	0.22
$\log(Al_2O_3)$	1.572	0.19	1.668	0.22
$\hat{\sigma}^2$	0.283		0.542	
$\hat{\theta}_3$	1.4228		1.6636	
$Corr(\hat{\beta}_1, \hat{\beta}_2)$	-0.9710		-0.9721	

CHAPTER 5: SPATIAL PREDICTION

5.1 Introduction

This topic was reviewed briefly in Chapter 2 (2.2.6). The main tool for prediction in spatial statistics is kriging, as introduced in Section 2.2.1. The different types of kriging have been developed to deal with various problems that occur with spatial data such as non-stationarity, non-Gaussian data and outliers. The variogram, as described in Chapter 4, is the key to deriving the kriging equations. It contains the information on the strength and structure of spatial autocorrelation at different distances or lags. The information carried by autocorrelation at a new site, $\mathbf{s}_0$, at known distances, d_i, from the sites $\mathbf{s}_i$, is needed to predict a new value. The analogy with time series continues here. Suppose a prediction, or forecast, was desired at time t=n+l, where n is the number of observations and l is a positive integer representing the number of lags into the future the forecast will project. If l is large, then there will be little autocorrelation effect and the model for the mean, μ, will be the forecast, under the assumption of stationarity. If l is small, then some adjustment to the forecast will have to be made for autocorrelation. This is what the kriging equations and the variogram accomplish for spatial predictions.

In this Chapter, kriging is used to make predictions. The quasi-likelihood model with dependent errors is used to handle the non-stationarity and the variogram is used to model the autocorrelation structure of the residuals. First, simple kriging and ordinary kriging will be developed in some detail following Cressie (1993, Ch. 3). Then, median polish kriging will be described so that the analogy can be made with the approach taken here; i.e. kriging of residuals to allow predictions at new sites for which explanatory variables are available. The approach will be applied to the examples from the previous chapter and the prediction error will be estimated. The models developed so far will also be cross-validated.

5.2 Kriging

5.2.1 Simple Kriging

As introduced in Chapter 2, there are many types of kriging. They are all based on finding an optimal predictor that minimizes the squared-error loss:

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

where $L(\cdot)$ is the loss function

$p(\mathbf{Z};\mathbf{s}_0)$ is the prediction for the process $\mathbf{Z}(\cdot)$ at site $\mathbf{s}_0$

and $\mathbf{Z}(\mathbf{s}_0)$ is the true value of the process at site $\mathbf{s}_0$.

The predictor, $p^{o}(\mathbf{Z};\mathbf{s}_0)$ which minimizes the expectation:

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

is $E(\mathbf{Z}(\mathbf{s}_0)|\mathbf{Z})$, the conditional expectation.

Cressie (1993, p. 109) notes that this conditional expectation is not always linear. However, if the best linear predictor that minimizes the expectation is sought, the results are the kriging equations. If the predictor is:

$$p(\mathbf{Z};\mathbf{s}_0) = \sum_{i=1}^{n} l_i \mathbf{Z}(\mathbf{s}_i) + k,$$

$l_1,...,l_n,k$ are sought such that $E(\mathbf{Z}(\mathbf{s}_0) - p(\mathbf{Z};\mathbf{s}_0))^2$ is minimized. This is equivalent to minimizing over $l_1,...,l_n,k$:

$$E[(\mathbf{Z}(\mathbf{s}_0) - \sum_{i=1}^{n} l_i \mathbf{Z}(\mathbf{s}_i) - k)^2] = var(\mathbf{Z}(\mathbf{s}_0) - \sum_{i=1}^{n} l_i \mathbf{Z}(\mathbf{s}_i)) + (\mu(\mathbf{s}_0) - \sum_{i=1}^{n} l_i \mu(\mathbf{s}_i) - k)^2$$

where $\mu(\mathbf{s}) = E(\mathbf{Z}(\mathbf{s}))$.

The optimal solution is found by differentiating this expectation with respect to $l_1,...,l_n,k$ and equating the result to zero:

$$k = \mu(\mathbf{s}_0) - \sum_{i=1}^{n} l_i \mu(\mathbf{s}_i),$$

and $\mathbf{l} = \mathbf{c}'\Sigma^{-1}$

where $\mathbf{c} = (C(\mathbf{s}_0,\mathbf{s}_1),...,C(\mathbf{s}_0,\mathbf{s}_n))'$

and Σ is an n x n matrix whose (i,j) element is $C(\mathbf{s}_i,\mathbf{s}_j)$.

Therefore, the optimal linear predictor is:

$$p^*(\mathbf{Z};\mathbf{s}_0) = \mathbf{c}'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu}) + \mu(\mathbf{s}_0)$$

where $\boldsymbol{\mu} = (\mu(\mathbf{s}_1),\ldots,\mu(\mathbf{s}_n))'$.

The minimized mean-squared prediction error is:

$$\sigma^2(\mathbf{s}_0) = C(\mathbf{s}_0,\mathbf{s}_0) - \mathbf{c}'\Sigma^{-1}\mathbf{c}.$$

This type of spatial prediction is known as simple kriging and it assumes that the mean function, $\mu(\cdot)$ is known. Since the semivariogram is related to the covariance function by:

$$\gamma(\mathbf{s}) = C(0) - C(\mathbf{s}),$$

knowledge of the variogram and the mean function allows prediction via simple kriging.

5.2.2 Ordinary Kriging

In practice, the mean function is rarely known. Like the variogram, the mean function is usually estimated from the data. Ordinary kriging assumes that the mean function, μ, is unknown, but, as in simple kriging, the predictor is still assumed to be a linear combination of observed values. An additional constraint, that the coefficients of the linear predictor sum to 1 ($\sum_{i=1}^{n} \lambda_i = 1$), is introduced to guarantee unbiasedness:

$$E(p(\mathbf{Z};\mathbf{s}_0)) = \mu = E(\mathbf{Z}(\mathbf{s}_0)) \text{ for all } \mu.$$

The optimal predictor that minimizes the squared-error loss over $\sum_{i=1}^{n} \lambda_i \mathbf{Z}(\mathbf{s}_i)$ subject to $\sum_{i=1}^{n} \lambda_i = 1$, is the ordinary kriging predictor. Equivalently, the quantity

$$E(\mathbf{Z}(\mathbf{s}_0) - \sum_{i=1}^{n} \lambda_i \mathbf{Z}(\mathbf{s}_i))^2 - 2m(\sum_{i=1}^{n} \lambda_i - 1) \tag{5.1}$$

is minimized with respect to $\lambda_1,...,\lambda_n$, and *m* (a Lagrange multiplier that ensures the constraint on the λs will be met). If the variogram of the process is:

$$2\gamma(\mathbf{h}) = \text{var}(\mathbf{Z}(\mathbf{s} + \mathbf{h}) - \mathbf{Z}(\mathbf{s})),$$

then the quantity to minimize becomes:

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j\gamma(\mathbf{s}_i - \mathbf{s}_j) + 2\sum_{i=1}^{n} \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - 2m(\sum_{i=1}^{n} \lambda_i - 1). \tag{5.2}$$

To see this, note that since $\sum_{i=1}^{n} \lambda_i = 1$, the following relations hold:

$$\mathbf{Z}(\mathbf{s}_0)^2 = \sum_{i=1}^{n} \lambda_i \mathbf{Z}(\mathbf{s}_0)^2 \tag{5.3}$$

$$\mathbf{Z}(\mathbf{s}_0) \sum_{i=1}^{n} \lambda_i \mathbf{Z}(\mathbf{s}_i) = \sum_{i=1}^{n} \lambda_i \mathbf{Z}(\mathbf{s}_0)\mathbf{Z}(\mathbf{s}_i) \tag{5.4}$$

$$\sum_{i=1}^{n} \lambda_i \mathbf{Z}(\mathbf{s}_i)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \mathbf{Z}(\mathbf{s}_i)^2 \tag{5.5}$$

$$\sum_{j=1}^{n} \lambda_j \mathbf{Z}(\mathbf{s}_j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \mathbf{Z}(\mathbf{s}_j)^2 \tag{5.6}$$

Note that equations 5.5 and 5.6 are equivalent, since only the subscript is

different. Also,

$$\{\sum_{i=1}^{n} \lambda_i \, Z(s_i)\}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \, Z(s_i) \, Z(s_j) \qquad (5.7)$$

and $$(Z(s_0) - \sum_{i=1}^{n} \lambda_i \, Z(s_i))^2 = Z(s_0)^2 - 2Z(s_0) \sum_{i=1}^{n} \lambda_i \, Z(s_i) + \{\sum_{i=1}^{n} \lambda_i \, Z(s_i)\}^2 \qquad (5.8)$$

which is the expansion of the expectation in eqation 5.1. Substituting equations 5.3-5.7 into equation 5.8,

$$(Z(s_0) - \sum_{i=1}^{n} \lambda_i \, Z(s_i))^2 = \sum_{i=1}^{n} \lambda_i \, Z(s_0)^2 - 2\sum_{i=1}^{n} \lambda_i \, Z(s_0) \, Z(s_i)$$

$$+ \sum_{i=1}^{n} \lambda_i \, Z(s_i)^2 - 1/2\{\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \, Z(s_i)^2$$

$$- 2\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \, Z(s_i) \, Z(s_j)$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \, Z(s_j)^2\}$$

$$= 2\sum_{i=1}^{n} \lambda_i (Z(s_0)^2 - 2Z(s_0) \, Z(s_i) + Z(s_i)^2)/2 -$$

$$1/2\{\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (Z(s_i)^2 - 2Z(s_i) \, Z(s_j) + Z(s_j)^2)\}$$

$$= - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (Z(s_i) - Z(s_j))^2/2 + 2\sum_{i=1}^{n} \lambda_i (Z(s_0) - Z(s_i))^2/2$$

Taking the expectation as indicated in equation 5.1 yields equation 5.2. If equation 5.2 is now differentiated with respect to λ and *m* and equated to zero, the optimal λ for ordinary kriging, λ_o, can be obtained from:

$$\lambda_o = \Gamma_o^{-1}\gamma_o$$

where $\lambda_o = (\lambda_1,...,\lambda_n,m)'$,

$\gamma_o = (\gamma(\mathbf{s}_0 - \mathbf{s}_1),...,\gamma(\mathbf{s}_0 - \mathbf{s}_n),1)'$

and $\Gamma_o = \gamma(\mathbf{s}_i - \mathbf{s}_j)$, for i=1,..,n, j=1,...,n

$= 1$, for i=n+1, j=1,...,n

$= 0$, for i=n+1, j=n+1

The minimized mean-squared prediction error, or kriging variance, is:

$$\sigma_k^2(\mathbf{s}_0) = \lambda_o'\gamma = \sum_{i=1}^{n} \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) + m. \qquad (5.9)$$

5.2.3 Median-Polish Kriging

The two types of kriging described so far both rely on the following decomposition of the process $Z(\cdot)$:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}), \qquad \mathbf{s} \in D,$$

where $\mu(\cdot) = E(Z(\cdot))$ is the mean structure

and $\delta(\cdot)$ is the error structure.

If $\mu(\cdot)$ is known, then the optimal linear predictor of $Z(\mathbf{s}_0)$ is the simple kriging estimator defined in Section 5.2.1. If $\mu(\cdot)$ is unknown, it may still be possible to

estimate $\mu(\cdot)$ in terms of its directional components. For example, in two dimensions, assume:

$$\mu(\mathbf{s}) = a + c(x) + r(y), \qquad \mathbf{s} = (x,y)' \in D.$$

If the $\mathbf{s}_i$ are on a grid, c(x) can be thought of as a column effect, r(y) as a row effect and a as an overall effect.

In median-polish kriging, the median-polish estimates $\tilde{\mu}(\mathbf{s}_i)$ are found from the data and the residuals, $Z(\mathbf{s}_i) - \tilde{\mu}(\mathbf{s}_i) = R(\mathbf{s}_i)$, are kriged using ordinary kriging (Section 5.2.2). The median-polish kriging predictor of $Z(\mathbf{s}_0)$ is:

$$\tilde{Z}(\mathbf{s}_0) = \tilde{\mu}(\mathbf{s}_0) + \hat{R}(\mathbf{s}_0) = \tilde{\mu}(\mathbf{s}_0) + \sum_{i=1}^{n} \lambda_i R(\mathbf{s}_i).$$

The kriging variance associated with $\tilde{Z}(\mathbf{s}_0)$ is defined as the ordinary-kriging variance of the median-polish residuals. This quantity, which Cressie (1993, p. 191) calls σ_m^2, is the same as equation 5.9 of the previous section when the residuals are kriged instead of the raw data. This is justified as follows:

$$\begin{aligned}(Z(\mathbf{s}_0) - \tilde{Z}(\mathbf{s}_0)) &= (Z(\mathbf{s}_0) - \tilde{\mu}(\mathbf{s}_0) - \hat{R}(\mathbf{s}_0)) \\ &= (R(\mathbf{s}_0) - \hat{R}(\mathbf{s}_0))\end{aligned}$$

and therefore:

$$E[(Z(\mathbf{s}_0) - \tilde{Z}(\mathbf{s}_0))^2] = E[(R(\mathbf{s}_0) - \hat{R}(\mathbf{s}_0))^2].$$

Conditional on $\tilde{\mu}$, the kriging variance is:

$$E\{(R(\mathbf{s}_0) - \hat{R}(\mathbf{s}_0))^2 | \tilde{\mu}(\cdot)\} = \sigma_m^2.$$

5.3 Kriging with Quasi-likelihood Residuals

5.3.1 Introduction

Cressie (1993, p. 190) points out that the median-polish residuals of the previous section can be thought of as a new spatial data set that has been detrended to allow ordinary kriging. Further, the analogy is made to time-series analysis where stationary models are fit to data after detrending. The quasi-likelihood model of Chapter 3 is another parameterization of the mean function, $\mu(\cdot)$, and the variograms fit in Chapter 4 were all based on residuals analogous to median-polish residual. It seems natural to attempt to make predictions based on the quasi-likelihood model for the mean and residuals that have been subjected to ordinary kriging.

Gotway and Hartford (1996) describe a similar approach known as "kriging with external drift." In this case, the mean function, $\mu(\cdot)$ is a linear combination of the explanatory variables and is incorporated into the universal kriging process in place of the usual functions of location. If the quasi-likelihood model of the present study were replaced by an ordinary least-squares model using the explanatory variables, the result would be "kriging with external drift."

5.3.2 Cross-Validation to Assess Fit

Cross-validation is used to assess the fit of the two quasi-likelihood models, with and without correlated errors. The procedure used was described in Section 3.3.3. The results presented in that section are repeated here for comparison. When an observation is deleted for the case of the spatially dependent model, one row and one column must be deleted from the **W** or **D** matrix, since there will be one less interaction. The corresponding row and column must also be deleted from Γ_o and the deleted values form Y_0.

5.3.3 Comparison of Quasi-Likelihood Models

5.3.3.1 **California Plant Species Example**

When the above cross-validation procedure was carried out for the quasi-likelihood models of the California plant species example, the results in Table 5.1 were found. The quantity R_s, the standardized prediction residual defined above, should be close to zero. $RMSR_s$, the root mean square prediction residual, should be close to one. PRESS, the prediction sum of squares, should be as small as possible. Based on this procedure, the quasi-likelihood model

TABLE 5.1

Cross-Validation Results for Quasi-likelihood Models of California Plant Species Data (With Table 3.10 Included for Comparison)

Model	R_s	$RMSR_s$	PRESS
OLS	-.020	1.095	187.8
QL (Poisson)	.048	5.428	157.2
QL (neg. bin.)	.107	1.572	178.2
Exponential	-.058	1.100	181.2
Wave	-.032	1.152	166.1
Relative Exp.	-.833	2.924	176.2
Relative Wave	-.531	1.580	164.3

with dependent errors with the semivariogram modeled by the wave function is the preferred model.

5.3.3.2 **Benthic Counts Example**

The cross-validation described above was also carried out for the models for benthic counts described in Chapter 4. In this case, the exponential model was used for the semivariogram for all parameterizations involving spatial dependence, but two different procedures for estimating intersite distance as well as the classical and robust estimator of the semivariogram were compared. Table 5.2 gives the results.

All of the spatial dependence models have better cross-validation statistics than the independent case, except for the robust semivariogram which has a slightly larger root mean-square standardized prediction residual ($RMSR_s$). Clearly, the lowest prediction sum of squares occurs for the case of water distance. Since the classical variogram model had lower R_s and $RMSR_s$, it was used in the following section to illustrate predictions and in Chapter 6 for simulations.

TABLE 5.2

Cross-Validation Results for Quasi-likelihood Models of Benthic Count Data (*Diporeia hoyi*)

Model	R_s	$RMSR_s$	PRESS
Independent	.112	1.932	9.290
Latitude/Longitude Distance:			
Classical Est.	-.041	1.231	7.024
Robust Est.	-.064	2.045	7.071
Water Distance:			
Classical Est.	-.041	1.269	6.435
Robust Est.	-.066	2.005	6.275

5.3.4 Predictions

5.3.4.1 **Introduction**

As discussed in Chapter 1, the need to establish biologically-based clean-up targets is the motivation for a predictive model of benthic counts. In this section, the procedure for predictions at new sites will be described. As for the case of cross-validation, ordinary kriging of quasi-likelihood residuals will be used, but block kriging will also be discussed as an alternative. Predictions will be made for *Diporeia hoyi* abundance at two sites in Georgian Bay and the North Channel of Lake Huron and prediction intervals will be constructed. Finally, the potential for adapting the predictive method to the multivariate case will be discussed.

5.3.4.2 **Prediction at New Sites**

Predictions at a new site require information about the new site and its spatial relationship to existing sites. In particular, the explanatory variables at the new site must be available to obtain the estimate of the mean. Then, as was done with cross-validation, the vector γ_o is constructed using the variogram and the location information about the new site. In this case, the entire Γ_o matrix is

used. The predicted value for a new site, $\mathbf{s}_0$, is:

$$\hat{\boldsymbol{Z}}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) + \sum_{i=1}^{n} \lambda_i \, R(\mathbf{s}_i)$$

where $\hat{\mu}(\mathbf{s}_0) = \exp(\mathbf{x}_0'\hat{\beta})$

$\mathbf{x}_0$ = the vector of explanatory variables for site $\mathbf{s}_0$

$\hat{\beta}$ = the quasi-likelihood parameter estimates

λ_i = the kriging coefficients based on the variogram and the location of site $\mathbf{s}_0$

and $R(\mathbf{s}_i)$ = the quasi-likelihood residuals at sites $\mathbf{s}_i$.

The kriging variance, σ_k^2, is still the result (equation 5.9) from Section 5.2.2. A nominal 95% prediction interval for $Z(\mathbf{s}_0)$ can be constructed (Cressie, 1993, p. 122):

$$A = \{\hat{\boldsymbol{Z}}(\mathbf{s}_0) - 1.96\,\sigma_k(\mathbf{s}_0), \hat{\boldsymbol{Z}}(\mathbf{s}_0) + 1.96\,\sigma_k(\mathbf{s}_0)\}.$$

5.3.4.3 **Predictions Using Block Kriging**

The ordinary kriging equations and quasi-likelihood parameter estimates are estimated using measured data from the sampled sites, $\mathbf{s}_i$. The explanatory variables, $\mathbf{x}_i$ and the measured counts, $Z(\mathbf{s}_i)$ are assumed to be point estimates. As described in Chapters 1 and 2, this is not really the case because

the samples come from a "box corer" which has a known area and volume. Since the volume is small compared to the volume of lake bottom being sampled, this difference is usually ignored. However, the predictions made using these point estimates may be representative of a volume of sediment whose magnitude cannot be ignored. For example, if 30,000 m^3 of contaminated lake bottom is removed from a harbour, a prediction of the average *Diporeia hoyi* count over the entire new lake bottom may be desired. The explanatory variables may vary substantially over this volume and distances from the known sites, $\mathbf{s}_i$ may depend on what location within the volume is used. In this case, the use of block kriging should be considered.

Cressie (1993, pp. 124-125) describes the general problem of kriging when the data and predictor have different supports. In the example described above, the data have point support, while the predictor may require block support. In this case, the ordinary kriging equations are modified to:

$$\lambda_o = \Gamma_o^{-1}\gamma_o(B)$$

and $$\sigma_k^2(B) = \lambda_o'\gamma_o(B) - \gamma(B,B)$$

where $$\gamma_o(B) = (\gamma(B,\mathbf{s}_1),\ldots,\gamma(B,\mathbf{s}_n),1)'$$

$$\gamma(B,\mathbf{s}_i) = \int_B \gamma(\mathbf{v} - \mathbf{s}_i)d\mathbf{v}/|B|$$

and $$\gamma(B,B) = \int_B\int_B \gamma(\mathbf{u} - \mathbf{v})d\mathbf{u}d\mathbf{v}/|B|^2.$$

This is a case of block kriging. If the geometry of the block, B is known, and is

significant relative to the geometry of the sites, $\mathbf{s}_i$, then block kriging should be considered. In the case of predictions for benthic counts, the geometry of B will not be known and the surface of explanatory variables within B would have to be predicted also.

5.3.4.4 Predictions for *Diporeia hoyi* Counts in Georgian Bay

In Georgian Bay and the North Channel of Lake Huron (Figures 4.7 and 4.8), there are three areas where contaminated sediments have accumulated and impacted the benthic community. These areas are Spanish River, Severn Sound and Collingwood Harbour. If some type of sediment remediation (such as dredging or *in situ* treatment) is attempted, a target *Diporeia hoyi* abundance for the area will be desirable to measure success of the effort. Since the explanatory variables for the quasi-likelihood are based on the physical and geological characteristics of the area, it should be possible to predict the abundance of *diporeia* for typical locations within these areas.

Since several reference sites are located within Severn Sound (Figure 4.7), impacted sites within this area would be candidates for prediction via block kriging. As mentioned in the previous section, this would require prediction of the explanatory variables. The other two sites, Collingwood Harbour (Figure 4.7) and

Spanish River (Figure 4.8), would be candidates for kriging at a point since they are both some distance (ie. 5 - 10 km) from the nearest reference site.

The procedure for obtaining predictions at these points is to measure the distance from nearby reference sites to the point for which the prediction is required. The water distance was measured using GIS as in Chapter 4. Representative values of the explanatory variables at the site to be predicted are also required. However, these can be based on previous measurements since these variables are more representative of geological and physical conditions than they are of contaminant loading. The prediction equations of Section 5.3.4.2 are then applied for the sites requiring prediction. The values of the explanatory variables used for each site are presented in Table 5.3 and the results of the predictions are found in Table 5.4.

The predictions in Table 5.4 are dominated by the effects of the explanatory variables. However, in the case of the Spanish River site, a negative adjustment was made due to the residuals. The interpretation of these predictions is straightforward. For Collingwood Harbour, when clean-up is complete, *Diporeia hoyi* counts should average between 5 and 26 counts per sample. For Spanish River, even when clean-up occurs no *diporeia* are expected because of the high aluminum levels in sediment and low alkalinity

TABLE 5.3

Values of Explanatory Variables at Predicted Sites

Variable	Collingwood Harbour	Spanish River
Depth	22 metres	22 metres
Al_2O_3	5.7 mg/g	11.4 mg/g
K_2O	2.8 mg/g	2.8 mg/g
Alkalinity	75.7 mg/l	17.2 mg/l

TABLE 5.4

Results of Predictions at Two Sites

Variable	Collingwood Harbour	Spanish River
$\hat{Z}(\mathbf{s}_0)$	15.92	0.08
$\hat{\mu}(\mathbf{s}_0)$	15.74	1.69
$\sum_{i=1}^{n} \lambda_i R(\mathbf{s}_i)$	0.18	-1.61
$\sigma_k(\mathbf{s}_0)$	5.35	5.76
Upper C.I. (95%) (for $Z(\mathbf{s}_0)$)	26.4 (two-sided)	11.4 (one-sided)
Lower C.I. (95%) (for $Z(\mathbf{s}_0)$)	5.4 (two-sided)	-

concentrations in the water. In the case of *Diporeia hoyi*, the upper confidence limit is probably not that important because *Diporeia hoyi* is an indicator of clean sediment. This limit may be more important with nuisance species.

These predictions have been made to illustrate the application of kriging with quasi-likelihood residuals. Because the goal of using biologically based clean-up standards is establishment of a healthy benthic community, more than one species will normally be of interest in prediction. This would require a multivariate analog to the model used here. An alternative would be to krige discriminant function scores instead of counts. However, if more than one discriminant function were important for a community, then this procedure would also require a multivariate prediction. Clearly, this is an area for further exploration.

CHAPTER 6: ASYMPTOTIC RESULTS AND SIMULATIONS

6.1 Introduction

The standard errors of the parameter estimates $\hat{\beta}$ in the models developed in Chapters 3 and 4 are all based on $V_{\hat{\beta}}$, the variance-covariance matrix for the parameter estimates. Zeger (1988) has shown that $\sqrt{n}(\hat{\beta}-\beta)$ is asymptotically multivariate Gaussian with zero mean and variance-covariance matrix:

$$V_{\hat{\beta}} = \lim_{n\to\infty} (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\ \partial\mu/\partial\beta\ /n)^{-1}.$$

He then simulated 100 realisations of the parameter-driven model to check if these results hold. In this chapter, the asymptotic results of Zeger will be extended to the case of spatially dependent data and simulations will be conducted using the cross-validated models from Chapter 5 for the case of spatially dependent benthic counts. The results of the simulations will be compared to the asymptotic results to check agreement with asymptotic normality for the parameter estimates, $\hat{\beta}$. The characteristics of the distribution of simulated results for the variogram parameters will also be discussed and compared to results of similar studies.

6.2 Asymptotic Results For Spatially Dependent Quasi-Likelihood Parameters

The following theorem is presented as a proposition in Zeger (1988) and the proof is sketched for the case of generalized estimating equations in Liang and Zeger (1986). The proof is analogous to proofs for asymptotic normality of maximum likelihood estimates (e.g. Zacks, 1971, pp. 246-247). The definitions used here are the same as those used throughout Chapters 2 - 5, but are repeated here for clarity and emphasis.

As in Section 2.1, let $\mathbf{z} = (z_1,\ldots,z_n)'$, $\mathbf{X} = (\mathbf{x}_1, \ldots,\mathbf{x}_n)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, $\mathbf{V} = \text{var}(\mathbf{Z}) = \mathbf{A} + \sigma^2 \mathbf{A}\boldsymbol{R}_e\mathbf{A}$,

where z_i are data values (counts) corresponding to $Z_1,\ldots,Z_n$ dependent random variables, not necessarily identically distributed, but each having a distribution in the exponential family

$\mathbf{X}$ is an $n \times p$ matrix of explanatory variables or a design matrix

$\boldsymbol{\mu} = E(\mathbf{Z}) = \exp(\mathbf{X}\beta)$

β is a $p \times 1$ vector of unknown, fixed parameters

$\mathbf{V}$ is a general symmetric positive definite matrix whose elements are functions of $\boldsymbol{\mu}$

$\mathbf{A} = \text{diag}(\mu_1, \ldots, \mu_n)$

ϵ is an unobserved, stationary process with $E(\epsilon_i) = 1$, $\text{var}(\epsilon_i) = \sigma^2$

and R_e is an n × n correlation matrix that describes the dependence among the ϵ_i.

The following regularity conditions (modified from Zacks, 1971, p.194) must be satisfied in order for Theorem 2, below, to hold:

(i) The parameter space B is either Euclidian r-space $E^{(r)}$ or a rectangle in $E^{(r)}$.

(ii) $\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}(\mathbf{z} - \mu)$ exists for each β_j, j = 1,..., p and all β.

(iii) $E_\beta(\partial/\partial\beta[\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}(\mathbf{z} - \mu)])$ exists and is finite for all β_i, β_j, $1 \le i, j \le p$ and all β.

(iv) The matrix $[\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}\ \partial\mu/\partial\beta]$ is positive definite for all $\beta \in B$.

In addition, the following conditions are needed for Lemmas 1 (Serfling, 1980, p. 27) and 2 (Yasui and Lele, 1997) , below, to hold:

(C1) The series $\sum_{i=1}^{\infty} \rho_i$ must converge, where ρ_i is the correlation between sites separated by distance i.

(C2) The series $\sum_{i=1}^{\infty} \sigma_i^2 (\log i)^2/i^2$ must converge, where σ_i^2 is the variance at site i.

(C3) $E(Y_i) = E(Z_i - \mu_i) = 0$ for each $i \in D$.

(C4) $\{ Y^2_i = (z_i - \mu_i)^2 ; i \in D\}$ is uniformly integrable.

(C5) there exist $\delta > 0$ and $K < \infty$ such that $\rho^d_{[Y_i; i \in D]}(r) \le K(\log r)^{-(2+\delta)}$, where

$\rho^{d}_{[Y_i;\, i\in D]}(r)$ is the maximal correlation coefficient and is equal to:

$$\sup_{\substack{D_1, D_2 \subset D \\ d(D_1, D_2) > r}} \; \sup_{\substack{x_1 \in L_2(\sigma[Y_i;\, i\in D_1]) \\ x_2 \in L_2(\sigma[Y_i;\, i\in D_2])}} |\mathrm{corr}(x_1, x_2)|$$

D is the set of sampling locations

d is a metric on D

and $L_2(\cdot)$ is the set of L_2 random variables measurable with respect to the σ-algebras.

The following lemmas are needed for the proof of Theorem 2, below.

LEMMA 1 (Serfling, 1980, p. 27). Let $Z_1, Z_2, \ldots$ have means $\mu_1, \mu_2, \ldots$, variances $\sigma_1^2, \sigma_2^2, \ldots$, and covariances $\mathrm{Cov}(Z_i, Z_j)$ satisfying

$$\mathrm{Cov}(Z_i, Z_j) \le \rho_{j-i}\, \sigma_i\, \sigma_j \quad (i \le j),$$

where $0 \le \rho_k \le 1$ for all $k = 0, 1, \ldots$.

If the series $\sum_{i=1}^{\infty} \rho_i$ and $\sum_{i=1}^{\infty} \sigma_i^2 (\log i)^2 / i^2$ are both convergent, then the following holds:

$$1/n \sum_{i=1}^{n} Z_i - 1/n \sum_{i=1}^{n} \mu_i \to 0 \text{ with probability} = 1.$$

Proof See Serfling (1970).

LEMMA 2 (Goldie and Greenwood as stated by Yasui and Lele, 1997). If a random field $\{Y_i; i \in D\}$ satisfies the following conditions: (1) $E(Y_i) = 0$ for each $i \in D$, (2) $\{Y^2_i; i \in D\}$ is uniformly integrable, (3) $\rho^d_{[Y_i; i \in D]}(r) \leq K(\log r)^{-(2+\delta)}$ with some $K < \infty$ and $\delta > 0$, then

(a) $V_n \leq nC(\sup_{i \in D} \text{var}[Y_i])$ with some $C < \infty$, and

(b) $\sum_{i \in D_n} Y_i / \sqrt{V_n}$ converges in distribution to the standard Gaussian distribution as $n \to \infty$,

where $\rho^d_{[Y_i; i \in D]}(r)$ is the maximal correlation coefficient and is equal to:

$$\sup_{\substack{D_1, D_2 \subset D \\ d(D_1, D_2) > r}} \; \sup_{\substack{x_1 \in L_2(\sigma[Y_i; i \in D_1]) \\ x_2 \in L_2(\sigma[Y_i; i \in D_2])}} |\text{corr}(x_1, x_2)|$$

D is the set of sampling locations

d is a metric on D

$D_n \subset D$ is a set of n spatial locations where observations are sampled

$L_2(\cdot)$ is the set of L_2 random variables measurable with respect to the σ-algebras

and $V_n = \text{var}[\sum_{i \in D_n} Y_i]$.

Proof See Goldie and Greenwood (1986a, 1986b).

LEMMA 3. (The Cramer-Wold Device) In R^p, Y_n converges in distribution to Y if and only if each linear combination of the components of Y_n converges in distribution to the corresponding linear combination of the components of Y.

Proof See Billingsley (1968, pp. 48-49).

LEMMA 4. The asymptotic distribution of $U\{\beta,\theta\}/\sqrt{n}$ is multivariate Gaussian with zero mean and covariance matrix:

$$\lim_{n\to\infty} (\partial\mu'/\partial\beta\ V^{-1}Cov(Z)V^{-1}\ \partial\mu/\partial\beta\ /n)$$

where $U\{\beta,\theta\}$ is the score function defined previously.

Proof Let $v^{-1}..$ be the elements of V^{-1}. $U\{\beta,\theta\}$ is a p-dimensional random vector whose components,

$$U_j\{\beta,\theta\} = \partial\mu'/\partial\beta_j\ V^{-1}(z-\mu) = \sum_{i=1}^{n}\sum_{k=1}^{n} v^{-1}_{i,k}\ \partial\mu_k/\partial\beta_j(z_i-\mu_i),$$

each involve the sum of the residuals, $z_i-\mu_i$ (i.e. $\sum_{i\in D_n} Y_i$ from Lemma 2). Assuming that the conditions of Lemma 2 are satisfied for each component, $U_j\{\beta,\theta\}/\sqrt{n}$ is asymptotically normal with variance:

$$var(U_j\{\beta,\theta\}/\sqrt{n}) = \lim_{n\to\infty} var(\sum_{i=1}^{n}\sum_{k=1}^{n} v^{-1}_{i,k}\ \partial\mu_k/\partial\beta_j(z_i-\mu_i))/n$$

$$= \lim_{n\to\infty}[\sum_{m=1}^{n} \text{var}(Z_m) \sum_{i=1}^{n}\sum_{k=1}^{n} v^{-1}{}_{m,i} v^{-1}{}_{k,i}\, \partial\mu_i/\partial\beta_j \partial\mu_k/\partial\beta_j$$

$$+ 2\sum_{m=1}^{n} \sum_{\substack{l=1 \\ l>m\neq i}}^{n} \text{cov}(Z_m Z_l) \sum_{i=1}^{n}\sum_{k=1}^{n} v^{-1}{}_{m,i} v^{-1}{}_{k,i}\, \partial\mu_i/\partial\beta_j \partial\mu_k/\partial\beta_j]/n \quad (6.1)$$

By Lemma 3, $\mathbf{U}\{\beta,\theta\}/\sqrt{n}$ is multivariate Gaussian with zero mean and covariance matrix:

$$\lim_{n\to\infty} (\partial\boldsymbol{\mu}'/\partial\beta\ \mathbf{V}^{-1}\text{Cov}(\mathbf{Z})\mathbf{V}^{-1}\ \partial\boldsymbol{\mu}/\partial\beta\ /n)$$

which is a p × p matrix formed from equation 6.1.

THEOREM 2. Let $\hat{\theta}$ be a $\sqrt{n}$-consistent estimate of θ depending on **z** and β. Let $\hat{\beta}$ be the solution of:

$$\mathbf{U}(\beta) = \mathbf{U}\{\beta,\hat{\theta}(\beta)\} = \partial\boldsymbol{\mu}'/\partial\beta\ \mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}(\mathbf{z} - \boldsymbol{\mu})$$

where $\mathbf{U}(\beta)$ is the score function of the quasi-likelihood function for the case of dependent errors (Sections 2.1.4 and 3.2.3)

and $\hat{\theta}(\beta)$ are parameter estimates such as $\hat{\sigma}^2$ and $\hat{\rho}$ or the sill and range of the semivariogram.

Suppose ϵ is a stationary process. Under regularity conditions (i) - (iv) and given that $\sqrt{n}(\hat{\theta}-\theta) = O_p(1)$ for some fixed θ, $\sqrt{n}(\hat{\beta}-\beta)$ is asymptotically multivariate Gaussian with zero mean and variance-covariance matrix:

$$V_{\hat{\beta}} = \lim_{n\to\infty} (\partial\mu'/\partial\beta\ V^{-1}\ \partial\mu/\partial\beta\ /n)^{-1}.$$

Proof Let $O_p(\cdot)$ and $o_p(\cdot)$ refer to the rate of convergence in probability. Let $X_n = o_p(Y_n)$ as $n \to \infty$ if $X_n/Y_n \xrightarrow{P} 0$; and let $X_n = O_p(Y_n)$ if there exists a constant K, $0 < K < \infty$, such that $\lim_{n\to\infty} P([|X_n/Y_n| \le K] = 1$. Expand $U\{\beta,\hat{\theta}(\beta)\}/\sqrt{n}$ around $\beta = \beta_0$:

$$U\{\beta,\hat{\theta}(\beta)\}/\sqrt{n} = U\{\beta_0,\hat{\theta}(\beta_0)\}/\sqrt{n} + [\delta/\delta\beta\ U\{\beta',\hat{\theta}(\beta')\}/n]\sqrt{n}(\beta-\beta_0) \quad (6.2)$$

where $\delta/\delta\beta\ U\{\beta',\hat{\theta}(\beta')\} = \partial/\partial\beta\ U\{\beta',\hat{\theta}(\beta')\}$

$$+ [\partial/\partial\theta U\{\beta',\hat{\theta}(\beta')\}][\partial\hat{\theta}(\beta')/\partial\beta]$$

and β' is a point on the line segment connecting β and β_0.

Substituting into equation 6.2, if $\hat{\beta}$ is a quasi-likelihood estimate of β,

$$U\{\hat{\beta},\hat{\theta}(\hat{\beta})\}/\sqrt{n} = U\{\beta,\hat{\theta}(\beta)\}/\sqrt{n}$$

$$+ [\delta/\delta\beta\ U\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\}/n]\sqrt{n}(\hat{\beta}-\beta) = 0. \quad (6.3)$$

where $\tilde{\beta}$ is a point on the line segment connecting $\hat{\beta}$ and β

and $[\delta/\delta\beta\ U\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\} = \partial/\partial\beta\ U\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\} + [\partial/\partial\theta U\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\}][\partial\hat{\theta}(\tilde{\beta})/\partial\beta]$

$$= \partial/\partial\beta\ U\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\} + AB. \quad (6.4)$$

Rearranging equation 6.3:

$$-U\{\beta,\hat{\theta}(\beta)\}/\sqrt{n} = [\delta/\delta\beta\ U\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\}/n]\sqrt{n}(\hat{\beta}-\beta) \quad (6.5)$$

By regularity condition (iii),

$$E_\beta[\partial/\partial\beta\ U\{\beta,\hat{\theta}(\beta)\}] = E_\beta(\partial/\partial\beta[\partial\mu'/\partial\beta\ V^{-1}\{\beta,\hat{\theta}(\beta)\}(z-\mu)])$$

$$= E_\beta(\partial^2\mu'/\partial\beta^2\ \mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}(\mathbf{z}-\mu)$$

$$+ \partial\mu'/\partial\beta[\partial/\partial\beta\ \mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}](\mathbf{z}-\mu)$$

$$- \partial\mu'/\partial\beta\ \mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}\partial\mu/\partial\beta)$$

$$= - \partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta.$$

So for each n = 1, 2, ...,

$$E_\beta[\partial/\partial\beta\ \mathbf{U}\{\beta,\hat{\theta}(\beta)\}/n] = - (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta)/n.$$

In equation 6.4 above, as $n \to \infty$, $A = o_p(1)$ by Lemma 1 since $\partial/\partial\theta\ \mathbf{U}\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\}$ is a linear function of $(\mathbf{z} - \mu)$ and $B = O_p(1)$ since θ is fixed and $\hat{\theta}$ is a consistent estimate of θ. Also, since $\hat{\beta} \to \beta$ as $n \to \infty$, so does $\tilde{\beta}$, since $\tilde{\beta} = v\hat{\beta} + (1 - v)\beta$, $0 \le v \le 1$. By Lemma 1,

$$[\partial/\partial\beta\ \mathbf{U}\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\}/n] = - (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta)/n + o_p(1).$$

By regularity condition (iv), $[\partial/\partial\beta\ \mathbf{U}\{\tilde{\beta},\hat{\theta}(\tilde{\beta})\}/n]^{-1}$ exists for n sufficiently large and so, as $n \to \infty$, the above asymptotic results can be substituted into equation 6.5 to obtain:

$$\sqrt{n}(\hat{\beta}-\beta) = - [(\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta)/n]^{-1}\ \mathbf{U}\{\beta,\hat{\theta}(\beta)\}/\sqrt{n}. \qquad (6.6)$$

Now let β be fixed and expand $\mathbf{U}\{\beta,\theta\}/\sqrt{n}$ around $\theta = \theta_0$:

$$\mathbf{U}\{\beta,\theta\}/\sqrt{n} = \mathbf{U}\{\beta,\theta_0\}/\sqrt{n} + [\partial/\partial\theta\ \mathbf{U}\{\beta,\theta'\}/n]\sqrt{n}(\theta - \theta_0)$$

where θ' is a point on the line segment connecting θ and θ_0.

Substituting $\hat{\theta}$ for θ_0:

$$\mathbf{U}\{\beta,\hat{\theta}(\beta)\}/\sqrt{n} = \mathbf{U}\{\beta,\theta\}/\sqrt{n} + [\partial/\partial\theta\ \mathbf{U}\{\beta,\tilde{\theta}\}/n]\sqrt{n}(\hat{\theta} - \theta)$$

where $\tilde{\theta}$ is a point on the line segment connecting $\hat{\theta}$ and θ. Since $\hat{\theta}$ is $\sqrt{n}$-consistent, $\tilde{\theta} = v\hat{\theta} + (1 - v)\theta$, $0 \leq v \leq 1$, is also $\sqrt{n}$-consistent and, as $n \to \infty$, $\tilde{\theta} \to \theta$ and $\partial/\partial\theta\ \mathbf{U}\{\beta, \tilde{\theta}\}/n \to \partial/\partial\theta\ \mathbf{U}\{\beta, \theta\}/n$, so

$$\mathbf{U}\{\beta, \hat{\theta}(\beta)\}/\sqrt{n}] = \mathbf{U}\{\beta, \theta\}/\sqrt{n} + [\partial/\partial\theta\ \mathbf{U}\{\beta, \theta\}/n]\sqrt{n}(\hat{\theta} - \theta) + o_p(1)$$

$$= A^* + B^*C^* + o_p(1)$$

Following the proof by Liang and Zeger (1986), B^* is $o_p(1)$ by Lemma 1, since $\partial/\partial\theta\ \mathbf{U}\{\beta, \theta\}$ is a linear function of $(\mathbf{z} - \mu)$, C^* is $O_p(1)$ and $\mathbf{U}\{\beta, \hat{\theta}(\beta)\}/\sqrt{n}$ is asymptotically equivalent to A^* whose asymptotic distribution is multivariate Gaussian with zero mean and covariance matrix:

$$\lim_{n \to \infty} (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\mathrm{Cov}(\mathbf{Z})\mathbf{V}^{-1}\ \partial\mu/\partial\beta\ /n)$$

by Lemma 4.

So, from equation 6.6, $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically multivariate Gaussian with zero mean and variance-covariance matrix:

$$\lim_{n \to \infty} n(\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta)^{-1}(\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\mathrm{Cov}(\mathbf{Z})\mathbf{V}^{-1}\partial\mu/\partial\beta)$$

$$(\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta)^{-1}$$

If $\mathbf{V} = \mathrm{Cov}(\mathbf{Z})$ as assumed in the quasi-likelihood model, then

$$\sqrt{n}(\hat{\beta} - \beta) \sim \mathrm{MVN}(0, \boldsymbol{V}_{\hat{\beta}}) \text{ (asymptotically)}$$

where $\boldsymbol{V}_{\hat{\beta}} = \lim_{n \to \infty} (\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta/n)^{-1}$.

This completes the proof.

6.3 Verification Of Conditions For Simulation

The model used for simulation in the next section is the spatially dependent quasi-likelihood model for counts with classical semivariogram estimator and water distance developed in Chapters 3 and 4 and cross-validated in Chapter 5. The above conditions are checked with respect to this model.

Condition (i) β is a $p \times 1$ vector and $-\infty < \beta_j < \infty$ for each $j = 1, \ldots, p$, therefore the parameter space is a real vector space, closed under addition and scalar multiplication. So the parameter space B is Euclidian r-space $E^{(r)}$ with $r=p$.

The following theorems from Anderson (1958, p. 337) are used to verify conditions (ii)-(iv) below:

(I) If **C** is positive definite and **B** is non-singular, then **B'CB** is positive definite.

(II) If **C** is positive definite, then $\mathbf{C}^{-1}$ is positive definite.

(III) If **C** with p rows and columns is positive definite and if **B** with p rows and q columns, $q \leq p$ is of rank q, then **B'CB** is positive definite.

Condition (ii) The matrix $\partial\boldsymbol{\mu}'/\partial\beta$ consists of elements $\partial\mu_i/\partial\beta_j = x_{ij}\exp(\mathbf{x}_i\beta)$ which are defined for all i = 1,...,n and j = 1,...,p. The elements of $\boldsymbol{R}_e$ are $\exp(-d(\mathbf{s},\mathbf{t})/\hat{\theta}_3)$, where $\hat{\theta}_3$ is the range of the semivariogram and d(**s**,**t**) is the water distance between sites **s** and **t** as defined in Chapter 4. $\hat{\sigma}^2$ is obtained from the sill of the semivariogram (see Chapter 4). $\hat{\sigma}^2\boldsymbol{R}_e$ is the exponential covariance function which is positive definite by definition (Haining, 1990, pp. 90-91). The variance-covariance matrix used in the estimating equations is:

$$\mathbf{V} = \mathbf{A} + \hat{\sigma}^2\mathbf{A}'\boldsymbol{R}_e\mathbf{A} \qquad (6.7)$$

where $\mathbf{A} = \mathrm{diag}(\mu_1, \mu_2, \ldots \mu_n)$ as defined previously.

Since **A** is non-singular (i.e. $\mu_i = \exp(\mathbf{x}_i\beta) > 0$), $\mathbf{A}'\boldsymbol{R}_e\mathbf{A}$ is positive definite by I. Likewise, **V** is an n × n, positive definite matrix and, by II, so also is $\mathbf{V}^{-1}$ and therefore $\mathbf{V}^{-1}$ exists. $\mathbf{z} - \boldsymbol{\mu}'$ is an n × 1 vector of differences between observed and expected counts,

$z_i - \mu_i$, and exists for all i = 1,...,n.

Condition (iii) $E_\beta(\partial/\partial\beta[\partial\boldsymbol{\mu}'/\partial\beta\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}(\mathbf{z} - \boldsymbol{\mu})]) =$

$$E_\beta(\partial^2\boldsymbol{\mu}'/\partial\beta^2\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}(\mathbf{z} - \boldsymbol{\mu})$$
$$+ \partial\boldsymbol{\mu}'/\partial\beta[\partial/\partial\beta\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}](\mathbf{z} - \boldsymbol{\mu})$$
$$- \partial\boldsymbol{\mu}'/\partial\beta\ \mathbf{V}^{-1}\{\beta, \hat{\theta}(\beta)\}\partial\boldsymbol{\mu}/\partial\beta).$$

The matrix $\partial^2\boldsymbol{\mu}'/\partial\beta^2$ consists of elements $\partial^2\mu_i/\partial\beta_j\beta_k = x_{ij}\,x_{ik}\exp(\mathbf{x}_i\beta)$ which are

defined for all i = 1,...,n and j, k = 1,...,p.

$[\partial/\partial\beta\ \mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}] = -\mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}[\partial/\partial\beta\ \mathbf{V}\{\beta,\hat{\theta}(\beta)\}]\mathbf{V}^{-1}\{\beta,\hat{\theta}(\beta)\}$ (Morrison, 1976, pp. 74-75). The matrix $[\partial/\partial\beta\ \mathbf{V}\{\beta,\hat{\theta}(\beta)\}]$ consists of elements:

$$\partial v_{ik}/\partial\beta_j = 3\hat{\sigma}^2 x_{ij} \exp(\mathbf{x}_i\beta) + [\partial/\partial\beta_j\,\hat{\sigma}^2](\exp(\mathbf{x}_i\beta))(1 + \exp(\mathbf{x}_i\beta)) \qquad i = k$$

$$= (\exp(\mathbf{x}_i\beta)\exp(\mathbf{x}_k\beta))\{\hat{\sigma}^2 \boldsymbol{R}_\epsilon(x_{ij} + x_{kj}) + [\partial/\partial\beta_j\,\hat{\sigma}^2]\boldsymbol{R}_\epsilon + [\partial/\partial\beta_j \boldsymbol{R}_\epsilon]\hat{\sigma}^2\} \quad i \neq k$$

Since $\hat{\sigma}^2$ and $\boldsymbol{R}_\epsilon$ are fixed, $[\partial/\partial\beta_j\,\hat{\sigma}^2]$ and $[\partial/\partial\beta_j \boldsymbol{R}_\epsilon]$ are zero and so,

$$\partial v_{ik}/\partial\beta_j = 3\hat{\sigma}^2 x_{ij} \exp(\mathbf{x}_i\beta) \qquad i = k$$

$$= (\exp(\mathbf{x}_i\beta)\exp(\mathbf{x}_k\beta))\{\hat{\sigma}^2 \boldsymbol{R}_\epsilon(x_{ij} + x_{kj})\} \qquad i \neq k$$

which are defined for all i, k = 1,...,n and j = 1,...,p.

Condition (iv) By III, since $\partial\mu/\partial\beta$ is a matrix of n rows and p columns with rank p (McCullagh and Nelder, 1983, p. 168), $\partial\mu'/\partial\beta\ \mathbf{V}^{-1}\partial\mu/\partial\beta$ is positive definite.

Condition (C1) Since the exponential covariance function has been used to account for spatial correlation among sites, the series $\sum_{i=1}^{\infty} \exp(-\theta_3 i)$ must converge. By the ratio test for convergence of series (Kaplan, 1973, p. 385), since the limit

$$\lim_{n\to\infty} |\exp(-\theta_3(n + 1))/\exp(-\theta_3 n)| = 0.881$$

is less than one, the series converges and Condition (C1) is satisfied.

Condition (C2) The expected count at site i, μ_i, is finite by definition. Let $M^* > \mu_i$ for i= 1... n, where n is the number of sites. Then $M = M^* + \sigma^2 M^{*2}$ (where σ^2 is estimated from the sill of the semivariogram). The variance at each site, σ_i^2, is $\mu_i + \sigma^2\mu_i^2$ and is less than M. If $\sum_{i=1}^{\infty} M (\log i)^2/i^2$ converges, then $\sum_{i=1}^{\infty} \sigma_i^2(\log i)^2/i^2$ converges. Since M is a constant, it suffices to show that $\sum_{i=1}^{\infty} (\log i)^2/i^2$ converges. The inequality

$$((\log i)^2/i^{3/4})\ (1/i^{5/4}) < (1/i^{5/4}) \text{ for all } i > 0$$

holds because $(\log i)^2/i^{3/4} < 1$ for all $i > 0$. Thus, by the comparison test for convergence of series (Kaplan, 1973, p. 383), $\sum_{i=1}^{\infty} (\log i)^2/i^2$ converges since $\sum_{i=1}^{\infty} 1/i^{5/4}$ converges (a harmonic series of order greater than 1 converges (Kaplan, 1973, p.384)). Therefore Condition (C2) is satisfied.

Condition (C3) The Y_i are the residuals, $z_i - \mu_i$, i = 1...79, whose expectations are zero. Thus each Y_i satisfies Condition (C3).

Condition (C4) Serfling (1980, pp. 13-14) presents a lemma that states, in part, that:

> Sufficient for uniform integrability of $\{\eta_i\}$ is that
>
> $$\sup_i E|\eta_i|^{1+\epsilon} < \infty$$
>
> for some $\epsilon > 0$.

If $\epsilon = 1$, this condition reduces to the requirement that the variance of the least

upper bound of the counts be finite (i.e. less than M in Condition (C2)) and so Condition (C4) is satisfied.

Condition (C5) Since the exponential covariance function has been used to account for spatial correlation among sites, the maximal correlation coefficient $\rho^{d}_{[Y_i;\, i\in D]}(r) = \exp(-\theta_3 r)$, where r is less than the water distance ($d(D_1, D_2)$) between two sites. For the example being simulated in the next section, the smallest water distance is 0.8 km which occurs twice among all of the site pairs. The largest water distance is 64.8 km which occurs for one pair (See Chapter 4). With $\hat{\theta}_3 = 0.1271$ substituted for θ_3, choose $\delta = 2$ and $K = 10$. For values of r near 1km, $K(\log r)^{-(2+\delta)}$ is quite large (>100) and $\exp(-\theta_3 r) < 1$ for all $r > 0$. Therefore the values of r that need to be checked are those for which $K(\log r)^{-(2+\delta)} < 1$. This occurs for the interval [5.92,64.79]. Over this range, the function $K(\log r)^{-(2+\delta)}$ is monotone decreasing from 0.9999 to 0.0330. Over the same range, $\exp(-\theta_3 r)$ is also monotone decreasing from 0.47122 to 0.0003. Therefore,

$$\rho^{d}_{[Y_i;\, i\in D]}(r) \le K(\log r)^{-(2+\delta)} \text{ for all } r \in [0.79,64.79]$$

and condition (C5) is satisfied.

The process ϵ is stationary by definition and by condition C5, it is

also a mixing process. Therefore, Theorem 2 may be applied in the following simulation (Section 6.4) if $\hat{\sigma}^2$ and $\hat{\theta}_3$ from the weighted least squares fit of the semivariogram are $\sqrt{n}$-consistent . The required consistency may be seen as follows.

The weighted least squares method for fitting the semivariogram model relies on minimizing the quantity

$$(2\hat{\gamma} - 2\gamma(\theta))' \mathbf{V2}(\theta)^{-1} (2\hat{\gamma} - 2\gamma(\theta)) \qquad (6.8)$$

where $\hat{\gamma}$ is the classical estimator of the semivariogram

$\gamma(\theta)$ is the semivariogram model to be fit

$\mathbf{V2}(\theta)$ is approximated by diag $\{\text{var}(2\hat{\gamma}(\mathbf{h}(j)))\}$, $j = 1,\ldots,K$

$\text{var}(2\hat{\gamma}(\mathbf{h}(j))) \approx 2\{2\gamma(\mathbf{h}(j);\theta)\}^2 / |N(\mathbf{h}(j))|$

$\mathbf{h}(j)$ is the jth lag or distance

K is the maximum number of lags in the semivariogram

and $|N(\mathbf{h}(j))|$ is the number of distinct pairs at each lag.

Gallant and White (1988, Ch. 2 and 3) define a broad class of nonlinear "optimization" estimators of the form

$$\hat{\theta}_n = \min \psi(\theta)' \hat{P}_n \psi(\theta)$$

where $\psi(\theta)$ is a function of θ, the parameter vector to be estimated and the data

and $\hat{P}_n$ is an estimate of the asymptotic variance of $\psi(\theta)$.

They then prove a theorem which states that under regularity conditions, $\hat{\theta}_n - \theta \to 0$ almost surely as $n \to \infty$. The regularity conditions are very similar to conditions (i) - (iv) and (C1) - (C5) in Section 6.2. Minimizing the quantity in equation 6.2 leads to an "optimization" estimator $\hat{\theta}_n$ which is consistent.

McShane et al. (1997) have used an estimating equation approach to obtain consistent estimates for the correlation structure parameters in Zeger's (1988) model. Their estimating equation is an alternate version of that used by Diggle et al. (1994, p. 165) and uses a working variance matrix which is analogous to $\mathbf{V2}(\theta)^{-1}$ in equation 6.8. The McShane et al. equation (with variable names changed to avoid confusion) is

$$\mathbf{E}(\theta_\rho, \sigma^2, \beta)' \, \mathbf{W}^{-1} \, (\mathbf{y}(\beta) - \mathbf{v}(\theta_\rho, \sigma^2, \beta)) = 0 \qquad (6.9)$$

where $\mathbf{E}(\theta_\rho, \sigma^2, \beta)' = \partial \mathbf{v}(\theta_\rho, \sigma^2, \beta) / \partial \theta_\rho$

$\mathbf{W} = \text{var}(\mathbf{y}(\beta))$

$\mathbf{y}(\beta)$ is a vector of n(n-1)/2 elements consisting of all entries below the diagonal of the matrix of residuals, **Y**, as defined in the previous section

and $\mathbf{v}(\theta_\rho, \sigma^2, \beta)$ is a vector of n(n-1)/2 elements consisting of all entries below the diagonal of the matrix **V** as defined above (equation 6.7).

McShane et al. (1997) chose **W** = **I**, but note that "...consistent estimates of the correlation structure parameters are obtained for any choice of **W**", as discussed by Prentice (1988). Taking the partial derivative of equation 6.8 with respect to θ results in estimating equations similar to equation 6.9:

$$-2\,(\partial\gamma(\theta)/\partial\theta)'\,\mathbf{V2}(\theta)^{-1}\,(2\hat{\gamma} - 2\gamma(\theta)) = 0 \qquad (6.10)$$

Equations 6.9 and 6.10 have similar components and would yield the same estimates for $\hat{\theta}$ if **W** is chosen to be equal to **V2**(θ). As Cressie (1993, p. 97) has stated, the use of **V2**(θ) is attractive because more weight is given to the smaller lags.

Equation 6.10 results in the following variance matrix for $\hat{\theta}$:

$$\mathrm{var}(\hat{\theta}) \approx [(\partial\gamma(\theta)/\partial\theta)'\,\mathbf{V2}(\theta)^{-1}\,(\partial\gamma(\theta)/\partial\theta)]^{-1}\,.$$

If **V2**(θ) is approximated as in equation 6.8, and the exponential model used for the variogram, with θ_2 as the sill and θ_3 as the range, then $\mathrm{var}(\hat{\theta}) \approx \mathbf{I}(\hat{\theta})^{-1}$

where $$I_{11} = (1/\hat{\theta}_2^2) \sum_{j=1}^{K} |N(\mathbf{h}(j))|$$

$$I_{12} = I_{21} = (1/\hat{\theta}_2) \sum_{j=1}^{K} |N(\mathbf{h}(j))|\,\mathbf{h}(j)\,/(\exp(\mathbf{h}(j)\,\hat{\theta}_3) - 1)$$

and $$I_{22} = \sum_{j=1}^{K} |N(\mathbf{h}(j))|\,\mathbf{h}(j)^2\,/(\exp(\mathbf{h}(j)\,\hat{\theta}_3) - 1)^2.$$

Using this approximation, the estimates for θ will be $\sqrt{n}$-consistent if the information, $\mathbf{I}(\hat{\theta})$ grows as n grows, or if the smallest latent root of $\mathbf{I}(\hat{\theta})$ tends to ∞ as $n \to \infty$ (Mardia and Marshall, 1984). The smallest latent root of $\mathbf{I}(\hat{\theta})$ can be found from the smaller root, λ_2, of the characteristic equation for $\mathbf{I}(\hat{\theta})$:

$$(\lambda - I_{11})(\lambda - I_{22}) - (I_{12})^2. \qquad (6.11)$$

λ_2 can be found from the quadratic formula. Rearranging equation 6.11,

$$\lambda^2 - \lambda(I_{11} + I_{22}) + (I_{11}I_{22} - (I_{12})^2).$$

Now λ is the solution of:

$$\lambda = -B/2 \pm \sqrt{B^2 - 4AC}/2$$

where $A = 1$

$$B = I_{11} + I_{22}$$

$$C = (I_{11}I_{22} - (I_{12})^2) = \det(\mathbf{I}(\hat{\theta})).$$

So,

$$\lambda_2 = (I_{11} + I_{22})/2 - \sqrt{(I_{11}+I_{22})^2 - 4\det(I(\hat{\theta}))}/2.$$

Since $I_{11} = (1/\theta_2^2)\sum_{j=1}^{K} |N(\mathbf{h}(j))|$ which is equal to $n(n-1)/2\theta_2^2$, λ_2 will tend to ∞ as $n \to \infty$ if

$$I_{11} + I_{22} > \sqrt{(I_{11}+I_{22})^2 - 4\det(I(\hat{\theta}))}.$$

Squaring both sides of the inequality,

$$(I_{11} + I_{22})^2 > (I_{11} + I_{22})^2 - 4\det(\mathbf{I}(\hat{\theta}))$$

or, $$4\det(\mathbf{I}(\hat{\theta})) > 0.$$

By the Cauchy inequality for sums,

$$\det(I(\hat{\theta})) = [\{(1/\hat{\theta}_2^2)\sum_{j=1}^{K}|N(h(j))|\}\{\sum_{j=1}^{K}|N(h(j))|\,h(j)^2/(\exp(h(j)\,\hat{\theta}_3)-1)^2\}$$

$$-\{(1/\hat{\theta}_2)\sum_{j=1}^{K}|N(h(j))|\,h(j)/(\exp(h(j)\,\hat{\theta}_3)-1)\}^2] > 0$$

and therefore $\hat{\theta}$ is $\sqrt{n}$-consistent.

6.4 Simulation

6.4.1 Introduction

Ripley (1981, p. 16) motivates the need for simulation as follows: "Suppose we are interested in the distribution of a statistic, T, which may be unavailable analytically or have an asymptotic or approximate answer the validity of which is unknown." The asymptotic results of the previous section are a case in point. Then simulation is used to compare empirical, simulated results to theory. Haining (1990, p. 116) notes that: "The fit of a model can be assessed by comparing properties of the data with properties derived from realisations of the model." In this section, the usual technique of simulation of realisations from a spatial model assuming Gaussian data will be described. Next, the procedure will be adopted for quasi-likelihood residuals. Finally, simulations will be conducted

with the model of benthic counts from Chapter 4 to check the asymptotic results of the previous section and provide another assessment of the model fit.

6.4.2 Simulation of a Gaussian Spatial Process

General techniques are available (Ripley, 1981, pp. 16-18; Haining, 1990, pp. 116-117) for simulation of a spatial process whose underlying distribution is multivariate normal (MVN). Suppose we wish to generate $Z_1,...,Z_n$ joint normal random variables with means $\mu_1,...,\mu_n$ and covariance Σ, ie. a drawing (**z**) from MVN($\boldsymbol{\mu}, \Sigma$). The procedure is to find **L** with **LL'** = Σ, and then calculate $z_i = \mu_i + \sum_{j=1}^{n} l_{ij}\epsilon_j$ where $\epsilon_1,...,\epsilon_n$ are independent, identically distributed, N(0,1) random variables. **L** may be found in general by the Cholesky decomposition of Σ, or specified directly as in the case of Upton and Fingleton's model where $\Sigma = \sigma^2((\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W}))^{-1}$ and $\mathbf{L} = \sigma(\mathbf{I} - \rho\mathbf{W})^{-1}$.

6.4.3 Simulation of a Spatial Quasi-Likelihood Process

Zeger (1988) conducted simulations with his parameter-driven model using a log-Gaussian error process with the same moments as were found for the polio example data set. To continue the analogy with time series, a similar process will be used for simulation of the spatial quasi-likelihood process. Zeger

used the process:

$$\delta_t = \log \epsilon_t$$

which was assumed to be Gaussian AR(1) with moments for ϵ_t equal to the example. In order to apply this to the example of spatially dependent benthic counts, the method of Aitchison and Brown (1963, p. 28) for generating log-Gaussian random variables is used with the procedure described in Section 6.4.2 to introduce spatial correlation.

The procedure of Aitchison and Brown is to generate log-Gaussian deviates, w_i using

$$w_i = \exp(\mu_\delta + \sigma_\delta y_i) = \exp(\delta_i)$$

where $\mu_\delta, \sigma_\delta$ are the first and second moments of the process δ

and y_i are N(0,1) random deviates.

Because the expectation and variance of a log-normally distributed random variable with $\mu_\delta = 0$ are $\exp(\sigma_\delta^2/2)$ and $\exp(\sigma_\delta^2)(\exp(\sigma_\delta^2) - 1)$, respectively, adjustments have to be made to obtain the required moments for ϵ_i. Since the quasi-likelihood model assumes $E(\epsilon_i) = 1$ and $Var(\epsilon_i) = \sigma^2$, choose $\epsilon_i = \exp(\delta_i - \sigma_\delta^2/2)$ and choose σ_δ so that $\exp(\sigma_\delta^2) - 1 = \sigma^2$. To introduce spatial autocorrelation in ϵ_i, let $\delta_i = \sigma_\delta y_i$ as above, and form

$$\epsilon_i = \exp(-\sigma_\delta^2/2 + \sum_{j=1}^{n} l_{ij}\delta_j)$$

where l_{ij} is the ith,jth element of the L matrix which is obtained from the

Cholesky decomposition of the correlation matrix of $R(\mathbf{s}_i)R(\mathbf{s}_j)$ from Section 5.3.

Once the vector ϵ has been formed as above, it can be multiplied elementwise by the $\hat{\mu}$ vector from the quasi-likelihood model to obtain $E(\mathbf{z}|\epsilon)$. Since $\mathbf{z}$ conditional on ϵ is assumed to be independent with both mean and variance equal to $E(\mathbf{z}|\epsilon)$, the simulated z_is are generated as Poisson deviates with means $\hat{\mu}_i\epsilon_i$.

6.4.4 Simulation Results

The above simulation procedure was used to obtain 1000 realisations of the quasi-likelihood model with spatially correlated errors to check the asymptotic results of Section 6.2. The case of distance determined by measurement using a GIS (water distance) and variogram estimated by the classical estimator was used because this combination performed well under cross-validation (Chapter 5). For comparison, 1000 realisations assuming independent errors were also obtained. Ideally, more simulations should be performed to fully explore the behaviour of the model. However, the variogram fitting is time (CPU) consuming and is dependent on initial values for efficiency. Automated selection of initial values for variogram fitting is necessary for lengthy

simulation runs, but this is beyond the scope of this study.

The results of the simulations are presented in Tables 6.1 - 6.4. Tables 6.1 and 6.2 contain information on bias for the simulated parameters, $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\theta}_3$ for the case of spatially correlated error and independent errors, respectively. Tables 6.3 and 6.4 contain the asymptotic 95% confidence intervals, coverage proportions and the Shapiro-Wilk test for normality for each element of $\hat{\beta}$. The bias for all of the estimated regression coefficients was quite low except for the intercept, $\hat{\beta}_0$. However, bias in the variogram parameters, $\hat{\sigma}^2$ and $\hat{\theta}_3$ was quite large. In the case of independent errors, where $\hat{\sigma}^2$ is estimated by the method of moments, its bias was still substantial. These results are consistent with the simulations in Zeger (1988) who found bias of about 15% in the estimated $\hat{\sigma}^2$.

Harris and Johnson (1996) conducted simulations with a regression model with spatially correlated errors to study the distributions of $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\theta}$. In this study, the explanatory variables were *in situ* measurements of temperature at a grassland site and the modelled variable was the remotely sensed temperature. $\hat{\beta}$ were maximum likelihood regression parameters and $\hat{\sigma}^2$ was the estimated variance of a Gaussian error process analogous to Upton and Fingleton's model. However, $\hat{\theta}$ was the range parameter of an exponential

TABLE 6.1

Simulation Results (Bias) for *Diporeia hoyi* (Water Distance and Classical Variogram Estimator)

m = 1000 Realisations

Parameter	True Value	Mean	Bias	Percent
$\hat{\beta}_0$	-2.45	-2.76	-0.31	-12.6
$\hat{\beta}_1$	0.758	0.745	-0.013	-1.72
$\hat{\beta}_2$	-1.41	-1.37	0.04	2.84
$\hat{\beta}_3$	1.75	1.80	0.05	2.86
$\hat{\beta}_4$	0.83	0.87	0.04	4.82
$\hat{\sigma}^2$	0.456	0.341	-0.115	-25.2
$\hat{\theta}_3$	0.1271	0.3243	0.1972	155.2

TABLE 6.2

Simulation Results (Bias) for *Diporeia hoyi* (Independent Errors)

m = 1000 Realisations

Parameter	True Value	Mean	Bias	Percent
$\hat{\beta}_0$	-2.45	-2.84	-0.39	-15.9
$\hat{\beta}_1$	0.758	0.750	-0.008	-1.06
$\hat{\beta}_2$	-1.41	-1.35	0.06	4.26
$\hat{\beta}_3$	1.75	1.78	0.03	1.71
$\hat{\beta}_4$	0.83	0.88	0.05	6.02
$\hat{\sigma}^2$	0.456	0.384	-0.072	-15.8

TABLE 6.3

Simulation Results (Coverage) for *Diporeia hoyi* (Water Distance and Classical Variogram Estimator)

m = 1000 Realisations

Parameter	S.E.(Asy)	S.E.(Sim)	LowerC.I.	UpperC.I.	Coverage	W	Pr<W
$\hat{\beta}_0$	4.02	4.17	-10.33	5.43	0.95	0.984	0.092
$\hat{\beta}_1$	0.17	0.17	0.42	1.09	0.95	0.984	0.128
$\hat{\beta}_2$	0.67	0.66	-2.72	-0.10	0.95	0.989	0.877
$\hat{\beta}_3$	0.68	0.70	0.42	3.08	0.94	0.986	0.407
$\hat{\beta}_4$	0.92	0.96	-0.97	2.63	0.95	0.985	0.222

TABLE 6.4

Simulation Results (Coverage) for *Diporeia hoyi* (Independent Errors)

m = 1000 Realisations

Parameter	S.E.(Asy)	S.E.(Sim)	LowerC.I.	UpperC.I.	Coverage	W	Pr<W
$\hat{\beta}_0$	4.02	3.69	-10.33	5.43	0.97	0.983	0.045
$\hat{\beta}_1$	0.17	0.16	0.42	1.09	0.95	0.986	0.455
$\hat{\beta}_2$	0.67	0.57	-2.72	-0.10	0.98	0.989	0.818
$\hat{\beta}_3$	0.68	0.61	0.42	3.08	0.97	0.986	0.442
$\hat{\beta}_4$	0.92	0.85	-0.97	2.63	0.97	0.985	0.186

semivariogram as used in Chapters 4 and 5. The results of 1000 replications were summarized for the variance-covariance matrix of $\hat{\beta}$, and the mean, standard error and skewness of $\hat{\sigma}^2$ and $\hat{\theta}$.

Harris and Johnson report very little bias in their estimates of $\hat{\beta}$ when the exponential variogram was used to account for spatial dependence. However, their estimates of $\hat{\sigma}^2$ were found to have substantial negative bias. Similarly, their $\hat{\theta}$, which is the inverse of $\hat{\theta}_3$ in this work, had a strong negative bias when $\hat{\beta}$ included a trend parameter. This corresponds to the large positive bias in $\hat{\theta}_3$ in this study. Harris and Johnson point out that in some cases the biases in $\hat{\sigma}^2$ and $\hat{\theta}$ counteract each other in their effects on the confidence intervals of $\hat{\beta}$.

Table 6.5 gives the characteristics of the simulated variogram parameters ($\hat{\sigma}^2$, $\hat{\theta}_3$ in the dependent case) and method of moments estimate ($\hat{\sigma}^2$ in the independent case). The distributions of the sill and the range are not normal and are quite skewed, especially the sill. Upper and lower 95% confidence levels were determined empirically from the simulation results. The simulated values of the sill and range had a Pearson correlation coefficient of -0.0978 which was significantly different from zero (p=0.002).

TABLE 6.5

Simulation Results for Variogram Parameters for *Diporeia hoyi*

m = 1000 Realisations

	$\hat{\sigma}^2$(Indep.)	$\hat{\sigma}^2$(Depen.)	$\hat{\theta}_3$
True Value	0.456	0.456	0.1271
Maximum	2.327	8.576	1.8038
75% (Q3)	0.456	0.404	0.4301
50% (Med.)	0.326	0.272	0.2334
25% (Q1)	0.233	0.193	0.1373
Minimum	0.069	0.031	0.0094
Lower C.I.	0.137	0.096	0.0556
Upper C.I.	0.931	0.877	1.0819
Skewness	2.904	12.78	1.704
Pearson Corr. Coef.		-0.0978	
p value		0.002	

The method of Cherry (1997) is available to reduce the bias in the sill. This would probably have a beneficial effect on the bias of the range parameter also, since these parameters often compensate for one another (Harris and Johnson, 1996; Ecker and Gelfand, 1997). If the goal of the study is to examine the characteristics of the spatial dependence, the effort to reduce or eliminate the bias in variogram parameters is justified. In this work, where the goal is spatial prediction, the improvement in predictive capability of the model may not be enough to warrant the extra effort. For the case of universal kriging, unbiased variogram parameters are not needed to obtain unbiased predictors (Cressie, 1993, p. 295).

In both of the simulation cases in this study, the results of the estimation of $\hat{\beta}$ using the quasi-likelihood model where consistent with Theorem 2. The standard errors were obtained from the diagonals of the variance-covariance matrices and used to construct the confidence intervals in Tables 6.3 and 6.4. The coverage proportions of the nominal 95% confidence intervals are close to the nominal value for the dependent case (Table 6.3) and in Table 6.4 four of the proportions exceed the nominal value (independent case). The attained significance level of the Shapiro-Wilk test for normality was greater than 0.05 in all but one case (for $\hat{\beta}_0$ in Table 6.4). Visual examination of histograms and normal probability plots confirmed this finding. The effect of spatially

dependent errors is to allow more realistic estimation of standard errors and approximately correct coverage proportions for interval estimates. For the independent case, the standard errors are underestimated and the coverage proportions inflated similar to Zeger's findings (1988) regarding the effect of incorporation of time dependence on confidence intervals for parameter estimates in his model. These results are also consistent with the results of simulations for spatial lattice data discussed in Section 2.2.6.

BIBLIOGRAPHY

Abraham, B. and J. Ledolter (1983), *Statistical Methods for Forecasting*. Wiley, New York.

Agterberg, F. P. (1970), "Autocorrelation functions in geology". In: *Geostatistics - a Colloquium*, D. F. Merriam (ed.), 113-141. Plenum Press, New York.

Aitchison, J. and J. A. C. Brown (1963), *The Lognormal Distribution*. University Press, Cambridge.

Anderson, T. W. (1958), *Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Armitage, P. D., D. Moss, J. F. Wright and M. T. Furse (1983), "The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites". *Water Research* 17(3), 333-347.

Armitage, P. D., R. J. M. Gunn, M. T. Furse, J. F. Wright and D. Moss (1987), "The use of prediction to assess macroinvertebrate response to river regulation". *Hydrobiologia* 144, 25-32.

Armstrong, M. (1984), "Improving the estimation and modelling of the variogram". In: *Geostatistics for Natural Resources Characterization Part I*, G. Verly, M. David, A. Journel, and A. Marechal (eds.), 1-19. D. Reidel Publishing, Dordrecht.

Bailey, R. C., K. E. Day, R. H. Norris and T. B. Reynoldson (1995), "Macroinvertebrate community structure and sediment bioassay results from nearshore areas of North American Great Lakes". *Journal of Great Lakes Research* 21(1), 42-52.

Banerjee, B. (1976), "Variance to mean ratio and the spatial distribution of animals". *Experientia* 32(8), 993-994.

Barton, D. R. (1986), "Nearshore benthic invertebrates of the Ontario waters of Lake Ontario". *Journal of Great Lakes Research* 12(4), 270-280.

Barton, D. R. (1988), "Distribution of some common benthic invertebrates nearshore Lake Erie, with emphasis on depth and type of substratum". *Journal of Great Lakes Research* 14(1), 34-43.

Barton, D. R. and M. Griffiths (1984), "Benthic invertebrates of the nearshore zone of eastern Lake Huron, Georgian Bay and North Channel". *Journal of Great Lakes Research* 10(4), 407-416.

Bickel, P. J. and K. A. Doksum (1977), *Mathematical Statistics.* Holden-Day, San Francisco.

Billingsley, P. (1968), *Convergence of Probability Measures.* Wiley, New York.

Bousfield, E. L. (1989), "Revised morphological relationships within the amphipod genera *Pontoporeia* and *Gammaracanthus* and the 'glacial relict' significance of their postglacial distributions". *Canadian Journal of Fisheries and Aquatic Sciences* 46, 1714-1725.

Box, G. E. P. and G. M. Jenkins (1970), *Time Series Analysis Forecasting and Control.* Holden-Day, San Francisco.

Brus, D. J. and J. J. de Gruijter (1993), "Design-based versus model-based estimates of spatial means: theory and application in environmental soil science". *Environmetrics* 4(2), 123-152.

Carroll, S. S. and N. Cressie (1996), "A comparison of geostatistical methodologies used to estimate snow water equivalent". *Water Resources Bulletin* 32(2), 267-278.

Cherry, S. (1997), "Non-parametric estimation of the sill in geostatistics". *Environmetrics* 8(1), 13-28.

Christensen, R. (1991), *Linear Models for Multivariate, Time Series and Spatial Data.* Springer-Verlag, New York.

Cole, R. A. and D. L. Weigmann (1983), "Relationships among zoobenthos, sediments, and organic matter in littoral zones of western Lake Erie and Saginaw Bay". *Journal of Great Lakes Research* 9(4), 568-581.

Corkum, L. D. and D. C. Currie (1987), "Distributional patterns of immature Simuliidae (Diptera) in northwestern North America". *Freshwater Biology* 17, 201-227.

Corkum, L. D. (1989), "Patterns of benthic invertebrate assemblages in rivers of northwestern North America". *Freshwater Biology* 21, 191-205.

Cox, D. R. (1981), "Statistical analysis of time series: some recent developments". *Scandinavian Journal of Statistics* 8, 93-115.

Cressie, N. A. C. (1985), "Fitting variogram models by weighted least squares". *Mathematical Geology* 17(5), 563-586.

——— (1993), *Statistics for Spatial Data*, Revised Edition. Wiley, New York.

Cui, H., A. Stein and D. E. Myers (1995), "Extention of spatial information, Bayesian kriging and updating of prior variogram parameters". *Environmetrics* 6(4), 373-384.

Delfiner, P. and J. P. Delhomme (1975), "Optimum interpolation by kriging". *Display and Analysis of Spatial Data*. J. C. Davis and M. J. McCullagh (eds.), 96-114. Wiley, New York.

Diggle, P. J., K.-Y. Liang, and S. L. Zeger (1994), *Analysis of Longitudinal Data*. Oxford University Press.

Draper, N. R. and H. Smith (1981), *Applied Regression Analysis*, Second Edition. Wiley, New York.

Ecker, M. D. and A. E. Gelfand (1997), "Bayesian variogram modeling for an isotropic spatial process". *Journal of Agricultural, Biological, and Environmental Statistics* 2(4), 347-369.

Elliott, J. R. and A. H. El-Shaarawi (1995), "Sediment lead loads in the Niagara River, 1986-1992". *Journal of Great Lakes Research* 21(3), 324-332.

Foutz, R. V. (1977), "On the unique consistent solution to the likelihood equations". *Journal of the American Statistical Association* 72(357), 147-148.

Fratt, T. W., D. W. Coble, F. Copes and R. E. Bruesewitz (1997), "Diet of burbot in Green Bay and western Lake Michigan with comparison to other waters". *Journal of Great Lakes Research* 23(1), 1-10.

Gallant, A. R. and H. White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell Ltd., Oxford.

Geisser, S. (1975), "The predictive sample reuse method with applications". *Journal of the American Statistical Association* 70, 320-328.

Goldie, C. M. and P. E. Greenwood (1986a), "Characterisations of set-indexed Brownian motion and associated conditions for finite-dimensional convergence". *The Annals of Probability*, 14, 802-816.

——— (1986b), "Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes". *The Annals of Probability*, 14, 817-839.

Gossiaux, D. C., P. F. Landrum and V. N. Tsymbal (1992), "Response of the amphipod *Diporeia* spp. to various stressors: cadmium, salinity, and temperature". *Journal of Great Lakes Research* 18(3), 364-371.

Gotway, C. A. and A. H. Hartford (1996), "Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables". *Journal of Agricultural, Biological, and Environmental Statistics* 1(1), 17-39.

Gotway, C. A. and W. W. Stroup (1997), "A generalized linear model approach to spatial data analysis and prediction". *Journal of Agricultural, Biological, and Environmental Statistics* 2(2), 157-178.

Guttorp, P. and Sampson, P. D. (1994), "Methods for estimating heterogeneous spatial covariance functions with environmental applications". In: *Handbook of Statistics, Vol. 12*, G. P. Patil and C. R. Rao (eds.), 661-689. Elsevier Science, North-Holland.

Haining, R. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press.

Harris, T. R. and D. E. Johnson (1996), "A regression model with spatially correlated errors for comparing remote sensing and in-situ measurements of a grassland site". *Journal of Agricultural, Biological, and Environmental Statistics* 1(2), 190-204.

Hartley, H. O. (1961), "The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares". *Technometrics* 3, 269-280.

Harvey, A. C. (1981), *The Econometric Analysis of Time Series*. Wiley, New York.

Hawkins, D. M. and N. Cressie (1984), "Robust kriging - a proposal". *Mathematical Geology*, 16(1), 3-18.

International Joint Commission (1987), *The Great Lakes Water Quality Agreement of 1978 (As Amended by Protocol, 1987)*. International Joint Commission, Washington, D.C. and Ottawa, Ontario. Reprinted 1994.

Jennrich, R. I. (1969), "Asymptotic properties of non-linear least squares estimators". *The Annals of Mathematical Statistics* 40(2), 633-643.

Johnson, M. P., L. G. Mason and P. H. Raven (1968), "Ecological parameters and plant species diversity". *The American Naturalist* 102(926), 297-306.

Johnson, R. K. and T. Wiederholm (1989), "Classification and ordination of profundal macroinvertebrate communities in nutrient poor, oligo-mesohumic lakes in relation to environmental data". *Freshwater Biology* 21, 375-386.

Journel, A. G. and Ch. J. Huijbregts (1978), *Mining Geostatistics*. Academic Press, London.

Kaplan, W. (1973), *Advanced Calculus*, Second Edition. Addison-Wesley, Reading, Massachusetts.

Kaufmann, H. (1987), "Regression models for nonstationary categorical time series: asymptotic estimation theory". *The Annals of Statistics*, 15(1), 79-98.

King, J. M. (1981), "The distribution of invertebrate communities in a small South African river". *Hydrobiologia* 83, 43-65.

Kitanidis, P. K. (1983), "Statistical estimation of polynomial generalized covariance functions and hydrologic applications". *Water Resources Research*, 19(4), 909-921.

Kraft, K. J. (1979), "*Pontoporeia* distribution along the Keweenaw shore of Lake Superior affected by copper tailings". *Journal of Great Lakes Research* 5(1), 28-35.

Kraft, K. J. and R. H. Sypniewski (1981), "Effect of sediment copper on the distribution of benthic macroinvertebrates in the Keweenaw Waterway". *Journal of Great Lakes Research* 7(3), 258-263.

Laslett, G. M. (1994), "Kriging and splines: an empirical comparison of their predictive performance in some applications". *Journal of the American Statistical Association* 89(426), 391-400.

Lehmann, E. L. (1991), *Theory of Point Estimation*. Wadsworth, Inc., Belmont, California.

Liang, K.-Y. and S. L. Zeger (1986), "Longitudinal data analysis using generalized linear models". *Biometrika*, 73(1), 13-22.

Mardia, K. V. and R. J. Marshall (1984), "Maximum likelihood estimation of models for residual covariance in spatial regression". *Biometrika*, 71(1), 135-146.

Matheron, G. (1962), "Traite de geostatistique appliquee, Tome I". *Memoires du Bureau de Recherches Geologiques et Minieres* No.14. Editions Technip, Paris.

——— (1963), "Principles of Geostatistics". *Economic Geology* 58, 246-266.

——— (1967), "Kriging or polynomial interpolation procedures?". *Transactions of the Canadian Institute of Mining and Metallurgy* 70, 240-244.

——— (1971), "The theory of regionalized variables and its applications". *Cahiers du Centre de Morphologie Mathematique* No. 5. Fontainebleau, France.

McCullagh, P. (1983), "Quasi-likelihood functions". *The Annals of Statistics* 11(1), 59-67.

McCullagh, P. and J. A. Nelder (1983), *Generalized Linear Models*. Chapman and Hall, London.

McShane, L. M., P. S. Albert and M. A. Palmatier (1997), "A latent process regression model for spatially correlated count data". *Biometrics* 53, 698-706.

Millard S. P., J. R. Yearsley and D. P. Lettenmaier (1985), "Space-time correlation and its effect on methods for detecting aquatic ecological change". *Canadian Journal of Fisheries and Aquatic Sciences* 42, 1391-1400.

Monestiez, P., J. Audergon and R. Habib (1990), "Spatial dependences and sampling in a fruit tree: a geostatistical approach". *SIMS Technical Report* No. 163. Stanford University.

Moore, D. F. (1986), "Asymptotic properties of moment estimators for overdispersed counts and proportions". *Biometrika* 73(3), 583-588.

Morrison, D. F. (1976), *Multivariate Statistical Methods*. McGraw-Hill, New York.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage (1987), "The prediction of the macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data". *Freshwater Biology* 17, 41-52.

Murphy, G. M. (1960), *Ordinary Differential Equations and their Solutions*. Van Nostrand, Princeton, New Jersey.

Nalepa, T. F. (1989), "Estimates of macroinvertebrate biomass in Lake Michigan". *Journal of Great Lakes Research* 15(3), 437-443.

Nalepa, T. F. and M. A. Quigley (1983), "Abundance and biomass of the meiobenthos in nearshore Lake Michigan with comparisons to the macrobenthos". *Journal of Great Lakes Research* 9(4), 530-547.

Owens, R. W. and P. G. Weber (1995), "Predation on *Mysis relicta* by slimy sculpins (*Cottus cognatus*) in southern Lake Ontario". *Journal of Great Lakes Research* 21(2), 275-283.

Parker, J. I. (1980), "Predation by *Mysis relicta* on *Pontoporeia hoyi*: a food chain link of potential importance in the Great Lakes". *Journal of Great Lakes Research* 6(2), 164-166.

Poole, R. W. (1978), "The statistical prediction of population fluctuations". *Annual Review of Ecology and Systematics* 9, 427-448.

Poulton, D. J., K. J. Simpson, D.R. Barton and K. R. Lum (1988), "Trace Metals and benthic invertebrates in Sediments of nearshore Lake Ontario at Hamilton Harbour". *Journal of Great Lakes Research* 14(1), 52-65.

Prentice, R. L. (1988), "Correlated binary regression with covariates specific to each binary observation". *Biometrics* 44, 1033-1048.

Quigley, M. A. (1988), "Gut fullness of the deposit-feeding amphipod, *Pontoporeia hoyi*, in southeastern Lake Michigan". *Journal of Great Lakes Research* 14(2), 178-187.

Rathbun, S. L. (1998), "'Spatial modelling in irregularly shaped regions: kriging estuaries". *Environmetrics* 9(2), 109-129.

Reynoldson, T. B., R. C. Bailey, K. E. Day and R. H. Norris (1995), "Biological guidelines for freshwater sediment based on benthic assessment of sediment using a multivariate approach for predicting biological state". *Australian Journal of Ecology* 20, 198-219.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg (1997), "The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates". *Journal of the North American Benthic Society* 16(4), 833-852.

Ripley, B. D. (1981), *Spatial Statistics*. Wiley, New York.

Sampson, P. D. and P. Guttorp (1992), "Nonparametric estimation of nonstationary spatial covariance structure". *Journal of the American Statistical Association* 87(417), 108-119.

SAS Institute Inc., (1982), *SAS User's Guide: Statistics*. Cary, North Carolina.

SAS Institute Inc., (1990), *SAS/IML Software: Usage and Reference*. Version 6. Cary, North Carolina.

Serfling, R. J. (1970), "Convergence proprties of S_n under moment restrictions". *The Annals of Mathematical Statistics* 41(4), 1235-1248.

——— (1980), *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shapiro, A. and J. D. Botha (1991), "Variogram fitting with a general class of conditionally nonnegative definite functions". *Computational Statistics and Data Analysis* 11, 87-96.

Shapiro, S. S. and M. B. Wilk (1965), "An analysis of variance test for normality (complete samples)". *Biometrika* 52, 591-611.

Shrivistava, H. N. (1974), "Macrobenthos of Lake Huron". *Journal of the Fisheries Research Board of Canada*, Technical Report 449.

Stone, M. (1974), "Cross-validatory choice and assessment of statistical predictions". *Journal of the Royal Statistical Society B*, 36, 111-133.

Switzer, P. (1984), "Inference for spatial autocorrelation functions". In: *Geostatistics for Natural Resources Characterization Part I*, G. Verly, M. David, A. Journel, and A. Marechal (eds.), 127-140. D. Reidel Publishing, Dordrecht.

Taylor, L. R. (1961), "Aggregation, variance, and the mean". *Nature, London* 189, 732-735.

Taylor, L. R. and R. A. J. Taylor (1977), "Aggregation, migration, and population mechanics". *Nature* 265, 415-420.

Thall, P. F. and S. C. Vail (1990), "Some covariance models for longitudinal count data with overdispersion". Biometrics 46, 657-671.

Upton, G. J. G. and B. Fingleton (1985), *Spatial Data Analysis by Example.* Vol. 1: "*Point pattern and quantitative data*". Wiley, New York.

Wedderburn, R. W. M. (1974), "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method". *Biometrika* 61(3), 439-447.

Whittle, P. (1954), "On stationary processes in the plane". *Biometrika* 41, 434-449.

Woodroofe. M. (1975). *Probability with Applications.* McGraw-Hill, New York.

Wright, J. F., D. Moss, P. D. Armitage, and M. T. Furse (1984), "A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data". *Freshwater Biology* 14, 221-256.

Yasui, Y. and S. Lele (1997), "A regression method for spatial disease rates: an estimating function approach". *Journal of the American Statistical Association* 92, 21-32.

Zacks, S. (1971), *The Theory of Statistical Inference.* Wiley, New York.

Zarull, M. A. and T. B. Reynoldson (1992), "A management strategy for contaminated sediment: assessment and remediation". *Water Pollution Research Journal of Canada* 27, 871-882.

Zeger, S. L. and K-Y. Liang (1986), "Longitudinal data analysis for discrete and continuous outcomes". *Biometrics* 42, 121-130.

Zeger, S. L. and B. Qaqish (1988), "Markov regression models for time series: a quasi-likelihood approach". *Biometrics* 44, 1019-1031.

GLOSSARY

Autocorrelation - Correlation exhibited by neighboring values of a time series or spatial process.

Cholesky Decomposition - A procedure for finding a matrix $\Sigma^{1/2}$, such that $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})'$

Cluster Analysis - A procedure for deciding empirically whether individual observations "group" or "cluster" based on some concept of nearness or distance.

Expectation, or expected value, of X - Defined as:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx$$

where $f(x)$ is the density of x.

Gauss-Markov Theorem - If a regression model satifies the classical assumptions, the ordinary least squares estimator of the regression parameters, β, is the best linear unbiased estimator (BLUE).

Gauss-Newton Method - A computational procedure for estimating the regression parameters, β, of a non-linear model using a Taylor series expansion. A starting value is chosen for β and it is iteratively improved until the error sum of squares is minimized.

Geographic Information System (GIS) - A computerized system for mapping and analyzing spatial information.

Global Positioning System (GPS) - A system of satellites that provides accurate information on geographical location to ground-based receivers. These receivers are accessed by means of a hand-held instrument which can be used to locate a sampling site to an accuracy of within a few metres.

Inadmissible - Define the mean squared error of a parameter estimate as $R(\theta,T)$ where θ is the parameter and T is the estimate. If two estimators S and T are such that $R(\theta,T) \leq R(\theta,S)$ for all θ with strict inequality holding for some θ, then S is inadmissible. (Bickel and Doksum, 1977, p. 118)

Kriging - A minimum-mean-squared-error method of spatial prediction that (usually) depends on the second-order properties of the process. Named after D.G. Krige, a South African mining engineer who, in the 1950s, developed

empirical methods for determining true ore-grade distributions from distributions based on sampled ore grades. (Cressie, 1993, p. 106)

Kurtosis - The degree of peakedness of a distribution, taken relative to the normal distribution. It is usually expressed as the ratio of the fourth moment about the mean to the square of the second moment about the mean, minus 3:

$$\mu_4 / \mu_2^2 - 3$$

Lagrange Multipliers - A scalar or vector value that is introduced to add one or more constraints to the solution of the extrema for a quadratic form.

Marginal Density - If X and Y have joint density *f*, then X and Y have marginal densities given by:

$$g(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy \qquad \text{for } -\infty < x < +\infty \qquad \text{and,}$$

$$h(x) = \int_{-\infty}^{+\infty} f(x,y)\, dx \qquad \text{for } -\infty < y < +\infty .$$

Marginal Moment - Moments calculated using the marginal densities defined above.

Markov Process - A process in which the future state depends on the present state and not on past states given the present.

$\sqrt{n}$ - Consistent Estimate - An estimate, $\hat{\theta}_n$, of a vector of parameters, θ, such that $\sqrt{n}(\hat{\theta}_n - \theta)$ is bounded in probability so that $\hat{\theta}_n$ tends to θ at least at the rate of $1/\sqrt{n}$ (Lehmann, 1991, p. 422).

Ordinary Least Squares (OLS) - The estimator of regression parameters when the variance matrix satisfies classical assumptions, ie. independent, identically distributed errors.

Skewness - The degree of asymmetry of a distribution. It is usually expressed as the ratio of the third moment about the mean to the 3/2 root of the second moment about the mean:

$$\mu_3/\mu_2^{3/2}$$

Uniformly Integrable (Serfling, 1980, p.13) - A sequence of random variables $\{ Z_n \}$ is uniformly integrable if:

$$\lim_{c\to\infty} \sup_n E\{|Z_n|I(|Z_n|>c)\} = 0$$

where $E\{\cdot\}$ is the expectation

and $I(\cdot)$ is the indicator function.

Variogram - The mean-squared difference between the values of a characteristic at two points as a function of the distance between the points.