AN ABSTRACT OF THE THESIS OF

Caryn Marie Thompson for the degree of Doctor of Philosophy in Statistics presented on June 6, 1995 . Title : Diagnostics for the Evaluation of Spatial Linear Models.

Redacted for Privacy

Abstract approved: _____

Fred L. Ramsey

Geostatistical linear interpolation procedures such as kriging require knowledge of the covariance structure of the spatial process under investigation. In practice, the covariance of the process is unknown, and must be estimated from the available data. As the quality of the resulting predictions, and associated mean square prediction errors, depends on adequate specification of the covariance structure, it is important that the analyst be able to detect inadequacies in the specified covariance model. Case-deletion diagnostics are currently used by geostatisticians to evaluate spatial models.

The second chapter of the thesis describes a particular case-deletion diagnostic based on standardized PRESS residuals, and its use in assessing the predictive capacity of spatial covariance models. Distributional properties of this statistic, denoted T_{PR}, are discussed, and a saddlepoint approximation to its distribution is derived. Guidelines for calculating approximate p-values for the statistic under an hypothesized covariance model are also given. A simulation study demonstrates that the distributional and p-value approximations are accurate. The proposed method is illustrated through an example, and recommendations for calculation of T_{PR}, and associated approximate p-values on a regional basis are given.

The third chapter investigates the behavior of the standardized PRESS residuals under various misspecifications of the covariance matrix, $\boldsymbol{V}$. A series of simulation studies show consistent patterns in the standardized PRESS residuals under particular types of misspecifications of $\boldsymbol{V}$. It is observed that misspecification of $\boldsymbol{V}$ may lead to variability among

the standardized PRESS residuals greater or less than would be expected if $\boldsymbol{V}$ was correctly specified, depending on the nature of the misspecification. Based on this observation, an adjustment to normal probability plots of the standardized PRESS residuals is proposed. The adjusted normal probability plots may be used to identify potential improvements to covariance models, without requiring extensive further calculations.

Diagnostics for the Evaluation of Spatial Linear Models

by

Caryn Marie Thompson

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed June 6, 1995

Commencement June 1996

Doctor of Philosophy thesis of Caryn Marie Thompson presented on June 6, 1995

APPROVED :

Redacted for Privacy

Major Professor, representing Statistics

Redacted for Privacy

Chair of Department of Statistics

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Caryn Marie Thompson, Author

Dedication

This thesis is dedicated to my parents, Maurice C. and E. Jean Thompson, and to the loving memory of my aunt, Loretta Marie Fulton (1922 – 1995) :

"Her finely touched spirit had still its fine issues, though they were not widely visible. Her full nature, like that river of which Cyrus broke the strength, spent itself in channels which had no great name on the earth. But the effect of her being on those around her was incalculably diffusive: for the growing good of the world is partly dependent on unhistoric acts; and that things are not so ill with you and me as they might have been, is half owing to the number who lived faithfully a hidden life, and rest in unvisited tombs."

– from *Middlemarch*
by George Eliot

Acknowledgements

I wish to acknowledge several people who have contributed to the completion of this thesis. Firstly, I thank Dr. Fred L. Ramsey for serving as my supervisor, and for his effort in scheduling the thesis defense. Thanks also go to the members of my committee, Dr. David Birkes, Dr. Dudley Chelton, Dr. Virginia Lesser, and Dr. Scott Urquhart for their thoughtful comments and recommendations for improvements to the thesis.

I thank Dr. Noel Cressie of Iowa State University for suggesting the problem addressed in the second chapter of the thesis. Dr. Qing Liu made several recommendations regarding the saddlepoint approximation, and also supplied some MATLAB functions. Dr. David Birkes supplied a GAUSS program for calculation of REML estimates for the parameters of universal kriging models.

Prof. Bryan Manly of the Department of Mathematics and Statistics of the University of Otago kindly read an earlier draft of the thesis, and made useful suggestions. Special thanks go to Genevieve Downing of Oregon State University; without her help, it would have been impossible to meet the various deadlines for submission of the thesis.

I am grateful to Dr. Justus Seely and the Department of Statistics of Oregon State University for providing funding and support during my graduate studies, and to the University of Otago for granting me leave to return to Oregon State during the summer of 94/95 (NZ).

I acknowledge my friends from the "good old days" in Corvallis and my new friends in Dunedin for their companionship and encouragement. Lastly, I wish to express my sincere gratitude to my parents, Maurice and Jean Thompson, and to my aunt, Marie Fulton, for their unfailing belief in me, and for the emotional and financial support they have provided over the many years of my education. I dedicate this thesis to them.

> "I was pushed back and about to fall, but the Lord helped me.
> The Lord is my strength and my song; he has become my
> salvation".
>
> Psalm 118 : 13 – 14

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

DIAGNOSTICS FOR THE EVALUATION OF SPATIAL LINEAR MODELS

INTRODUCTION

Most statistical procedures and models are based on certain assumptions. In order to have confidence in the analysis achieved using a particular model, the validity of the associated assumptions must be verified. Typically, verification consists of the informal examination of diagnostic statistics or plots.

Much effort has been devoted to the development of diagnostic procedures in many areas of statistics. Consider regression analysis, where a plethora of diagnostic tools are available to the analyst, with which the various assumptions underlying the multiple regression model may be verified, and candidate models evaluated. The criteria used depend on the objectives of the particular analysis. For instance, the Prediction Sum of Squares (PRESS) statistic (Allen, 1971) is used to evaluate models from a prediction standpoint. Mallow's C_p statistic identifies models which provide a balance between bias and variance in the regression parameter estimates and predictions. The coefficient of determination, R^2 (or adjusted R^2), provides a measure of how well the model fits the available data, by indicating how much of the total variance in the observations is explained by the model. Cook and Weisberg (1982) provide a comprehensive discussion of diagnostic techniques and statistics based on the analysis of residuals, where the primary purpose is to check for violations in the assumptions of the model. Other diagnostics are available to detect influential observations, and identify sources of collinearity in the data (Belsley *et al.*, 1980).

The goal of the research recorded in this thesis was to improve existing diagnostics for the evaluation of spatial linear models in particular, and to investigate techniques to highlight various inadequacies in the assumed covariance structure of the model. The research focused on the use of standardized PRESS residuals to achieve this end. While certain of the

results presented might find equal application in the regression setting, the primary objective was to develop diagnostics appropriate for use in spatial data analysis.

Spatial Linear Models – General

Spatial autocorrelation is a phenomenon associated with many geographically distributed random variables. Typically, these variables exhibit positive autocorrelation, meaning in practical terms that observations close together tend to be more similar in value than observations far apart. Due to the lack of spatial independence, classical statistical methods developed for analysis of variables assumed to be independently and identically distributed are not appropriate.

Spatial models incorporate the phenomenon of spatial autocorrelation into their error structure. These models have found widespread application because of the ubiquity of spatial autocorrelation. Spatial models are an essential part of the analysis of data arising in many subject areas, including geology, soil science, agronomy, forestry, oceanography, epidemiology, and image processing.

One goal of the analysis of spatially referenced data is to develop models which adequately describe global and local variation within some region of interest, and provide a means of predicting values of one or more random variables at unmeasured locations. This thesis concentrates on geostatistics as a subset of the broader category of spatial statistics, and focuses on the geostatistical linear interpolation procedure known as kriging. Typically, in the analysis of spatial data, a "superpopulation" model-based approach to inference is taken, whereby the data, recorded at spatial locations, are considered to be a partial realization of a random process.

In the geostatistical setting, the models selected to describe spatial variability are of a particular parametric form, and must satisfy certain stationarity conditions. Under these

assumptions, the available data may be used to obtain best linear unbiased predictors (BLUP's) of the random variable at any location within some region of interest. Functions such as the *covariance function* or *variogram* describe the spatial structure within this region.

Suppose measurements corresponding to a random variable, Y, have been obtained at n spatial locations $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, where $\boldsymbol{u}_i$ is a vector containing the coordinates for the i^{th} location. Let D define a specific geographical region. The BLUP of $y(\boldsymbol{u}_0)$, the value of y at any location $\boldsymbol{u}_0 \in D$, is obtained by taking a weighted linear combination of the observations $\boldsymbol{Y} = [y(\boldsymbol{u}_1), \ldots, y(\boldsymbol{u}_n)]'$. The weight assigned to each $y(\boldsymbol{u}_i)$ in predicting $y(\boldsymbol{u}_0)$, depends on the spatial structure, as defined by the covariance function or variogram.

A process is said to be second-order stationary if the covariance

$$\mathrm{Cov}[y(\boldsymbol{u}_i),\, y(\boldsymbol{u}_j)] = C(\boldsymbol{u}_i,\boldsymbol{u}_j)$$

depends only on the relative locations of $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$. In this situation, the variogram is defined by

$$2\gamma(\boldsymbol{u}_i,\boldsymbol{u}_j) = \mathrm{Var}[y(\boldsymbol{u}_i) - y(\boldsymbol{u}_j)].$$

The variogram function $2\gamma(\cdot)$ is also defined under the weaker assumption of intrinsic stationarity (Cressie, 1991), whereby

$$2\gamma(\boldsymbol{u}_i,\boldsymbol{u}_j) = \mathrm{E}[y(\boldsymbol{u}_i) - y(\boldsymbol{u}_j)]^2 \,.$$

The ordinary kriging model is

$$\boldsymbol{Y} = \mathbf{1}\beta + \boldsymbol{\epsilon} \,,$$

where $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{V}$, with elements $v_{ij} = C(\boldsymbol{u}_i,\, \boldsymbol{u}_j)$, $\mathbf{1}$ is an $n \times 1$ vector of 1's, and $\beta = \mathrm{E}[Y(\boldsymbol{u})]$, for all $\boldsymbol{u} \in D$. In terms of the covariance function, the BLUP of $y(\boldsymbol{u}_0)$ under this

model is given by

$$\hat{y}(\boldsymbol{u}_0) = \boldsymbol{\lambda}'\boldsymbol{Y} ,$$

with vector of weights

$$\boldsymbol{\lambda} = \left[\frac{(1 - \boldsymbol{V}_0'\boldsymbol{V}^{-1}1)}{1'\boldsymbol{V}^{-1}1} 1' + \boldsymbol{V}_0' \right] \boldsymbol{V}^{-1},$$

where $\boldsymbol{V}_0' = [C(\boldsymbol{u}_0,\boldsymbol{u}_1), \ldots, C(\boldsymbol{u}_0,\boldsymbol{u}_n)]$ (Christensen, 1990). A brief description of the more general universal kriging model is given in Chapter 2 of this thesis.

Strictly speaking, for the kriging predictor to be a BLUP, the covariance matrix $\boldsymbol{V}$, and therefore either the covariance function or variogram, must be known. In reality these functions are unknown, and so the usual practice is to estimate their values at specific lags, $\boldsymbol{l}$. Under the assumption of *isotropy*, $\boldsymbol{l}$ is usually defined to be the Euclidean distance between $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$. The two most common estimators of the variogram function are a method-of-moments estimator (Matheron, 1963), and a robust estimator proposed by Cressie and Hawkins (1980). Typically, a parametric form of the variogram or covariance model is then fit to these estimates, using a procedure such as weighted least squares, or restricted maximum likelihood (Cressie, 1991; Zimmerman and Zimmerman, 1991). Popular variogram models include the exponential, spherical, Gaussian, and power models (Cressie, 1991). Each of these models has particular features, related to its parameters, which correspond to properties of the surface defined by the spatial process.

Two of the models in particular, the exponential and Gaussian models, are considered at various points throughout this thesis. The exponential model is given by

$$C(\boldsymbol{l}) = \begin{cases} \theta_1 exp(-\theta_2 \boldsymbol{l}) & \text{for } \boldsymbol{l} > 0 \\ \theta_0 + \theta_1 & \text{for } \boldsymbol{l} = 0 \end{cases},$$

while the Gaussian model has the form

$$C(l) = \begin{cases} \theta_1 exp(-(\theta_2 l)^2) & \text{for } l > 0 \\ \theta_0 + \theta_1 & \text{for } l = 0 \end{cases} .$$

The spherical model is also involved in one of the simulation studies of Chapter 3.

It is beneficial to have an intuitive understanding of the parameters of the covariance function or variogram models. Isaaks and Srivastava (1989) and Cressie (1991) give interpretations of each of these parameters. For convenience, the discussion here will be framed in terms of the variogram function, its features and properties. Similar principles apply to the covariance function. When the condition of second-order stationarity is satisfied, a simple relationship exists between the variogram and the covariance function, namely $2\gamma(l) = 2(C(0) - C(l))$.

Important features of the variogram include the sill, nugget effect and range. Figure 1.1 shows a generic variogram, and highlights each of these features. The sill of the theoretical variogram, the sum of the parameters θ_0 and θ_1 in the models given above, corresponds to the variance of the process. The range is the separation lag at which the sill is reached, beyond which two locations are considered to be spatially uncorrelated. The exponential and Gaussian models approach a sill only asymptotically, and therefore the *effective range* is defined to be the lag at which a value equal to 95% of the sill is reached (Journel and Huijbregts, 1978). For the exponential model, the effective range is approximately equal to $\frac{3}{\theta_2}$, while for the Gaussian model, the effective range is approximately $\frac{\sqrt{3}}{\theta_2}$. The nugget effect, θ_0, refers to microscale variability, or variability on a scale smaller than that which can be identified using the available data. The nugget effect consists of two components, one of which is measurement error .

Another important consideration when modeling spatial variability is that of anisotropy. In practice, the model assumption of isotropy often does not hold, meaning that spatial

structure depends on direction as well as distance. Models are available which specify varying degrees of spatial correlation in several directions. A particular type of anisotropy, geometric anisotropy, may be corrected by a linear transformation of the lag vector l.

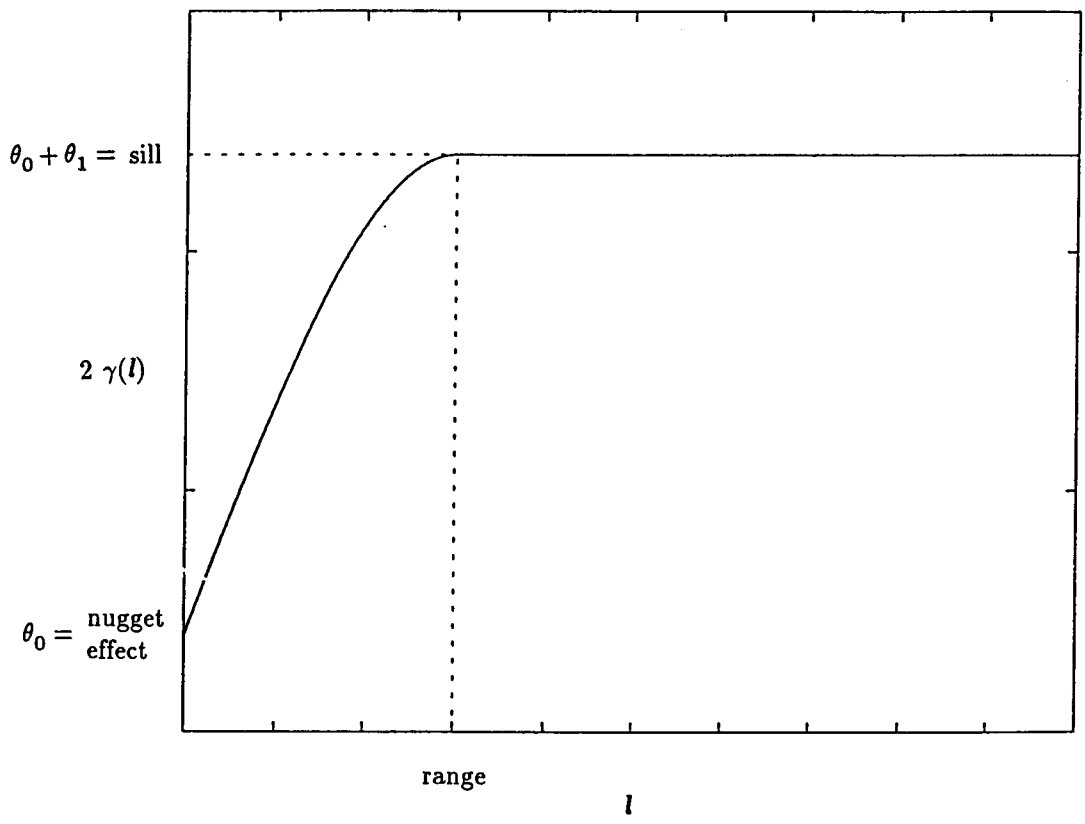
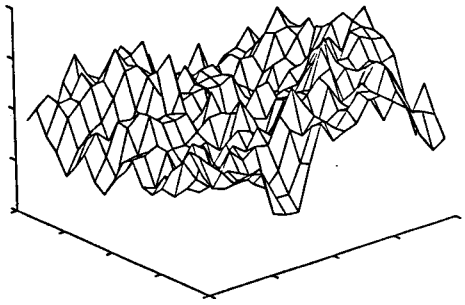
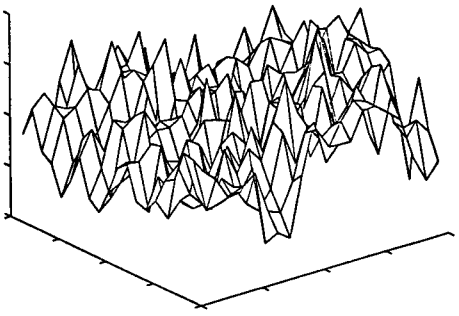


Figure 1.1 Features of the variogram :- the sill, nugget effect, and range.

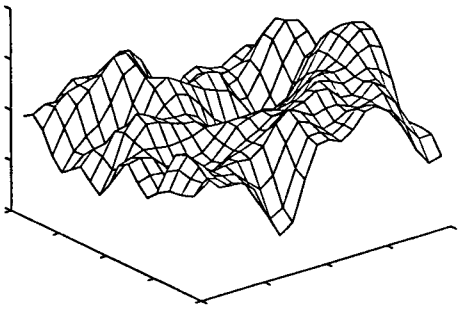
It is also advantageous to understand the relationship between the parameters of the covariance function or variogram model and the features of the underlying process. Figures 1.2, 1.3, and 1.4 illustrate the changes in the smoothness of the surface as each of the parameters θ_0, θ_1, and θ_2 is varied. Examples for both the exponential and Gaussian models are given. Each plot shows one realization from a process with the model indicated, generated using the spectral decomposition method (Cressie, 1991) on a 20×20 grid. In general, the



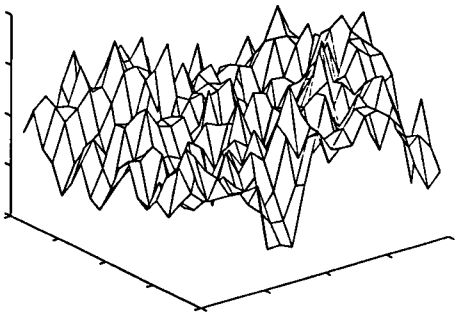
(a) exponential model, $\theta_0 = 0.0$



(b) exponential model, $\theta_0 = 0.6$

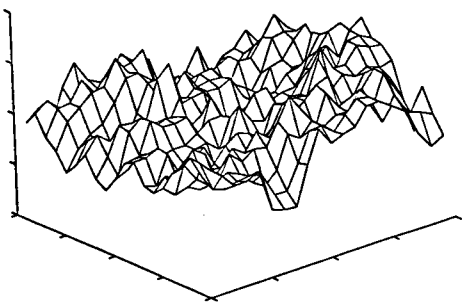


(c) Gaussian model, $\theta_0 = 0.0$

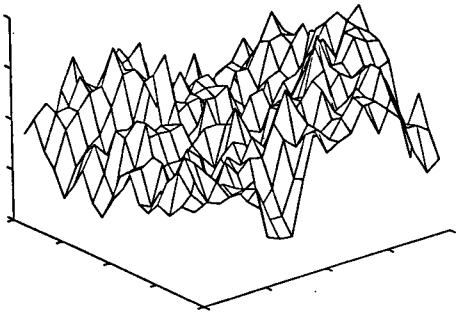


(d) Gaussian model, $\theta_0 = 0.6$

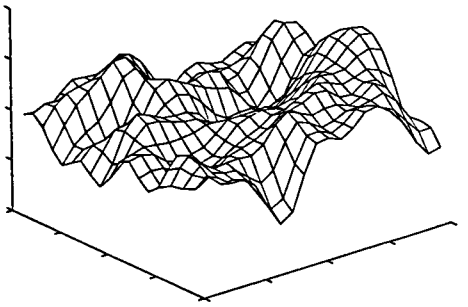
Figure 1.2 Example surfaces generated from exponential and Gaussian models, varying the value of the parameter θ_0. In all cases, $\theta_1 = 1.0$ and $\theta_2 = 0.4$.



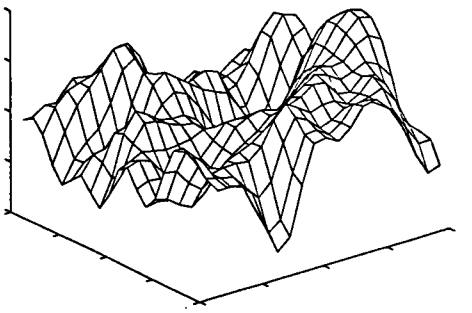
(a) exponential model, $\theta_1 = 0.75$



(b) exponential model, $\theta_1 = 1.5$

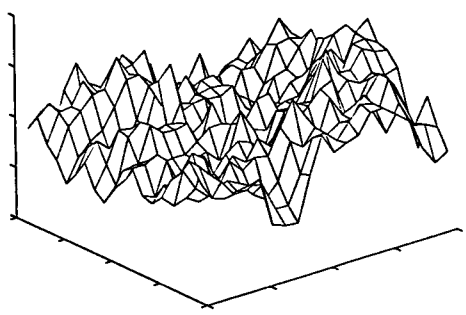


(c) Gaussian model, $\theta_1 = 0.75$

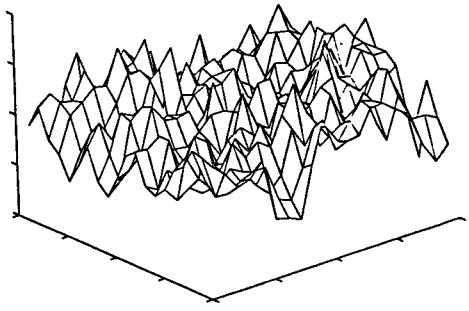


(d) Gaussian model, $\theta_1 = 1.5$

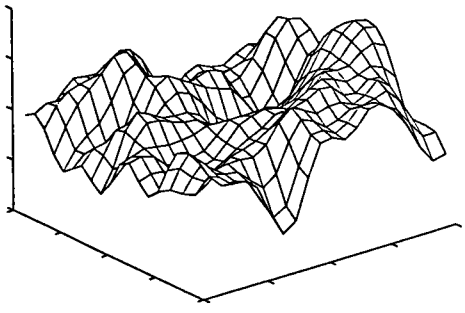
Figure 1.3 Example surfaces generated from exponential and Gaussian models, varying the value of the parameter θ_1. In all cases, $\theta_0 = 0.0$ and $\theta_2 = 0.4$.



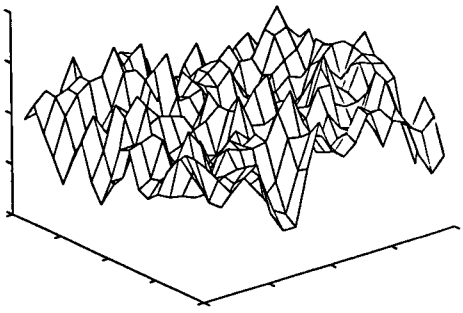
(a) exponential model, $\theta_2 = 0.4$



(b) exponential model, $\theta_2 = 0.8$



(c) Gaussian model, $\theta_2 = 0.4$



(d) Gaussian model, $\theta_2 = 0.8$

Figure 1.4 Example surfaces generated from exponential and Gaussian models, varying the value of the parameter θ_2. In all cases, $\theta_0 = 0.0$ and $\theta_1 = 1.0$.

process becomes "noisier" as the parameters increase in value. However, as is evident in the figures, processes defined by an exponential covariance function are far less sensitive to changes in the parameter θ_2 than are those defined by a Gaussian covariance model. This issue is discussed in greater detail in Chapter 3.

Further information on these and other aspects of geostatistical models may be found in the texts by Cressie (1991) and Christensen (1991).

Diagnostics for Spatial Linear Models

Diagnostic plots and statistics feature prominently in the exploratory data analysis toolbox of any statistician. These diagnostics are designed to summarize and highlight various characteristics of the data which might be of concern, or to suggest which analysis procedures are appropriate. Belsley *et al.* (1980), Cook and Weisberg (1982), Chambers *et al* (1983), Atkinson (1985), and Belsley (1991), among others, discuss useful diagnostics and plotting techniques, with a particular focus on regression analysis. In the regression context, many of the more popular diagnostics are based on examination of the residuals obtained from the model.

Similar techniques are available for the evaluation of spatial linear models. Although its utility is often criticized, cross-validation has long been popular with geostatisticians (Davis, 1987) as a means of assessing the predictive capacity of candidate models. Christensen, Johnson, and Pearson (1992, 1993) lay the groundwork for the development of more versatile case-deletion diagnostics. Table 1.1 summarizes the currently available case-deletion diagnostics for geostatistics, and their regression analysis counterparts.

Christensen *et al* (1992) develop influence diagnostics which are useful in the spatial setting. They define an extended version of Cook's (1977) D, which provides a measure of the influence of individual observations on the vector of coefficients, β, of the universal kriging model. A second statistic, serving the same purpose as DFFITS (Cook and Weisberg, 1982) in

the regression context, is introduced as a diagnostic to determine the influence of individual observations on the predictions produced by the model.

Table 1.1 : Currently available case-deletion diagnostics.

Regression	Spatial linear models
DFFITS	Christensen et al (1992)
DFBETAS	Influential observations
Cook's D	(prediction, $\hat{\beta}$'s)
	Christensen et al (1993)
	Influential observations
	(estimates of covariance
	function parameters)
PRESS	Various "PRESS"-like statistics
	(predictive ability of the model)

Christensen *et al* (1993) present diagnostics for identifying observations which are influential in estimating the parameters of the covariance function model. As such, these diagnostics have no direct counterparts in regression analysis, where the usual assumption is that $\text{Var}(\epsilon) = \sigma^2\mathbf{I}$. Working in the context of a restricted maximum likelihood approach to parameter estimation, Christensen *et al* (1993) define statistics to measure the influence of each observation on the vector of covariance function parameters $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2]'$, and on the individual components of $\boldsymbol{\theta}$. Thus, the diagnostic to detect influence on $\boldsymbol{\theta}$ bears some resemblance to Cook's D, while the diagnostics to detect influence on the individual parameters serve a similar purpose to DFBETAS (Cook and Weisberg, 1992). Recall, however, that the purpose of the diagnostics developed by Christensen *et al* (1993) is to detect influence on parameters of the covariance function, as opposed to parameters of the mean function of the universal kriging model.

Various versions of a statistic analogous to the PRESS statistic (Allen, 1971) of regression analysis have also found widespread applications in geostatistics (Knudsen and Kim, 1978; Davis, 1987; Cressie, 1991). These case-deletion statistics are used to cross-validate the chosen variogram or covariance function model and are measures of the predictive capacity of a given model. Cross-validation procedures are generally viewed as useful exploratory data analysis techniques, and may be used to select a "best" model for the covariance structure from a finite number of candidates.

Graphical techniques may also be employed to assess the performance of a model from a prediction standpoint, identify influential observations, and detect certain anomalies. Cressie (1991) gives a good overview of some commonly used exploratory plotting techniques for geostatistical data, while Bradley and Haslett (1992) describe more recent advancements. Several graphical techniques have been described in the geostatistical literature. Hamlett *et al* (1986) describe simple tools for the exploration of spatial structure of the data, including assessment of stationarity properties. Cressie (1984, 1991) uses simple "pocket plots" to detect localized regions of non-stationarity, while Chauvet (1982) and Cressie (1991) promote the use of variogram cloud plots as a powerful exploratory tool. Haslett *et al* (1991) stress the importance of detecting both global and local anomalies in the data. Plots based on standardized versions of the PRESS residuals have also found application in geostatistics. Cressie (1991) describes the use of histograms, stem-and-leaf plots, and normal probability plots based on standardized PRESS residuals to detect outliers.

Outline of the Thesis

The thesis is presented in manuscript format, so that the second and third chapters consist of two papers to be submitted for publication in the journal "Mathematical Geology", while the final chapter summarizes the conclusions of the two preceding chapters. The results of the second chapter were previously presented as a contributed paper at the 1995 Winter Meeting

of the American Statistical Association. This paper deals with examination of the distribution of a particular case-deletion diagnostic statistic, of the type which has frequently been used by geostatisticians to evaluate spatial linear models. A saddlepoint approximation to this distribution is presented, which provides the analyst with a more informative yardstick for comparing candidate models than the currently available *ad hoc* decision rules. Recommendations for calculating this statistic on a regional basis are also presented.

The third chapter focuses on the behavior of the modified PRESS residuals under various misspecifications of the spatial covariance model. Additional properties of the statistic described in the second chapter are also examined. The behavior of this statistic, T_{PR}, and other summaries related to the distribution of the PRESS residuals are investigated through a simulation study, in an attempt to identify consistent patterns under particular model misspecifications. Investigation of patterns in these residuals leads to modifications of normal probability plots which point to over- or under-estimation of the parameters of the covariance function model.

USE OF A MODIFIED PRESS STATISTIC FOR SPATIAL LINEAR MODEL EVALUATION

Caryn M. Thompson and Fred L. Ramsey

Abstract

Case-deletion diagnostics are currently used by geostatisticians to evaluate spatial covariance models. This paper focuses on a diagnostic based on a standardized version of the PRESS statistic, and its use in assessing the predictive capacity of covariance models. A saddlepoint approximation to the distribution of the standardized PRESS statistic is proposed. Using this approximation, approximate p-values for the observed statistic under an hypothesized covariance model may be obtained. The adequacy of the distributional and p-value approximations are evaluated in a simulation study. Implementation of the proposed method is illustrated through an example.

Introduction

Cross-validation is a common method for the evaluation of statistical models. The procedure involves the sequential deletion of observations, followed by prediction at the deleted points using the remaining data. This technique may be used to construct "leave-one-out" or "case-deletion" diagnostics. In the linear regression setting, the PRESS statistic (Allen, 1971) is a popular tool for evaluating the predictive ability of candidate models. Other case-deletion diagnostics are available for a variety of purposes, including identification of influential observations (Cook and Weisberg, 1982).

Cross-validation has also seen fairly wide application in the geostatistical setting as a diagnostic tool for assessing variogram and covariance function models. While several authors (Davis, 1987; Isaaks and Srivastava, 1989; Solow, 1990; Cressie, 1991) caution against the use of cross-validation for any purpose beyond that of a simple diagnostic technique, some attempt has

been made to develop inferential procedures based on cross-validation statistics (Carr and Roberts, 1989). Recently, Christensen, Pearson and Johnson (1992, 1993) have proposed case-deletion diagnostics which may be used to identify observations which are influential, either with respect to estimation of the parameters of spatial models, or in terms of the resulting predictions. As part of their development, Christensen *et al* (1992) show that case-deletion diagnostics are easily computed for universal kriging models.

In this paper, a modification of the PRESS statistic is recommended as a diagnostic to evaluate spatial models. The development is presented in terms of the universal kriging model, with simplifications for ordinary kriging highlighted. Distribution theory for the statistic is presented. It is demonstrated that a saddlepoint approximation may be used to identify feasible values of the standardized PRESS statistic under an hypothesized model. Procedures for determining approximate p-values associated with this statistic are also described.

Background and Notation

Following the notation of Christensen, Johnson, and Pearson (1993), suppose the data consist of a vector $\boldsymbol{Y} = (y(\boldsymbol{u}_1)\ y(\boldsymbol{u}_2), \ldots, y(\boldsymbol{u}_n))'$ of n observations, taken at locations $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$, within some domain of interest, D. At each location $\boldsymbol{u}_i$, let $y(\boldsymbol{u}_i)$ be a random process of the form

$$y(\boldsymbol{u}_i) = m(\boldsymbol{u}_i) + \epsilon(\boldsymbol{u}_i),$$

where $m(\boldsymbol{u}_i)$ is a function representing the mean of $y(\boldsymbol{u}_i)$ and $\epsilon(\boldsymbol{u}_i)$ is a zero-mean error process. Under the universal kriging model, it is assumed that $m(\boldsymbol{u}_i)$ may be expressed as a linear combination of p known functions of the $\boldsymbol{u}_i$'s,

$$m(\boldsymbol{u}_i) = \sum_{j=1}^{p} \beta_j x_j(\boldsymbol{u}_i),$$

where the β_j's are unknown parameters. Let

$$x_{ij} = x_j(\boldsymbol{u}_i), \quad \boldsymbol{x}_i' = (x_{i1},...,x_{ip}) ,$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{x}_n' \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix},$$

and

$$\epsilon_i = \epsilon(\boldsymbol{u}_i) \text{ and } \boldsymbol{\epsilon} = (\epsilon_1,...,\epsilon_n)',$$

so that

$$m(\boldsymbol{u}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$$

and

$$y(\boldsymbol{u}_i) = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i ,$$

for $i = 1,...,n$. Thus, the universal kriging model may be expressed as the linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} ,$$

where $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{V}$, with elements $v_{ij} = \text{Cov}[\epsilon(\boldsymbol{u}_i), \epsilon(\boldsymbol{u}_j)] = C(\boldsymbol{u}_i,\boldsymbol{u}_j)$. It is assumed that $\boldsymbol{V}$ is nonsingular, and $\boldsymbol{X}$ has full column rank. The model is easily simplified to the ordinary kriging setting by specifying $\boldsymbol{X}$ to be an $n \times 1$ vector of 1's.

Under this model the best linear unbiased predictor (BLUP) of y at any location $u_0 \in D$ is given by

$$\hat{y}(\boldsymbol{u}_0) = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} + \boldsymbol{V}_0'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

where $\boldsymbol{x}_o' = (x_1(\boldsymbol{u}_0), \ldots, x_p(\boldsymbol{u}_0))$, $\boldsymbol{V}_0' = [C(\boldsymbol{u}_0,\boldsymbol{u}_1), \ldots, C(\boldsymbol{u}_0,\boldsymbol{u}_n))$ and $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y}$ (Christensen, 1990). The mean square prediction error (*mspe*), or *kriging variance* at any location $\boldsymbol{u}_0$ is given by

$$\sigma_k^2(\boldsymbol{u}_0) = C(\boldsymbol{0}) - \boldsymbol{V}_0'\boldsymbol{V}^{-1}\boldsymbol{V}_0 + [\boldsymbol{x}_0 - \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}_0]'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}[\boldsymbol{x}_0 - \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{V}_0] .$$

A Standardized Version of the PRESS Statistic

A popular criterion for the evaluation of multiple linear regression models is the prediction sum of squares (PRESS) statistic (Allen, 1971). A similar statistic may be used as a diagnostic for more general linear models. The PRESS statistic is calculated using a "leave-one-out" cross-validation procedure to generate prediction errors, or PRESS residuals. Christensen *et al* (1992) show how to calculate these prediction errors using relatively straightforward updating formulae. The i^{th} prediction error $e_{-i}(\boldsymbol{u}_i)$ is equal to $y(\boldsymbol{u}_i) - \hat{y}_{-i}(\boldsymbol{u}_i)$, where $\hat{y}_{-i}(\boldsymbol{u}_i)$ is the BLUP of $y(\boldsymbol{u}_i)$ obtained when the ith observation is removed from the data, and the remaining $n-1$ observations used to obtain an estimate of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}_{-i}$. The vector $\boldsymbol{Y}_{-i}$ and the matrix $\boldsymbol{X}_{-i}$ are defined to be $\boldsymbol{Y}$ and $\boldsymbol{X}$ with the i^{th} row removed, while $\boldsymbol{V}_{-i}$ is the covariance matrix for $\boldsymbol{Y}_{-i}$. Christensen *et al* (1992) show that

$$e_{-i}(\boldsymbol{u}_i) = (\tilde{y}_i(\boldsymbol{u}_i) - \tilde{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}})\frac{s_i}{s_i - \tilde{h}_i} ,$$

where , $\tilde{y}_i(\boldsymbol{u}_i) = y_i - \boldsymbol{Y}_{-i}'\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_i$, $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \boldsymbol{X}_{-i}'\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_i$, $s_i = v_{ii} - \boldsymbol{v}_i'\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_i$, $\tilde{h}_i = \tilde{\boldsymbol{x}}_i'(\boldsymbol{X}\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\tilde{\boldsymbol{x}}_i$ and $\boldsymbol{v}_i$ is the i^{th} column of $\boldsymbol{V}$ with the element v_{ii} deleted.

In order for case-deletion procedures to be feasible from a computational standpoint, the inverses of the matrices $\boldsymbol{V}_{-i}$ must be easily obtained. Fortunately, the inverse matrices $\boldsymbol{V}_{-i}^{-1}$ may be calculated directly from the elements of $\boldsymbol{V}^{-1}$. Christensen *et al* (1992) show

that if V^{-1} is written as

$$V^{-1} = \begin{bmatrix} v^{ii} & \lambda_i' \\ \lambda_i & \Lambda_{-i} \end{bmatrix},$$

then $V_{-i}^{-1} = \Lambda_{-i} - \lambda_i \lambda_i' / v^{ii}$.

The PRESS statistic is then equal to $\sum_{i=1}^{n} e_{-i}(\boldsymbol{u}_i)^2$. Various modifications of this statistic have been suggested (Cook and Weisberg, 1982), among them a version based on the standardized PRESS residuals $t_i = e_{-i}(\boldsymbol{u}_i)/\sigma_{e_{-i}}$, where

$$\sigma^2_{e_{-i}} = \mathrm{Var}(e_{-i}(\boldsymbol{u}_i)) = \frac{s_i^2}{s_i - \tilde{h}_i} \quad .$$

Therefore,

$$t_i = \frac{(\tilde{y}_i(\boldsymbol{u}_i) - \tilde{\boldsymbol{x}}_i' \hat{\boldsymbol{\beta}})}{\sqrt{s_i - \tilde{h}_i}} \quad .$$

The standardized PRESS statistic will be denoted by $T_{PR} = \sum_{i=1}^{n} t_i^2$.

Similar case-deletion statistics based on prediction errors, sometimes known as cross-validation diagnostics, are commonly used in geostatistics. Cressie (1991) mentions $\frac{1}{n}\sum_{i=1}^{n} t_i$ and $\left(\frac{1}{n}\sum_{i=1}^{n} t_i^2\right)^{1/2}$ as two possibilities. Typically, the examination of these statistics has been *ad hoc* in nature, with the rule of thumb being that if the proposed model is appropriate, the value of the first statistic should be close to 0, while the second should be approximately equal to 1. Although their values under a particular hypothesized model might be deemed to be either "large" or "small", the question of whether such statistics are sufficiently extreme to suggest inconsistencies with the proposed model cannot be answered without accompanying distributional theory.

In the development which follows, T_{PR} is proposed as an appropriate diagnostic for the evaluation of the predictive ability of spatial covariance models, and certain of its distributional

properties are determined. An easily obtained approximation to the distribution of T_{PR}, along with accompanying approximate p-values, is presented.

Properties of the Distribution of the Statistic T_{PR}

First we consider distributional properties of the standardized PRESS residuals, the t_i's. Under the additional assumption that $\epsilon \sim N(\mathbf{0}, \boldsymbol{V})$ and $\boldsymbol{V}$ is known, $t_i \sim N(0,1)$ for $i = 1, 2, \ldots, n$. The covariance between any two standardized PRESS residuals t_i and t_j is given by

$$\begin{aligned}\mathrm{Cov}(t_i, t_j) &= \frac{1}{\sqrt{s_i - \tilde{h}_{\,i}}\sqrt{s_j - \tilde{h}_{\,j}}}\,\mathrm{Cov}[\tilde{y}_i - \tilde{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}},\ \tilde{y}_j - \tilde{\boldsymbol{x}}_j'\hat{\boldsymbol{\beta}}] \\ &= \mathrm{Cov}[\boldsymbol{k}_i'\boldsymbol{Y},\ \boldsymbol{k}_j'\boldsymbol{Y}]\ .\end{aligned}$$

The vector $\boldsymbol{k}_i'$ is given by $[(\boldsymbol{a}_i' - \boldsymbol{v}_i'\boldsymbol{V}_{-i}^{-1}\boldsymbol{B}_i) - \tilde{\boldsymbol{x}}_i'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}]/(\sqrt{s_i - \tilde{h}_{\,i}})$, where $\boldsymbol{a}_i$ is the i^{th} column of an $n \times n$ identity matrix, and $\boldsymbol{B}_i$ is an $(n-1) \times n$ matrix consisting of an $n \times n$ identity matrix with the i^{th} row removed. The covariance expression simplifies to

$$\mathrm{Cov}(t_i, t_j) = \frac{s_{ij} - \tilde{h}_{\,ij}}{\sqrt{s_i - \tilde{h}_{\,i}}\sqrt{s_j - \tilde{h}_{\,j}}}\ ,$$

where $s_{ij} = v_{ij} - \boldsymbol{v}_i'\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_{-i,j} - \boldsymbol{v}_j'\boldsymbol{V}_{-j}^{-1}\boldsymbol{v}_{-j,i} + \boldsymbol{v}_i'\boldsymbol{V}_{-i}^{-1}\boldsymbol{V}_{-i,-j}\boldsymbol{V}_{-j}^{-1}\boldsymbol{v}_j$,
$\tilde{h}_{\,ij} = \tilde{\boldsymbol{x}}_i'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\tilde{\boldsymbol{x}}_j$, $\boldsymbol{v}_{-i,j}$ is the j^{th} column of $\boldsymbol{V}_{-i}$, and $\boldsymbol{V}_{-i,-j}$ is the matrix $\boldsymbol{V}$ with its i^{th} row and j^{th} column removed.

The variance-covariance matrix of $\boldsymbol{t} = [t_1, \ldots, t_n]'$ is then given by

$$\mathrm{Cov}(\boldsymbol{t}) = \boldsymbol{\Sigma} = \boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\ , \tag{2.1}$$

where $\boldsymbol{K}'$ has as its n rows the vectors of constants $\boldsymbol{k}_i'$, $i = 1, \ldots, n$ as defined above.

It is readily seen that the t_i's are not independent, even in the case where $\boldsymbol{V} = \sigma^2\mathbf{I}$. Therefore, $T_{PR} = \sum_{i=1}^{n} T_i = \sum_{i=1}^{n} t_i^2$ is the sum of n correlated χ_1^2 random variables. The joint distribution of $T_1, \ldots, T_n$ has been referred to as the "multivariate chi-squared distribution" (Krishnaiah *et al*, 1963). Gordon and Ramig (1983) give results for the cumulative distribution function (cdf) of a random variable such as T_{PR}. In this case, the joint characteristic function of $T_1, \ldots, T_n$ is given by

$$\psi_{T_1, \ldots, T_n}(w_1, \ldots, w_n) = |\mathbf{I} - 2i\boldsymbol{\Sigma}\boldsymbol{D}_w|^{-1/2}$$

(Lukacs and Laha, 1964), where $\boldsymbol{D}_w$ is a diagonal matrix with diagonal elements $w_1, \ldots, w_n$, and $\mathbf{I}$ is the identity matrix of dimensions $n \times n$. Thus, T_{PR} has a characteristic function (Gnedenko, 1968)

$$\psi_{T_{PR}}(w) = \psi_{T_1, \ldots, T_n}(w, \ldots, w) = |\mathbf{I} - 2iw\boldsymbol{\Sigma}|^{-1/2} = |\mathbf{I} - 2iw\boldsymbol{\Lambda}|^{-1/2},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}$ (Gordon and Ramig, 1983).

Unfortunately, $\psi_{T_{PR}}(w)$ is not easily inverted, and thus the exact cdf of T_{PR} is not readily obtained. The distribution of T_{PR} is likely to be similar to a chi-squared distribution, as $\boldsymbol{\Sigma}$ may be expressed in quadratic form. Therefore, it is reasonable to assume that the moment generating function (mgf) of T_{PR} exists, and is given by

$$M_{T_{PR}}(w) = \psi_{T_{PR}}(-iw) = |\mathbf{I} - 2w\boldsymbol{\Lambda}|^{-1/2} = \prod_{i=1}^{n} (1 - 2w\lambda_i)^{-1/2}.$$

Gordon and Ramig (1983) propose a numerical approximation to the inversion of the cdf of T_{PR}. Mathai and Provost (1992) present a general formulation for the distribution of random variables having quadratic forms such as that of the statistic T_{PR}, in terms of a power series expansion. An alternative procedure would be to develop an approximation through a "matching of moments" technique, such as that described by Solomon and Stephens (1977). As

will be shown is next section, however, the distribution of T_{PR} may be approximated easily and accurately using a saddlepoint approximation. This approximation requires knowledge of only the eigenvalues of the matrix Σ, and is readily implemented under general conditions.

A Saddlepoint Approximation

Saddlepoint approximations have been shown to be effective in a variety of applications because of their accuracy and ease of use. Reid (1988) reviews the theory and development of the various saddlepoint methods. The technique described here requires knowledge of the cumulant generating function of the statistic of interest.

One of the original usages of the saddlepoint approximation was to obtain approximate probability density functions for statistics through inversion of their characteristic or moment generating functions. Daniels (1954) describes an approximation to the density function of the mean $\overline{x}$ of m independent, continuous and identically distributed random variables. A similar technique may be employed to find an approximate density function for the statistic T_{PR}. The development which follows does not involve an assumption of independence of the individual terms of T_{PR}. Rather, T_{PR} is considered to be the mean of a single observation (thus, $m = 1$ in this case). An alternative approach to obtaining a saddlepoint approximation to the distribution of the mean of m dependent, but identically distributed, variables was developed by Robinson (1982), and discussed briefly by Reid (1988).

In order to implement the procedure, it is first necessary to locate the saddlepoints of T_{PR}. These occur where

$$K'(w) = T_{PR} \tag{2.2}$$

where $K'(w)$ is the first derivative of the the cumulant generating function of T_{PR}. The saddlepoint corresponding to a particular value of T_{PR} is denoted w_0. Applying the

approximation described by Daniels (1954), the density function $f(T_{PR})$ may be expanded to give

$$f(T_{PR}) \sim g(T_{PR}) \left\{1 + [1/8\ \lambda_4(w_0) - 5/24\ \lambda_3^2\ (w_0)] + \ldots \right\}$$

where $\lambda_j(w) = K^{(j)}(w)/[K''(w)]^{j/2}$ for $j \geq 3$ and

$$g(T_{PR}) = \left\{\frac{1}{2\pi K''(w_0)}\right\}^{1/2} e^{[K(w_0) - w_0 T_{PR}]}.$$

The function $g(T_{PR})$ is known as the saddlepoint approximation to $f(T_{PR})$.

This procedure is easily implemented. First, it is necessary to calculate the first two derivatives of $K(w)$. Assuming the moment generating function exists, the cumulant generating function for T_{PR} is given by

$$K(w) = ln\ M_{PR}(w) = -\frac{1}{2}\, ln\, |\, \mathbf{I} - 2w\mathbf{\Lambda}\, |$$

$$= -\frac{1}{2}\, ln \prod_{j=1}^{n} (1 - 2w\lambda_j)$$

$$= -\frac{1}{2} \sum_{j=1}^{n} ln(1 - 2w\lambda_j) \ .$$

Therefore, the first and second derivatives of $K(w)$ are given by

$$K'(w) = \sum_{j=1}^{n} \left(\frac{\lambda_j}{1 - 2w\lambda_j} \right)$$

and

$$K''(w) = \sum_{j=1}^{n} \frac{2\lambda_j^2}{(1 - 2w\lambda_j)^2}$$

respectively. Thus, the saddlepoint approximation to the probability density function of T_{PR} is given by

$$g(T_{PR}) = \left[\frac{1}{2\pi \sum_{j=1}^{n} \frac{2\lambda_j^2}{(1-2w_0\lambda_j)^2}}\right]^{1/2} e^{\left[-1/2 \sum_{j=1}^{n} ln(1-2w_0\lambda_j) - w_0 T_{PR}\right]} .$$

This function is easily programmed using, for example, the software package MATLAB (The Math Works, 1992).

In order to use a statistic such as T_{PR} effectively for model evaluation, it is necessary to calculate upper and lower tail probabilities for the approximate density function. Although one approach would be to calculate tail probabilities through numerical integration of the saddlepoint approximation $g(T_{PR})$, accurate approximations to the tail probabilities are available. Daniels (1987) discusses three methods for the estimation of tail probabilities, including one due to Lugannani and Rice (1980). The Luganani and Rice method was selected, since it is valid for both upper and lower tails. Using this method, the approximate quantiles for T_{PR} are given by

$$Q_{LR}(T_{PR}) = \phi(\hat{\zeta})\left\{R(\hat{\zeta}) + \frac{1}{\hat{z}} - \frac{1}{\hat{\zeta}}\right\} .$$

where $\phi(x)$ is the standard normal density function, $\Phi(x)$ is the standard normal cumulative distribution function, $R(x) = \{1-\Phi(x)\}/\phi(x)$, $\hat{z} = w_0\sqrt{K''(w_0)}$, and $\hat{\zeta} = \sqrt{2(w_0 T_{PR} - K(w_0)}$. The Lugannani and Rice approximation has been shown to be relatively easy to use, and extremely accurate (Reid, 1988) .

As for the saddlepoint approximation itself, the Lugannani and Rice method requires knowledge of the eigenvalues of the covariance matrix Σ, and its derivatives. If it is desired to report an estimated tail probability for a given a value of T_{PR}, it is also necessary to locate the corresponding saddlepoint, w_0. This may be achieved, for instance, using the Newton-Raphson procedure.

Simulation Results

The performance of the approximation was assessed for several covariance function models, for both transects and gridded data. The intent was to illustrate the behavior of the distribution of T_{PR}, and the associated approximation, for varying degrees of spatial correlation in the underlying process, with moderate sample sizes ($n \simeq 50$). For simplicity, ordinary kriging models were used, with zero nugget effect. An illustrative example involving a universal kriging model is presented in the following section.

For each null covariance model, 10,000 simulated values of T_{PR} were obtained directly by calculating Σ using (2.1), and then employing the spectral decomposition simulation method (Cressie, 1991). Figures 2.1 – 2.3 show plots of a histogram for the 10,000 simulated values of T_{PR}, approximate density $g(T_{PR})$, and the χ^2_{50} density as a point of reference, under three different hypothesized covariance models, each defined on transects of length 50 units. The models had effective ranges of 5, 20 and 2 respectively (Journel and Huijbregts, 1978), and reflect an increasing degree of spatial autocorrelation (with Figure 2.3 corresponding to the strongest spatial autocorrelation). Observation of Figures 2.1 – 2.3 shows good agreement between the histograms for the simulations and the saddlepoint approximations, particularly in the tails.

Approximate upper tail probabilities were calculated for various models using the method described above, and processes defined on both transects and two-dimensional grids. The observed percentiles of the 10,000 simulated values were used to estimate quantiles of the distribution of T_{PR}. Estimated upper tail probabilities were then calculated for the estimated quantiles, and denoted $\widehat{p}_{LR}$. Appropriate saddlepoints were obtained by solving (2.2) using the Newton-Raphson method. All results are tabulated in Tables 2.1 – 2.6. The effective ranges associated with the models were 5, 20, 2, 1, 10, and 2 respectively. Tables 2.1 – 2.3 are presented in order of increasing degree of spatial autocorrelation, as are Tables 2.4 – 2.6.

Tail probabilities based on a χ^2_n distribution, denoted $\widehat{p}_{CHI}$, are given for comparison in these tables. In most cases, the saddlepoint approximation performs reasonably well, particularly in the upper tails. These results also demonstrate the inadequacies of the χ^2_n distribution as an approximation to the distribution of T_{PR}.

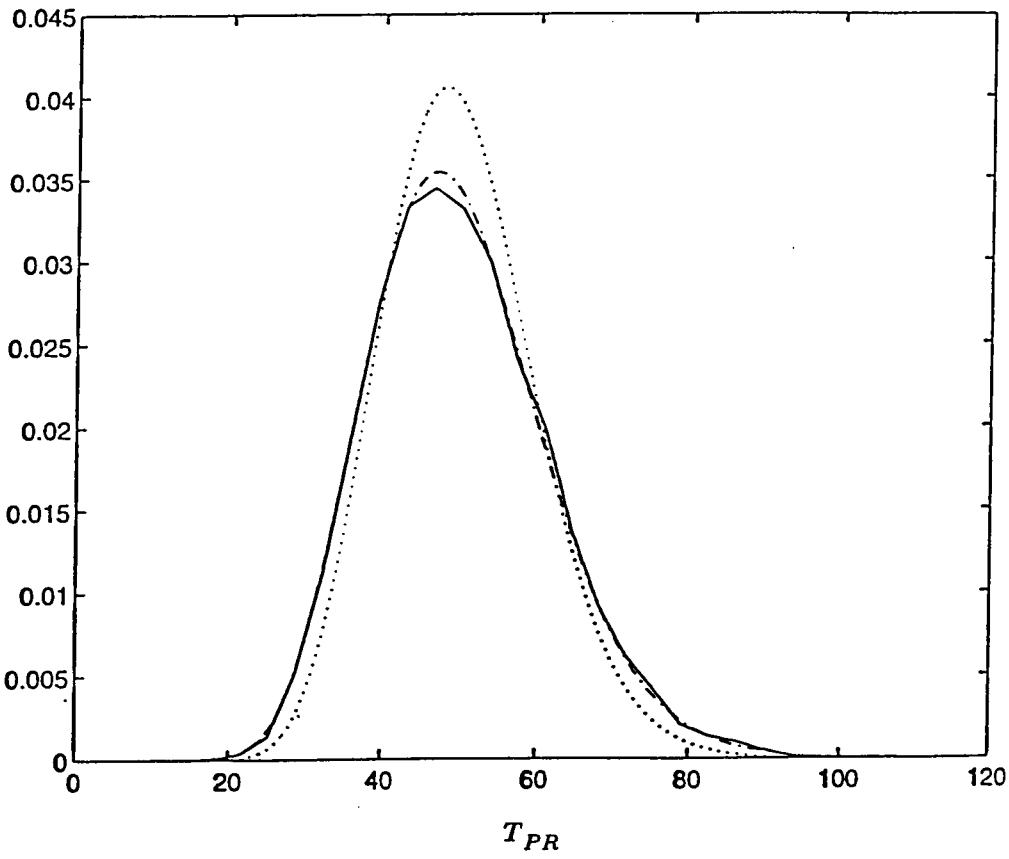


Figure 2.1 Exponential model : $C(l) = exp(-0.6l)$, $l = \| \boldsymbol{u}_i - \boldsymbol{u}_j \|$. Transects of length 50.

Chi-square distribution, 50 df : .....
Saddlepoint approximation : — · — ·
Histogram for simulated values of T_{PR} : ————

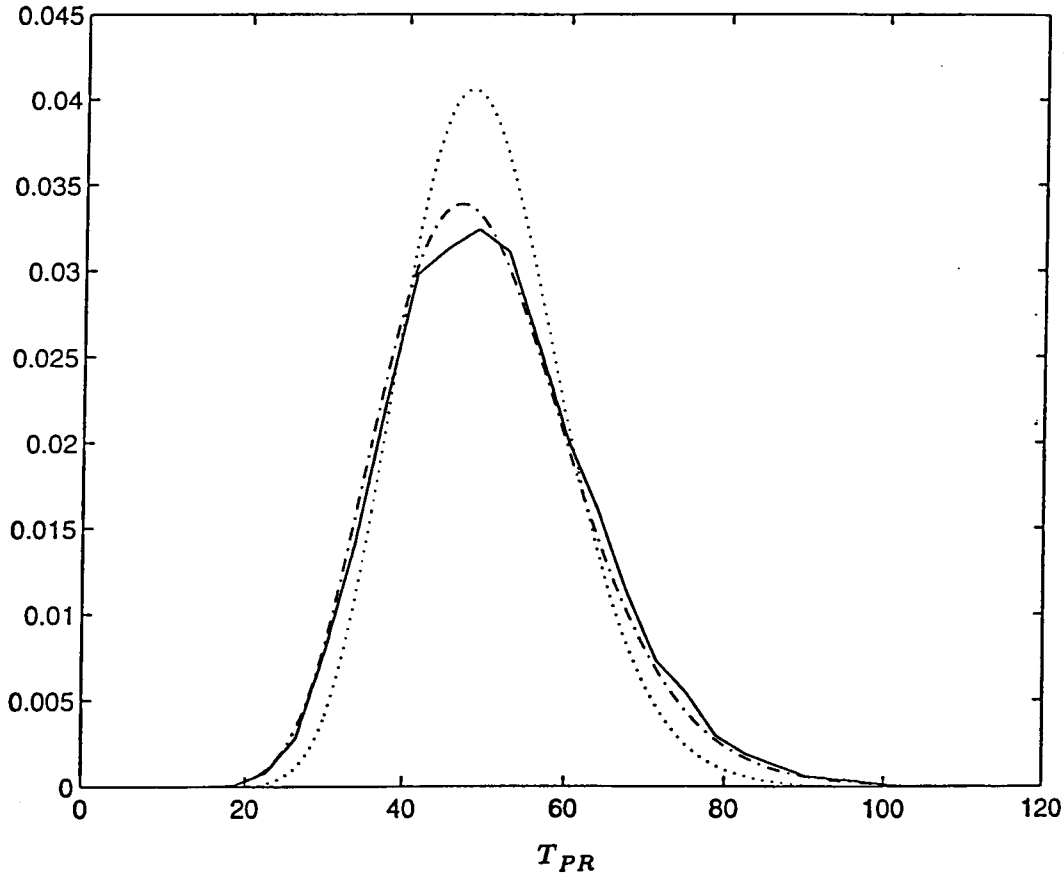


Figure 2.2 Exponential model : $C(l) = exp(-0.15l)$, $l = \| \boldsymbol{u}_i - \boldsymbol{u}_j \|$. Transects of length 50.

Chi-square distribution, 50 df : .....
Saddlepoint approximation : $- \cdot - \cdot$
Histogram for simulated values of T_{PR} : ————

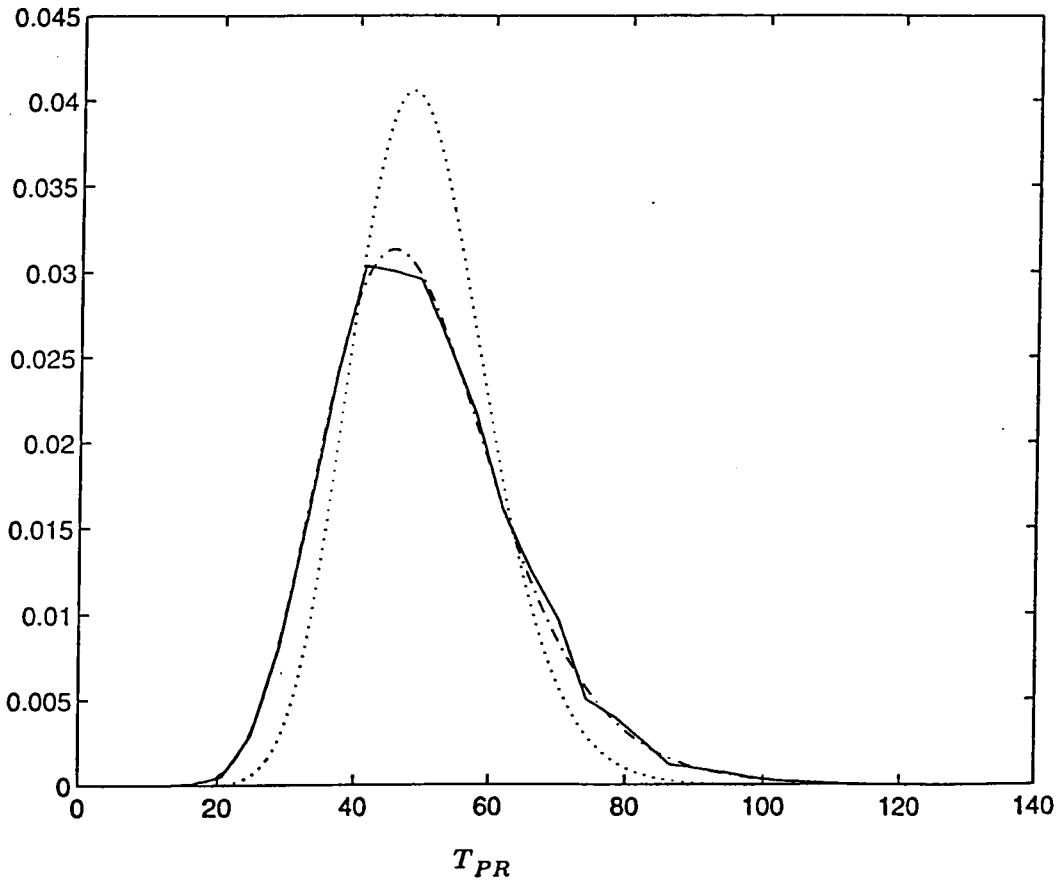


Figure 2.3 Gaussian model : $C(l) = exp\Big(-\Big(\frac{\sqrt{3}}{2}\, l\Big)^2\Big)$, $l = \| \boldsymbol{u}_i - \boldsymbol{u}_j \|$. Transects of length 50.

Chi-square distribution, 50 df : .....

Saddlepoint approximation : – · – ·

Histogram for simulated values of T_{PR} : ————

Table 2.1 Tail probability estimates for exponential model : $C(l) = exp(-0.6l)$, transects of length 50.

T_{PR}	$\widehat{p}$	$\widehat{p}_{LR}$	$\widehat{p}_{CHI}$
25.95	.9950	.9941	.9959
27.47	.9900	.9894	.9930
30.61	.9750	.9711	.9812
32.99	.9500	.9463	.9640
36.00	.9000	.8975	.9272
41.82	.7500	.7465	.7928
57.15	.2500	.2518	.2300
65.49	.1000	.0996	.0638
70.63	.0500	.0510	.0233
75.61	.0250	.0251	.0075
80.58	.0100	.0118	.0021
85.38	.0050	.0054	.0010

Table 2.2 Tail probability estimates for exponential model : $C(l) = exp(-0.15l)$, transects of length 50.

T_{PR}	$\widehat{p}$	$\widehat{p}_{LR}$	$\widehat{p}_{CHI}$
25.30	.9950	.9940	.9968
26.70	.9900	.9899	.9949
29.97	.9750	.9722	.9844
32.65	.9500	.9456	.9670
35.78	.9000	.8970	.9306
41.85	.7500	.7475	.7919
58.23	.2500	.2508	.1998
67.23	.1000	.0983	.0462
72.79	.0500	.0500	.0145
77.78	.0250	.0259	.0044
83.65	.0100	.0113	.0009
88.16	.0050	.0057	.0002

Table 2.3 Tail probability estimates for Gaussian model : $C(l) = exp\Big(-\Big(\frac{\sqrt{3}}{2}\, l\Big)^2\Big)$, transects of length 50.

T_{PR}	$\widehat{p}$	$\widehat{p}_{LR}$	$\widehat{p}_{CHI}$
24.01	.9950	.9936	.9981
25.62	.9900	.9883	.9964
28.10	.9750	.9744	.9914
31.02	.9500	.9458	.9788
34.11	.9000	.8981	.9526
40.36	.7500	.7482	.8349
58.10	.2500	.2470	.2033
68.04	.1000	.0998	.0395
74.81	.0500	.0491	.0091
80.94	.0250	.0246	.0019
89.23	.0100	.0092	.0002
94.60	.0050	.0047	2.8×10^{-5}

Table 2.4 Tail probability estimates for exponential model : $C(l) = exp(-l)$, 7×7 grid.

T_{PR}	$\widehat{p}$	$\widehat{p}_{LR}$	$\widehat{p}_{CHI}$
26.13	.9950	.9944	.9957
27.73	.9900	.9895	.9924
30.29	.9750	.9751	.9829
32.55	.9500	.9528	.9679
35.59	.9000	.9043	.9334
41.23	.7500	.7528	.8104
55.82	.2500	.2477	.2709
63.06	.1000	.1030	.0970
67.85	.0500	.0521	.0410
72.20	.0250	.0264	.0166
77.51	.0100	.0107	.0047
81.70	.0050	.0050	.0016

Table 2.5 Tail probability estimates for exponential model : $C(l) = exp(-0.3l)$, 7×7 grid.

T_{PR}	$\widehat{p}$	$\widehat{p}_{LR}$	$\widehat{p}_{CHI}$
25.19	.9950	.9948	.9970
26.89	.9900	.9898	.9943
29.46	.9750	.9761	.9867
31.99	.9500	.9520	.9723
35.01	.9000	.9052	.9415
40.76	.7500	.7580	.8239
56.03	.2500	.2527	.2642
63.78	.1000	.1052	.0860
68.84	.0500	.0538	.0337
73.31	.0250	.0281	.0129
79.29	.0100	.0110	.0030
83.64	.0050	.0053	.0009

Table 2.6 Tail probability estimates for Gaussian model : $C(l) = exp\left(-\left(\frac{\sqrt{3}}{2}\, l\right)^2\right)$, 7×7 grid.

T_{PR}	$\widehat{p}$	$\widehat{p}_{LR}$	$\widehat{p}_{CHI}$
20.89	.9950	.9942	.9995
22.53	.9900	.9890	.9990
25.09	.9750	.9747	.9971
27.81	.9500	.9487	.9922
31.02	.9000	.9015	.9788
37.63	.7500	.7531	.8985
57.64	.2500	.2583	.2160
70.75	.1000	.1002	.0227
79.78	.0500	.0499	.0026
89.10	.0250	.0240	.0002
99.85	.0100	.0103	4.3×10^{-6}
107.12	.0050	.0058	2.6×10^{-7}

Example

To illustrate the application of the methods, consider the aquifier data analyzed by L. W. Lenfest (unpublished work), Jones (1989), and Christensen *et al* (1992). Jones (1989) compares an isotropic model with an anisotropic model, while Christensen *et al* use the restricted maximum likelihood procedure (REML) to fit a universal kriging model with Gaussian covariance structure and incorporating measurement error to the data.

Figure 2.4 shows sampling locations for the 93 observations in this data set. A simplified version of the universal kriging model selected by Christensen *et al*, with isotropic Gaussian covariance structure of the form

$$C(l) = \begin{cases} \theta_1 exp(-(\theta_2 l)^2) & \text{for } l > 0 \\ \theta_0 + \theta_1 & \text{for } l = 0 \end{cases},$$

and ignoring measurement error, was fitted to the data using the REML procedure. The resulting estimates were $\widehat{\boldsymbol{\theta}} = [26.35,\ 1146,\ 0.1241]'$ and $\widehat{\boldsymbol{\beta}} = [2250, -0.7370, -3.037]'$ respectively, where $p = 3$, $\boldsymbol{x}_0$ is a column of 1's, and $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are vectors containing the latitude and longitude of each of the 93 observations. PRESS residuals were obtained under this model, and the value of T_{PR} found to be 92.99 with $\widehat{p}_{LR} = 0.4741$. To illustrate an additional potential usage of T_{PR} and its approximate distribution, the Saratoga Valley region was arbitrarily split into 5 subregions (Figure 2.4). The modified PRESS statistic was calculated for each of these regions, with results presented in Table 2.7.

Christensen *et al* (1992) identified observations 49 and 50 from Region I as being potentially influential for prediction purposes. Using the model described above, the value of the standardized PRESS residuals for these two observations were found to be $t_{49} = -3.54$ and $t_{50} = 3.62$ respectively. The estimation procedure was repeated, first with these two observations eliminated for the purposes of obtaining REML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and then

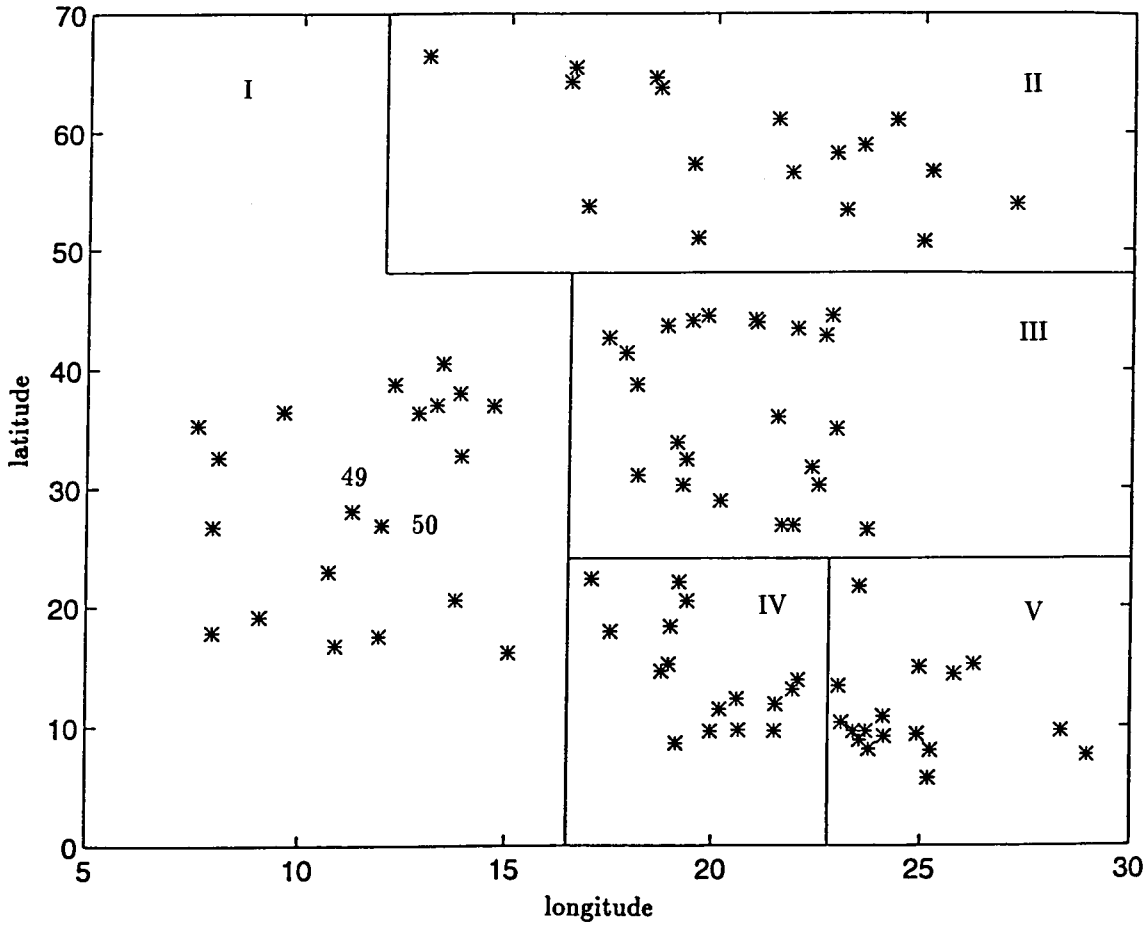


Figure 2.4 Sampled locations for the Saratoga Valley aquifer.

with observations 49 and 50 eliminated entirely. With observations 49 and 50 removed for estimation purposes only, the REML estimates were $\widehat{\boldsymbol{\theta}} = [19.64, 1154, 0.1261]'$ and $\widehat{\boldsymbol{\beta}} = [2248, -0.6991, -3.005]'$. All results, for the entire region, and calculated separately by subregion, are summarized in Tables 2.8 and 2.9.

The results show no indication of an overall problem with the prediction properties of the model fitted to all 93 observations. However, considering the results for the five regions, there is some suggestion that the model performs poorly in Region I ($\widehat{p}_{LR} = 0.0541$) and Region III ($\widehat{p}_{LR} = 0.9557$). Low values of relatively low T_{PR}, as obtained in Region III, suggest that

the estimated mean square errors of prediction obtained from the model may be too high for predicted values within that region.

Table 2.7 All observations included during estimation of θ and β

Region	T_{PR}	$\widehat{p}_{LR}$
Overall	92.99	0.4741
I	34.10	0.0541
II	18.41	0.3548
III	11.53	0.9557
IV	20.69	0.2070
V	8.27	0.9370

Table 2.8 Observations 49 and 50 omitted during estimation of θ and β, but included during prediction.

Region	T_{PR}	$\widehat{p}_{LR}$
Overall	117.32	0.0826
I	44.27	0.0090
II	21.83	0.2286
III	14.43	0.8706
IV	26.22	0.0693
V	10.56	0.8384

Table 2.9 Observations 49 and 50 eliminated.

Region	T_{PR}	$\widehat{p}_{LR}$
Overall	84.23	0.6362
I	11.96	0.7748
II	21.84	0.2285
III	14.04	0.8844
IV	25.92	0.0740
V	10.48	0.8427

Tables 2.7 – 2.9 present quantitative summaries of regional differences in the predicted surfaces depending on whether the two influential observations, 49 and 50, are eliminated. When observations 49 and 50 are eliminated for purposes of estimating $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, yet included during prediction, the results of Table 2.8 indicate that predictions were generally poorer. In particular, a large value of T_{PR} was observed in Region I ($T_{PR} = 44.27$; $\widehat{p}_{LR} = 0.0090$). Eliminating these two observations entirely leads to the most acceptable model from a cross-validation standpoint (with $0.05 \leq \widehat{p}_{LR} \leq 0.95$ for all regions). However, a decision regarding whether to omit these two points should not be based on examination of these and other diagnostic statistics alone.

Discussion and Concluding Remarks

The preceding example shows the utility of the statistic T_{PR}, and knowledge of an approximation to its distribution. Subdividing the region of interest is a useful technique for identifying regions of poor prediction under a given model, provided a reasonable number of observations are available. While T_{PR} for the overall region may show no reason to suspect poor predictive capacity, or under- or over-estimation of prediction errors, it is possible that the chosen model performs poorly in one or more subregion.

Solow (1990) argues that cross-validation statistics such as T_{PR} indicate only how well the proposed model for the covariance structure corresponds to the observed covariance estimates, rather than to the true covariance of the process. However, by considering distributional theory, it is possible to identify feasible values of T_{PR} under any hypothesized model. Note that examination of the approximate p-values corresponding to a particular hypothesized model amounts to a significance test for that model, since a particular alternative to the null model is not specified. Additional characteristics of T_{PR}, including its power to detect certain misspecifications of the covariance structure, are determined and discussed in a subsequent study.

It is important to note that both large and small values of T_{PR} are indicative of undesirable prediction properties in the model. This is in contrast to the usual unstandardized version of the PRESS statistic encountered in regression analysis, for which large values correspond to poor predictive models. Because T_{PR} is based on standardized PRESS residuals, extremely small values of the statistic suggest that the variance of the process has been over-estimated. This in turn indicates that the mean-squared prediction error estimates associated with such a model are inflated.

The modified PRESS statistic, T_{PR} provides an easily calculated statistic which may be calibrated using the saddlepoint approximation to its distribution. However, extreme values of T_{PR}, while suggesting that a model is inadequate, provide no guidelines for selection of a model which would be expected to have improved prediction properties. Methods for detecting more specific failures of the covariance model, using T_{PR} and additional statistics based on the modified PRESS residuals, are discussed in a subsequent study.

Bibliography

Allen, D. M. 1971. The prediction sum of squares as a criterion for selecting predictor variables. Technical Report No. 23, Department of Statistics University of Kentucky, Lexington, KY.

Carr, J. R. and Roberts, K. P. 1989. Application of universal kriging for estimation of earthquake ground motion : Statistical significance of results. Mathematical Geology 21 : 255 – 265.

Christensen, R. 1990. Linear Models for Multivariate, Time Series, and Spatial Data. Springer-Verlag, New York.

Christensen, R., Johnson, W. and Pearson, L. M. 1992. Prediction diagnostics for spatial linear models. Biometrika 79(3) : 583 – 591.

Christensen, R., Johnson, W. and Pearson, L. M. 1993. Covariance function diagnostics for spatial linear models. Mathematical Geology 25(2) : 145 – 160.

Cook, R. D. and Weisberg, S. 1982. Residuals and Influence in Regression. Chapman and Hall, New York.

Cressie, N. 1991. Statistics for Spatial Data. J. Wiley and Sons, New York.

Daniels, H. E. 1954. Saddlepoint approximations in statistics. Annals of Mathematical Statistics 25 : 631 – 649.

Daniels, H. E. 1987. Tail probability approximations. International Statistical Review 55 : 37 – 48.

Davis, B. M. 1987. Uses and abuses of cross-validation in geostatistics. Mathematical Geology 19 : 241 – 248.

Gnedenko, B. 1968. The Theory of Probability. Chelsea Publishing Co., New York.

Gordon, N. H. and Ramig, P. F. 1983. Cumulative distribution function of the sum of correlated chi-squared random variables. Journal of Statistical Computing and Simulation 17 : 1 – 9.

Isaaks, E. H. and Srivastava, R. M. 1989. Applied Geostatistics. Oxford University Press, New York.

Jones, R. H. 1989. Fitting a stochastic partial differential equation to aquifier data. Stochastic Hydrology and Hydraulics 3 : 85 – 96.

Journel, A. G., and Huijbregts, C. J. 1978. Mining Geostatistics. Academic Press, New York.

Krishnaiah, P. R., Hagis, P. and Steinberg, L. 1963. A note on the bivariate Chi-squared distribution. SIAM Review 5 : 140 – 144.

Lugannani, R. and Rice, R. O. 1980. Saddlepoint approximation for the distribution of the sum of independent random variables. Advances in Applied Probability 12 : 475 – 490.

Lukacs, E. and Laha, R. G. 1964. Applications of the Characteristic Functions. Griffin's Statistical Monograph No. 14, Griffin, London.

Mathai, A. M. and Provost, S. B. 1992. Quadratic Forms in Random Variables : Theory and Applications. Marcel Dekker, New York.

The Math Works, Inc. 1992. MATLAB Reference Guide. The Math Works, Inc., Massachusetts.

Reid, N. 1988. Saddlepoint methods and statistical inference. Statistical Science 3 : 213 – 238.

Robinson, J. 1982. Saddlepoint approximation for permutation tests and confidence intervals. Journal of the Royal Statistical Society, Series B 44 : 91 – 101.

Solomon, H. and Stephens, M. A. 1977. Distribution of a sum of weighted chi-square variables. Journal of the American Statistical Association 72 : 881 – 885.

Solow, A.R. 1990. Geostatistical cross-validation : A cautionary note. Mathematical Geology 22 : 637 – 639.

BEHAVIOR OF RESIDUALS UNDER MISSPECIFICATION OF COVARIANCE STRUCTURE IN SPATIAL LINEAR MODELS

Caryn M. Thompson

Abstract

Simulation studies were conducted to investigate the behavior of the standardized PRESS residuals under various misspecifications of the covariance matrix, $\boldsymbol{V}$. It is observed that misspecification of $\boldsymbol{V}$ tends to produce variability among the standardized PRESS residuals greater or less than would be expected if the covariance structure was known. The nature of the misspecification determines whether variability among the standardized PRESS residuals tends to be higher or lower than expected under correct specification of $\boldsymbol{V}$. Based on this observation, an adjustment to the normal probability plots commonly used to examine the standardized PRESS residuals in the geostatistical setting is proposed.

Introduction

Exploratory data analysis, and the use of statistical and graphical diagnostics, is an important component of any statistical investigation, including the analysis of spatial data. Bradley and Haslett (1992) list several objectives which may be partially achieved through the use of diagnostic techniques in the geostatistical setting. These include assessing the ability of the model to provide adequate predictions over the entire region of interest, identifying potential outliers and influential observations, and more general exploration of the data to detect potentially anomalous features.

A number of diagnostic techniques, most based on residuals obtained through cross-validation, are available to meet the first two of these objectives. Standardized PRESS residuals give an idea of the fitted covariance function or variogram model's predictive capacity at any given spatial location, and summary statistics based on these residuals may be used to calibrate

various candidate models (Thompson and Ramsey, 1995). Christensen *et al* (1992, 1993) propose case-deletion diagnostics, analogous to DFFITS, DFBETAS, and Cook's D in the regression setting, to detect influential observations in spatially referenced data. Bradley and Haslett (1992) also develop statistics which may highlight influential observations, and locations at which prediction under the proposed model is poor.

In this paper, we focus on the use of diagnostics based on the standardized PRESS residuals to detect particular misspecifications of the covariance structure of spatial linear models. Simulation studies are used to investigate the behavior of various summaries of the standardized PRESS residuals when one or more of the parameters of the covariance function have been over- or underestimated in the ordinary kriging model. The power of tests based on the statistic T_{PR} (Thompson and Ramsey, 1995) to detect misspecifications of particular parameters of the covariance function model is examined. Useful modifications to normal probability plots, which may act as aids in distinguishing among candidate models for the covariance function, are also proposed.

The Ordinary Kriging Model

Suppose the data consist of measurements corresponding to a random variable, Y, obtained at n spatial locations $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, where $\boldsymbol{u}_i$ is a vector containing the coordinates for the i^{th} location. Let the vector $\boldsymbol{Y} = [y(\boldsymbol{u}_1), y(\boldsymbol{u}_2), \ldots, y(\boldsymbol{u}_n)]'$ contain the observations for the n locations. Under the ordinary kriging model,

$$\boldsymbol{Y} = 1\beta + \boldsymbol{\epsilon}\,,$$

where $\mathrm{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$, $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{V}$, with elements $v_{ij} = C(\boldsymbol{u}_i, \boldsymbol{u}_j)$, $\boldsymbol{1}$ is an $n \times 1$ vector of 1's, and $\beta = \mathrm{E}[Y(\boldsymbol{u})]$, for all $\boldsymbol{u} \in D$, where D is the domain of interest. The best linear unbiased predictor (BLUP) of $y(\boldsymbol{u}_0)$, the value of Y at any location $\boldsymbol{u}_0$ within D, is given by

$$\hat{y}(\boldsymbol{u}_0) = \boldsymbol{\lambda}'\boldsymbol{Y} ,$$

with

$$\boldsymbol{\lambda} = \left[\frac{(1 - \boldsymbol{V}_0'\boldsymbol{V}^{-1}\mathbf{1})}{\mathbf{1}'\boldsymbol{V}^{-1}\mathbf{1}} \mathbf{1}' + \boldsymbol{V}_0' \right] \boldsymbol{V}^{-1},$$

where $\boldsymbol{V}_0' = [C(\boldsymbol{u}_0,\boldsymbol{u}_1), \ldots, C(\boldsymbol{u}_0,\boldsymbol{u}_n)]$ (Christensen, 1990). In terms of the parameter β, the BLUP of $y(\boldsymbol{u}_0)$ is given by

$$\hat{y}(\boldsymbol{u}_0) = \hat{\beta} + \boldsymbol{V}_0'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \mathbf{1}\hat{\beta}) ,$$

where the generalized least squares estimator of β is

$$\hat{\beta} = \frac{\mathbf{1}'\boldsymbol{V}^{-1}\boldsymbol{Y}}{(\mathbf{1}'\boldsymbol{V}^{-1}\mathbf{1})} .$$

The mean square prediction error at location $\boldsymbol{u}_0$ is given by

$$\sigma_k^2(\boldsymbol{u}_0) = v_{ii} - \boldsymbol{\lambda}'\boldsymbol{V}_0 + \frac{(1 - \mathbf{1}'\boldsymbol{V}^{-1}\boldsymbol{V}_0)}{\mathbf{1}'\boldsymbol{V}^{-1}\mathbf{1}} .$$

In practice, the covariance matrix $\boldsymbol{V}$ is not known, and must be estimated from the available data. Typically, a parametric form is assumed for the covariance function, and the parameters of this function are estimated using procedures such as weighted least squares, or restricted maximum likelihood (REML). Misspecification of $\boldsymbol{V}$ may affect both the resulting predictions, and the mean square errors of the predicted values. As one step in determining whether $\boldsymbol{V}$ has been adequately specified, it is common to examine cross-validated residuals from the resulting kriging model.

The i^{th} prediction error $e_{-i}(\boldsymbol{u}_i)$ is equal to $y(\boldsymbol{u}_i) - \hat{y}_{-i}(\boldsymbol{u}_i)$, where $\hat{y}_{-i}(\boldsymbol{u}_i)$ is the BLUP of $y(\boldsymbol{u}_i)$ obtained when the ith observation is removed from the data, and the remaining $n-1$ observations used to estimate β. The vectors $\boldsymbol{Y}_{-i}$ and $\mathbf{1}_{-i}$ are defined to be $\boldsymbol{Y}$ and $\mathbf{1}$ with the

i^{th} row removed, while $\boldsymbol{V}_{-i}$ is the covariance matrix for $\boldsymbol{Y}_{-i}$. For the ordinary kriging model,

$$e_{-i}(\boldsymbol{u}_i) = (\tilde{y}_i(\boldsymbol{u}_i) - \tilde{\boldsymbol{x}}_i\,\widehat{\boldsymbol{\beta}})\,\frac{s_i}{s_i - \tilde{h}_i}\,,$$

where $\tilde{y}_i(\boldsymbol{u}_i) = y_i - \boldsymbol{Y}_{-i}\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_i$, $\tilde{\boldsymbol{x}}_i = 1 - \mathbf{1}'_{-i}\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_i$, $s_i = v_{ii} - \boldsymbol{v}_i'\boldsymbol{V}_{-i}^{-1}\boldsymbol{v}_i$, $\tilde{h}_i = \tilde{\boldsymbol{x}}_i^2/(1'\boldsymbol{V}^{-1}1)$ and $\boldsymbol{v}_i$ is the i^{th} column of $\boldsymbol{V}$ with the element v_{ii} deleted. Details for calculation of the inverse matrices $\boldsymbol{V}_{-i}^{-1}$ directly from the elements of $\boldsymbol{V}^{-1}$ are given by Christensen *et al* (1992).

These cross-validated prediction errors are sometimes known as PRESS residuals. For the ordinary kriging model, standardized versions of these PRESS residuals are given by

$$t_i = \frac{e_{-i}(\boldsymbol{u}_i)}{\sigma_{e_{-i}}} = \frac{(\tilde{y}_i(\boldsymbol{u}_i) - \tilde{\boldsymbol{x}}_i\,\widehat{\boldsymbol{\beta}})}{\sqrt{s_i - \tilde{h}_i}}\,.$$

Thompson and Ramsey (1995) discuss distributional properties of these standardized PRESS residuals. In particular, under the additional assumption that $\epsilon \sim N(\mathbf{0}, \boldsymbol{V})$, each of the t_i's has a marginal standard normal distribution. However, the t_i's are not independent. If $\boldsymbol{V}$ is known, then for $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]'$,

$$\mathrm{Cov}(\boldsymbol{t}) = \boldsymbol{\Sigma} = \boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\,,$$

where the i^{th} row of $\boldsymbol{K}'$ is given by $\boldsymbol{k}_i' = [(\boldsymbol{a}_i' - \boldsymbol{v}_i'\boldsymbol{V}_{-i}^{-1}\boldsymbol{B}_i) - \tilde{\boldsymbol{x}}_i(1'\boldsymbol{V}^{-1}1)^{-1}1'\boldsymbol{V}^{-1}]/(\sqrt{s_i - \tilde{h}_i})$, $\boldsymbol{a}_i$ is the i^{th} column of an $n \times n$ identity matrix, and $\boldsymbol{B}_i$ is an $(n-1) \times n$ matrix consisting of an $n \times n$ identity matrix with the i^{th} row removed.

Behavior of the Standardized PRESS Residuals Under Misspecifications of $\boldsymbol{V}$

As stated previously, in practice $\boldsymbol{V}$ is not known, and must be estimated from the data. Denote the estimated covariance matrix as $\boldsymbol{V}^*$. The matrix $\boldsymbol{V}^*$ is then used to calculate the standardized PRESS residuals, contained in the vector $\boldsymbol{t}$, and $\mathrm{Cov}(\boldsymbol{t}) = \boldsymbol{K}^{*\prime}\boldsymbol{V}\boldsymbol{K}^*$, where $\boldsymbol{k}_i^{*\prime} = [(\boldsymbol{a}_i' - \boldsymbol{v}_i'\boldsymbol{V}^{*-1}_{-i}\boldsymbol{B}_i) - \tilde{\boldsymbol{x}}_i(1'\boldsymbol{V}^{*-1}1)^{-1}1'\boldsymbol{V}^{*-1}]/(\sqrt{s_i - \tilde{h}_i})$. Using the estimated covariance

$\boldsymbol{V}^*$, $\mu_t = \mathrm{E}(t_i) = \mathrm{E}[(\tilde{y}(\boldsymbol{u}_i) - \tilde{x}_i\widehat{\beta})/(\sqrt{s_i - \widetilde{h}_i})] = 0$, for $i = 1, ..., n$. However, the dispersion of the standardized PRESS residuals is affected by misspecification of $\boldsymbol{V}$.

Consider the statistic

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{t}\,)^2$$

$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n}(t_i - \mu_t)^2 - n(\bar{t}\,-\mu_t)^2\right\}$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} t_i^2 - n\bar{t}^{\,2}\right)$$

as a measure of variability among the t_i's. Thompson and Ramsey (1995) discuss distributional properties of the related statistic, $T_{PR} = \boldsymbol{t}'\boldsymbol{t} = \sum_{i=1}^{n} t_i^2$, and the calculation of approximate "p-values" associated with this statistic. Extreme values of T_{PR}, and also of s^2, may indicate that $\boldsymbol{V}$ has been misspecified, but provide no information as to how the covariance model might be improved.

However, investigation of the dispersion of the standardized PRESS residuals under particular misspecifications of $\boldsymbol{V}$ suggests a link between the amount of variability among the t_i's, and the type of misspecification of the covariance matrix which has occurred. If $\boldsymbol{V}^* = \boldsymbol{V}$, that is, if the covariance matrix is correctly specified, the expectation of s^2 is given by

$$\mathrm{E}(s^2) = \frac{1}{n-1}\left\{\mathrm{E}(\boldsymbol{t}'\boldsymbol{t}) - \frac{1}{n}\,\mathrm{E}(\boldsymbol{t}'\boldsymbol{t})\right\}$$

$$= \frac{1}{n}\,\mathrm{E}(\boldsymbol{t}'\boldsymbol{t}) = \frac{1}{n}\,tr(\Sigma),$$

where tr represents the trace, or sum of the diagonal elements of Σ. Also, if $\boldsymbol{V}$ is correctly specified, then $t_i \sim N(0,1)$, and therefore $tr(\Sigma) = n$. Otherwise, if $\boldsymbol{V}^* \neq \boldsymbol{V}$, the covariance matrix is said to be misspecified, and $tr(\Sigma) \neq n$. Therefore, $\mathrm{E}(s^2) = 1$ when $\boldsymbol{V}^* = \boldsymbol{V}$, but will be greater than or less than 1 for $\boldsymbol{V}^* \neq \boldsymbol{V}$, depending on the nature of the misspecification. This

suggests that for $\boldsymbol{V}^* \neq \boldsymbol{V}$, variability among the standardized PRESS residuals may be either greater or less than would be expected if $\boldsymbol{V}$ was correctly specified.

Consider, for example, an exponential covariance model of the form

$$C(l) = \begin{cases} \theta_1 exp(-\theta_2 l) & \text{for } l > 0 \\ \theta_0 + \theta_1 & \text{for } l = 0 \end{cases}, \tag{3.1}$$

where $l = \| \boldsymbol{u}_i - \boldsymbol{u}_j \|$. The parameters θ_0, θ_1, and θ_2 are estimated from the available data to obtain the estimated covariance function $C^*(l)$, which has the same form as $C(l)$ above, but with θ_0, θ_1, and θ_2 replaced by their respective estimates. The elements of the estimated covariance matrix $\boldsymbol{V}^*$ are then given by $v^*_{ij} = C^*(l)$. Over- or under-estimation of any one of the three parameters θ_0, θ_1, and θ_2 leads to a misspecification of $\boldsymbol{V}$. A simulation study was conducted to assess the effect of these and other forms of misspecification of $\boldsymbol{V}$ on the behavior of the standardized PRESS residuals.

Description of the Simulation Studies

One thousand independent realizations were generated on 8×8 square grids for a zero-mean Gaussian process with an exponential covariance function model and parameters $\theta_0 = 0.2$, $\theta_1 = 1.0$ and $\theta_2 = 0.5$, using the spectral decomposition method (Cressie, 1991) . Thus, the elements of $\boldsymbol{V}$ are given by $v_{ij} = C(l)$, for $\theta_0 = 0.2$, $\theta_1 = 1.0$ and $\theta_2 = 0.5$. Cross-validation was

Table 3.1 Parameters of exponential models involved in the first simulation study.

	Specified model for $C(l)$						
	I	II	III	IV	V	VI	VII
θ_0	0.2	0.1	0.3	0.2	0.2	0.2	0.2
θ_1	1.0	1.0	1.0	0.75	1.25	1.0	1.0
θ_2	0.5	0.5	0.5	0.5	0.5	0.3	0.7

then performed for each realization, specifying several different models of exponential form as V^*, that is, assuming alternative values for θ_0, θ_1, and θ_2, but retaining the exponential model form given in (3.1) for the covariance structure. The respective parameter values for the seven models for V^* are presented in Table 3.1. Model I corresponds to the correctly specified model, while each of Models II – VII represents either over- or under-estimation of a single parameter of the covariance function. Summaries, including the mean, s^2, minimum, maximum and range of the standardized PRESS residuals, and the statistic T_{PR} were recorded for each cross-validated model. The results of the study are given in Table 3.2 and Table 3.3.

The contents of Table 3.2 represent average values for each of the summary statistics, obtained over the 1000 simulated realizations. As such, these mean values provide simulated estimates of the expected values for each summary statistic. Several consistent trends are evident in these results. As expected, for all specifications of V^* considered in the study, $\widehat{\mathrm{E}}(\overline{t}\,) = \frac{1}{1000}\sum_{i=1}^{1000}\overline{t}$ is approximately zero. When $V^* = V$, that is, when Model I is selected for V^*, $\widehat{\mathrm{E}}(s^2)$ is approximately equal to one, and $\widehat{\mathrm{E}}(T_{PR})$ is approximately equal to n ($= 64$, since each realization consists of 64 values generated on an 8×8 square grid). However, as suggested previously, when $V^* \neq V$, $\widehat{\mathrm{E}}(s^2) \neq 1$, while $\widehat{\mathrm{E}}(T_{PR}) \neq n$. Specifically, over-estimation of any of the three parameters of the covariance function (as occurs in Models III, V, and VII) leads to

Table 3.2 Average value of summary statistics calculated from 1000 realizations generated from exponential Model I.

	Specified model for $C(l)$						
	I	II	III	IV	V	VI	VII
$\overline{t}$	0.0001	0.0001	0.0001	0.0002	0.0001	0.0002	0.0001
s^2	1.0145	1.2645	0.8662	1.2168	0.8852	1.3341	0.8547
min(t)	– 2.367	– 2.629	– 2.179	– 2.577	– 2.201	– 2.703	– 2.164
max(t)	2.369	2.629	2.183	2.566	2.202	2.706	2.167
range(t)	4.736	5.258	4.362	5.143	4.303	5.409	4.331
T_{PR}	64.480	79.674	54.572	76.661	55.768	84.056	53.848

decreased variability among the standardized PRESS residuals, as reflected in the values of s^2, the range and T_{PR}, which all decrease in their estimated expected values. In contrast, under-estimation of θ_0, θ_1, or θ_2 leads to increased variability among the t_i's obtained for each surface, as indicated by the fact that the estimated expected values for the statistics s^2, range($\boldsymbol{t}$) and T_{PR} were greater for Models II, IV and VI than for the reference model, Model I.

The results of Table 3.3 may be used to make statements about the power of the statistic T_{PR} to detect particular misspecifications of $\boldsymbol{V}$. Note that tests based on T_{PR} amount to significance testing, in that a particular alternative to the covariance function $\boldsymbol{V}^*$ is not specified. As such, extremely large or small values of T_{PR}, as calibrated by the corresponding p-values, suggest that $\boldsymbol{V}^* \neq \boldsymbol{V}$, but provide no indication as to the structure of $\boldsymbol{V}$ (Thompson and Ramsey, 1995).

Table 3.3 Proportion of realizations generated from exponential Model I showing $\widehat{p}$ values within a particular range, for various specifications of $C(l)$.

	Specified model for $C(l)$						
	I	II	III	IV	V	VI	VII
$\widehat{p} \leq 0.01$	0.013	0.125	0.000	0.091	0.001	0.202	0.000
$\widehat{p} \leq 0.05$	0.053	0.312	0.005	0.246	0.009	0.443	0.009
$\widehat{p} \leq 0.10$	0.098	0.459	0.015	0.390	0.020	0.576	0.019
$\widehat{p} \leq 0.90$	0.902	0.994	0.695	0.990	0.737	0.998	0.727
$\widehat{p} \leq 0.95$	0.962	0.999	0.807	0.997	0.835	1.000	0.824
$\widehat{p} \leq 0.99$	0.996	1.000	0.955	1.000	0.961	1.000	0.959

The $\widehat{p}$-values given in Table 3.3 represent estimated upper tail probabilities, so that small values of $\widehat{p}$ correspond to large values of T_{PR}, while large values of $\widehat{p}$ correspond to small values of T_{PR}. The results indicate that T_{PR} has limited power to detect over- or under-estimation of the parameters of the exponential covariance function model. For instance, Models II, IV, and VI correspond to under-estimation of the parameters θ_0, θ_1, and θ_2

respectively, as indicated in Table 3.1. As discussed above, under-estimation of any one of these parameters tends to lead to large values of T_{PR}, and therefore small values of $\hat{p}$. For Models II, IV, and VI, values of $\hat{p} \leq 0.1$, corresponding to an upper tail probability of less than or equal to 10 %, were observed for 45.9 %, 39.0 % and 57.6 %, respectively, of the realizations. For over-estimation of θ_0, θ_1, or θ_2, as defined by Models III, V, and VII, values of $\hat{p} \geq 0.9$, corresponding to a lower tail probability of less than or equal to 10 %, were observed for 30.5 %, 26.3 %, and 27.3 % of the simulated realizations.

A second study was conducted, generating 1000 realizations on an 8×8 grid, from a Gaussian process with a Gaussian covariance model of the form

$$C(l) = \begin{cases} \theta_1 exp(-(\theta_2 l))^2 & \text{for } l > 0 \\ \theta_0 + \theta_1 & \text{for } l = 0 \end{cases},$$

with parameters $\theta_0 = 0.2$, $\theta_1 = 1.0$, and $\theta_2 = 0.8$. As in the previous study, six additional models for $\boldsymbol{V}^*$ were selected, with each representing over- or underestimation of one of the parameters θ_0, θ_1, or θ_2. The parameters of the models for $\boldsymbol{V}^*$ are summarized in Table 3.4.

Table 3.4 Parameters of Gaussian models involved in the second simulation study.

	Specified model for $C(l)$						
	I	II	III	IV	V	VI	VII
θ_0	0.2	0.1	0.3	0.2	0.2	0.2	0.2
θ_1	1.0	1.0	1.0	0.75	1.25	1.0	1.0
θ_2	0.8	0.8	0.8	0.8	0.8	0.6	1.0

The results of Table 3.5 show similar trends to the results for the exponential model (Table 3.2). In all cases, the estimated expected value of $\overline{t}$ is approximately equal to zero. For the models where one of the three parameters (θ_0, θ_1, θ_2) has been over-estimated (Models III, V, and VII), decreased variability among the standardized PRESS residuals is evident, meaning

average values for s^2, range($\boldsymbol{t}$), and T_{PR} are lower than for the correctly specified model (Model I). In contrast, for those models representing under-estimation of one of the three parameters (Models II, IV, and VI), increased variability in the t_i's was observed. That is, the estimated expected values for s^2, range($\boldsymbol{t}$), and T_{PR} were greater for Models II, IV and VI than for the correctly specified model.

Table 3.5 Average value of summary statistics calculated from 1000 realizations generated from Gaussian Model I.

	Specified model for $C(l)$						
	I	II	III	IV	V	VI	VII
$\overline{t}$	0.0002	0.0002	0.0001	0.0002	0.0002	0.0002	0.0001
s^2	1.024	1.452	0.830	1.177	0.918	1.599	0.749
min($\boldsymbol{t}$)	−2.358	−2.788	−2.129	−2.533	−2.229	−2.942	−2.030
max($\boldsymbol{t}$)	2.355	2.790	2.128	2.531	2.226	2.937	2.033
range($\boldsymbol{t}$)	4.713	5.578	4.256	5.064	4.456	5.879	4.063
T_{PR}	64.513	91.492	52.300	74.147	57.834	100.75	47.176

Table 3.6 Proportion of realizations generated from Gaussian Model I showing $\widehat{p}$ values within a particular range, for various specifications of $C(l)$.

	Specified model for $C(l)$						
	I	II	III	IV	V	VI	VII
$\widehat{p} \leq 0.01$	0.011	0.293	0.000	0.054	0.001	0.483	0.000
$\widehat{p} \leq 0.05$	0.056	0.508	0.002	0.158	0.027	0.704	0.000
$\widehat{p} \leq 0.10$	0.101	0.631	0.011	0.274	0.041	0.793	0.000
$\widehat{p} \leq 0.90$	0.905	0.993	0.658	0.985	0.799	1.000	0.479
$\widehat{p} \leq 0.95$	0.959	0.998	0.782	0.994	0.883	1.000	0.655
$\widehat{p} \leq 0.99$	0.995	1.000	0.936	1.000	0.968	1.000	0.880

Table 3.6 gives an indication of the power of the statistic T_{PR} to detect misspecifications of $\boldsymbol{V}$ when the covariance matrix is known to have Gaussian form. As

observed for the exponential model (Table 3.3), T_{PR} has limited power to detect over- or under-estimation of the parameter θ_1. Where θ_1 was under-estimated (Model IV), estimated upper tail probabilities, $\hat{p}$, less than 0.1 were observed for only 27.4 % of the simulated realizations. Over-estimation of θ_1 (Model V) led to $\hat{p} \geq 0.9$ for only 20.1 % of the simulated realizations. However, somewhat higher levels of power were apparent for misspecification of the parameter θ_0. Where θ_0 was under-estimated (Model II), $\hat{p}$ values ≤ 0.1 were observed for 63.1 % of the realizations, while over-estimation of θ_0 led to $\hat{p}$ values ≥ 0.9 in 34.2 % of the simulated realizations. The statistic T_{PR} was most sensitive to over- or under-estimation of the parameter θ_2. Under-estimation of θ_2 (Model VI) produced $\hat{p}$ values ≤ 0.1 for 79.3 % of the realizations generated through the simulation study, while $\hat{p}$-values ≥ 0.9 were observed for 52.1% of the realizations when θ_2 was over-estimated (Model VII).

The results presented in Tables 3.2, 3.3, 3.5, and 3.6 are consistent with the findings of Warnes (1986). In a sensitivity analysis of universal kriging, Warnes found that, for the exponential model, substantial changes in the parameter θ_2 have little effect on the predictions. However, for the Gaussian model, small changes to θ_2 result in significant perturbations of the predicted surface. This phenomenon is explained by Stein and Handcock (1989), who define compatibility classes to contain variograms (or covariance functions) showing asymptotically similar behavior in terms of the predictions and prediction error variances they produce. They showed that the predictions produced by two exponential models with different values of θ_2 will be asymptotically the same, whereas the prediction error variances resulting from the two models will be approximate multiples of one another. Hence, it is not surprising that the statistic T_{PR} has little power to identify misspecifications of the exponential covariance function. In contrast, two Gaussian models with unequal values of θ_2 cannot be compatible (Stein and Handcock, 1989), and as expected, T_{PR} had a more reasonable level of power for detecting this sort of misspecification of the covariance function. It should be noted that the

Gaussian covariance model corresponds to an infinitely differentiable random field, and therefore an unrealistically smooth surface.

A final simulation study was conducted to investigate the behavior of the standardized PRESS residuals under misspecification of the parametric form of the covariance model structure. The six combinations of specified and true models employed in this simulation study are shown in Table 3.7. Specifically, $\boldsymbol{V}$ was chosen to have elements arising from either an exponential model (3.1) or a spherical covariance function model, of the form

$$C(l) = \begin{cases} \theta_1\{1-(1.5(\theta_2 l)-0.5(\theta_2 l)^3)\} & \text{for } 0 < l < 1/\theta_2 \\ 0 & \text{for } l \geq 1/\theta_2 \\ \theta_0 + \theta_1 & \text{for } l = 0 \ . \end{cases}$$

The parameters for the models investigated in the study were chosen keeping in mind that the exponential model reaches a value equal to 95 % of its sill $(\theta_0 + \theta_1)$ at an effective range approximately equal to $3/\theta_2$ (Journel and Huijbregts, 1978). Therefore, an exponential model with $\theta_2 = 0.6$ has approximately the same range as a spherical model with $\theta_2 = 0.2$. However, the results of Stein (1988) indicate that it is the behavior of the covariance function (or variogram) near the origin which is crucial. Since both the exponential and spherical models exhibit fairly linear behavior at the origin, the first "true" model given in Table 3.7

Table 3.7 Models involved in the third simulation study. The study investigates misspecification of the parametric form of the covariance function (Exponential (E) versus Spherical (S)). In all cases, $\theta_0 = 0.0$, and $\theta_1 = 1.0$, while the value of θ_2 is varied.

	I	II	III	IV	V	VI
True model	E	E	E	S	S	S
θ_2	0.6	0.6	0.6	0.4	0.4	0.4
Specified model	S	S	S	E	E	E
θ_2	0.3	0.4	0.5	0.4	0.6	0.9

(exponential, $\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 0.6$) might be expected to be similar to a spherical model with similar nugget effect and slope at at the origin ($\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 0.4$).

Results for this study are summarized in Table 3.8 and Table 3.9. The results indicate that, as expected, an exponential covariance model with $\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 0.6$ is indistinguishable from a spherical model with $\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 0.4$ in terms of the behavior of the residuals when one of these models is specified in place of the other. However, the results for Models I and III indicate that values of $\theta \neq 0.4$ correspond to behavior in the residuals similar to that seen for over-specification (Model III) or under-specification (Model I) of the covariance structure for the exponential model (Table 3.2 and Table 3.3). In contrast, when the true model was of the spherical form, specifying an exponential model with similar nugget effect and range (model V) led to greater variability among the standardized PRESS residuals than would be expected for a correctly specified model. Again, as seen for Models I and III, specification of an exponential model with a slope at the origin greater or less than that for the true spherical model (Models IV and VI) led to behavior in the residuals similar to that observed for under- or over-specification, respectively, of the parameter θ_2 in the exponential model.

Table 3.8 Average value of summary statistics calculated from 1000 realizations generated from the combinations of true models and specified models given in Table 3.7.

	Specified/true model for $C(l)$					
	I	II	III	IV	V	VI
$\bar{t}$	0.0002	0.0001	0.0001	0.0002	0.0002	0.0001
s^2	1.3114	0.9978	0.7470	1.7201	1.2019	0.8900
min(t)	−2.674	−2.315	−2.179	−3.064	−2.563	−2.210
max(t)	2.670	2.312	2.183	3.069	2.569	2.218
range(t)	5.344	4.626	4.362	6.133	5.132	4.427
T_{PR}	82.6218	62.860	54.572	108.368	75.718	56.074

The statistic T_{PR} has more power to detect misspecifications of the type described in this simulation study (Table 3.9) than observed for the first simulation study (Table 3.3). Approximate $\widehat{p}$-values ≤ 0.1 were observed for 49.7 % of the realizations generated for the situation defined by model I. For the misspecification defined by Model III, approximate p-values $\geq$ 0.9 were observed for 57.3 % of the realizations. Power was still greater for misspecification of the spherical model. For the situation described by Model IV (specification

Table 3.9 Proportion of realizations generated from exponential model showing $\widehat{p}$ values in particular range, for the situations described in Table 3.7.

	Specified/true model for $C(l)$					
	I	II	III	IV	V	VI
$\widehat{p} \leq 0.01$	0.171	0.011	0.000	0.672	0.078	0.001
$\widehat{p} \leq 0.05$	0.359	0.050	0.001	0.849	0.214	0.007
$\widehat{p} \leq 0.10$	0.497	0.091	0.004	0.909	0.336	0.024
$\widehat{p} \leq 0.90$	0.993	0.874	0.427	1.000	0.990	0.757
$\widehat{p} \leq 0.95$	0.999	0.933	0.586	1.000	0.998	0.856
$\widehat{p} \leq 0.99$	1.000	0.985	0.821	1.000	1.000	0.967

of an exponential rather than a spherical model, and under-specification of θ_2), $\widehat{p} \leq 0.1$ were observed for 90.9 % of the realizations, while for the situation associated with Model VI (exponential rather than spherical model, and over-estimation of θ_2), $\widehat{p}$-values ≥ 0.9 were observed for 75.7 % of the realizations. The results indicate that the T_{PR} is more sensitive to misspecifications of the spherical covariance model than to similar misspecifications of the exponential model.

Normal Probability Plots as Diagnostic Tools

Summary statistics based on the standardized PRESS residuals are often used in geostatistical analysis as diagnostic tools to check for inadequacies in the covariance function model. Alternatively, stem-and-leaf plots, histograms, and normal probability plots based on

the t_i's may shed light on their distribution, and also indicate the presence of potential outliers. Cressie (1991) describes the use of normal probability plots based on the standardized PRESS residuals to detect potential outliers (that is, unusually large values of the t_i's) arguing that although the t_i's are highly dependent, each has a marginal standard normal distribution, assuming $\boldsymbol{V}$ has been correctly specified.

The results of the simulation studies described in the previous section suggest that an assessment of variability in the standardized PRESS residuals under a particular specification of the covariance matrix, $\boldsymbol{V}^*$, may provide information as to how $\boldsymbol{V}$ has been misspecified, if at all. As shown previously, for $\boldsymbol{Y} \sim N(\beta, \boldsymbol{V})$ and $\boldsymbol{V}^* = \boldsymbol{V}$, the expected value of $s^2 = 1$. Therefore, noting also that $\mathrm{E}(t_i) = 0$, for $i = 1, \ldots, n$, it seems reasonable to assume that the t_i's follow a distribution which is approximately standard normal when the covariance matrix is properly specified, despite the dependency among the t_i's. However, when $\boldsymbol{V}^* \neq \boldsymbol{V}$, the t_i's, still with an expected value of zero, would be expected to more closely follow a normal distribution with $\sigma^2 \neq 1$. Therefore, a modification to the normal probability plots is suggested, which may be used to detect particular misspecifications of $\boldsymbol{V}$.

As stated previously, for $\boldsymbol{V}^* = \boldsymbol{V}$, $\mathrm{E}(s^2) = tr(\Sigma)/n = tr(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K})/n$. However, for $\boldsymbol{V}^* \neq \boldsymbol{V}$, $\mathrm{E}(s^2) = tr(\boldsymbol{K}^{*'}\boldsymbol{V}\boldsymbol{K}^*)/n = tr(\boldsymbol{K}^*\boldsymbol{K}^{*'}\boldsymbol{V})/n$. Therefore, when $\boldsymbol{V}$ is not properly specified, the standardized PRESS residuals should at least approximately follow a normal distribution with mean 0, and variance equal to $tr(\boldsymbol{K}^*\boldsymbol{K}^{*'}\boldsymbol{V})/n$. Examination of a series of normal probability plots, each using $\boldsymbol{K}^*$ obtained from a particular $\boldsymbol{V}^*$, but varying $\boldsymbol{V}$, and thus the variance associated with the plots, may point to improved specification of $\boldsymbol{V}$.

To compare the specified covariance matrix to several candidate models for $\boldsymbol{V}$, $\boldsymbol{K}^*$ is calculated for the specified model $\boldsymbol{V}^*$, and $tr(\boldsymbol{K}^*\boldsymbol{K}^{*'}\boldsymbol{V})$ is calculated for each alternative. Each of the ordered t_i's is then adjusted by dividing by $\sqrt{tr(\boldsymbol{K}^*\boldsymbol{K}^{*'}\boldsymbol{V})/n}$ to obtain

$$r_i = \frac{t_i}{\sqrt{tr(\boldsymbol{K}^*\boldsymbol{K}^{*'}\boldsymbol{V})/\ n}} \quad .$$

The r_i's are then ordered, so that $r_{(i)}$ denotes the i^{th} smallest value from the vector $\boldsymbol{r} = [r_1, r_2, \ldots, r_n]'$, and the $r_{(i)}$'s plotted against their approximate expected values (Blom, 1958), given by

$$\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right),$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal distribution. These plots may then be used to suggest an appropriate alternative specification for $\boldsymbol{V}$. If $\boldsymbol{r}$ is calculated for each model, and the resulting plots of $r_{(i)}$ versus $\hat{E}(r_{(i)})$ overlaid, the covariance functions corresponding to plots showing a straight line appearance would be preferred. This plotting technique is an application of the animated probability plots suggested by Cook and Weisberg (1989); here, each plot corresponds to a different covariance model $\boldsymbol{V}$.

Some potential uses of the adjusted normal probability plots as diagnostic tools are illustrated through the following examples.

Example 1

The first example involves misspecification of the nugget effect parameter (θ_0) in the exponential model. Six realizations were generated from an exponential covariance model $\boldsymbol{V}$ with parameters $\theta_0 = 0.4$, $\theta_1 = 1.0$, and $\theta_2 = 0.9$. Standardized PRESS residuals for each realization were obtained through cross-validation, with the specified covariance matrix $\boldsymbol{V}^*$ having an exponential model, with parameters $\theta_0 = 0.0$, $\theta_1 = 1.0$, and $\theta_2 = 0.9$. Adjusted normal probability plots of the ordered adjusted standardized PRESS residuals, the $r_{(i)}$'s, were produced for each realization. The matrix $\boldsymbol{K}^*$ obtained from the specified covariance matrix $\boldsymbol{V}^*$, and three alternatives for $\boldsymbol{V}$, including $\boldsymbol{V}^*$, the actual covariance, and a third model, with $\theta_0 = 0.2$, $\theta_1 = 1.0$, and $\theta_2 = 0.9$ (Figure 3.1) were used to calculate appropriate adjustment

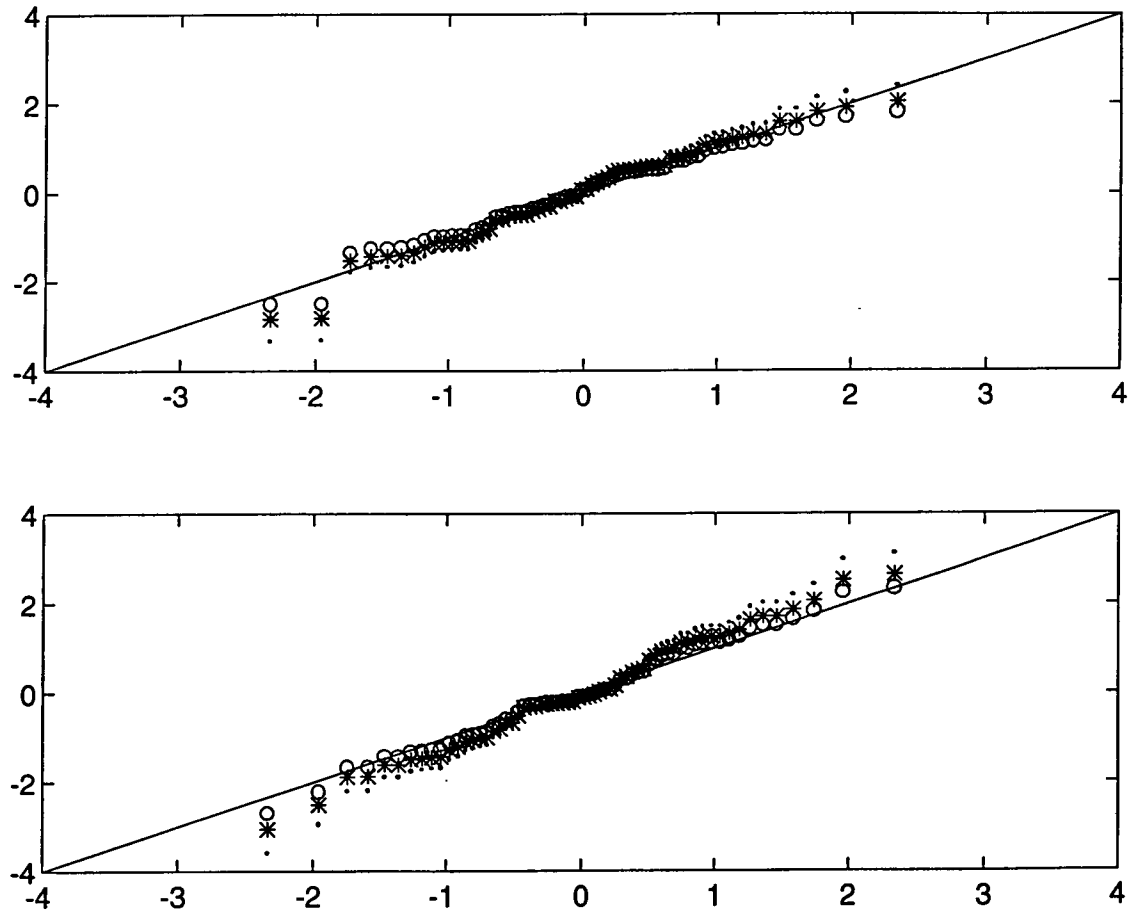


Figure 3.1 Adjusted normal probability plots for the models defined in Example 1. Each of the six individual plots (continued on following pages) represents a realization generated from the true model, corresponding to an exponential covariance function with $\theta = [0.4, 1, 0.9]'$. The three specified models also correspond to exponential covariance functions, with parameters $\theta = [\theta_0, \theta_1, \theta_2]' = [0, 1, 0.9]'$ ($\cdot\ \cdot\ \cdot\ \cdot$); $\theta = [0.4, 1, 0.9]'$ (o o o o), and $\theta = [0.2, 1, 0.9]'$ (****).

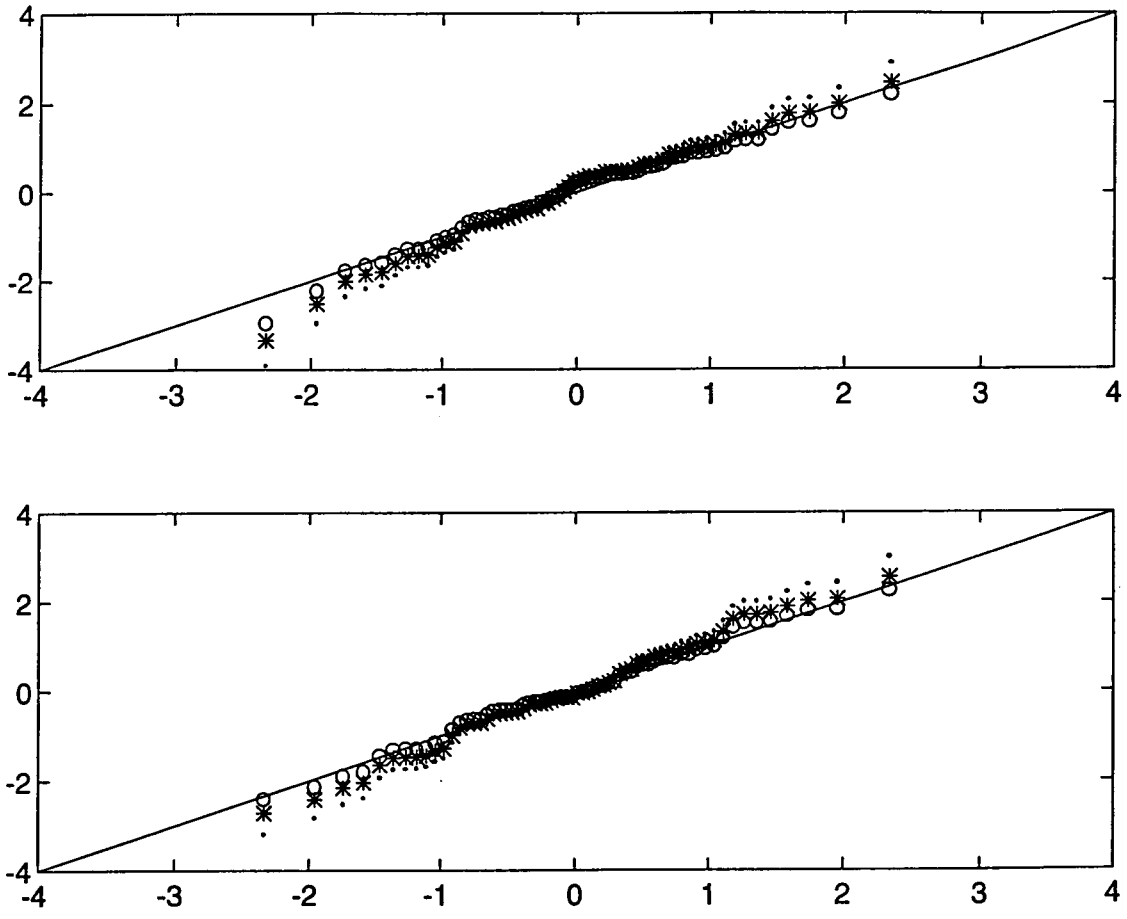


Figure 3.1 (Continued)

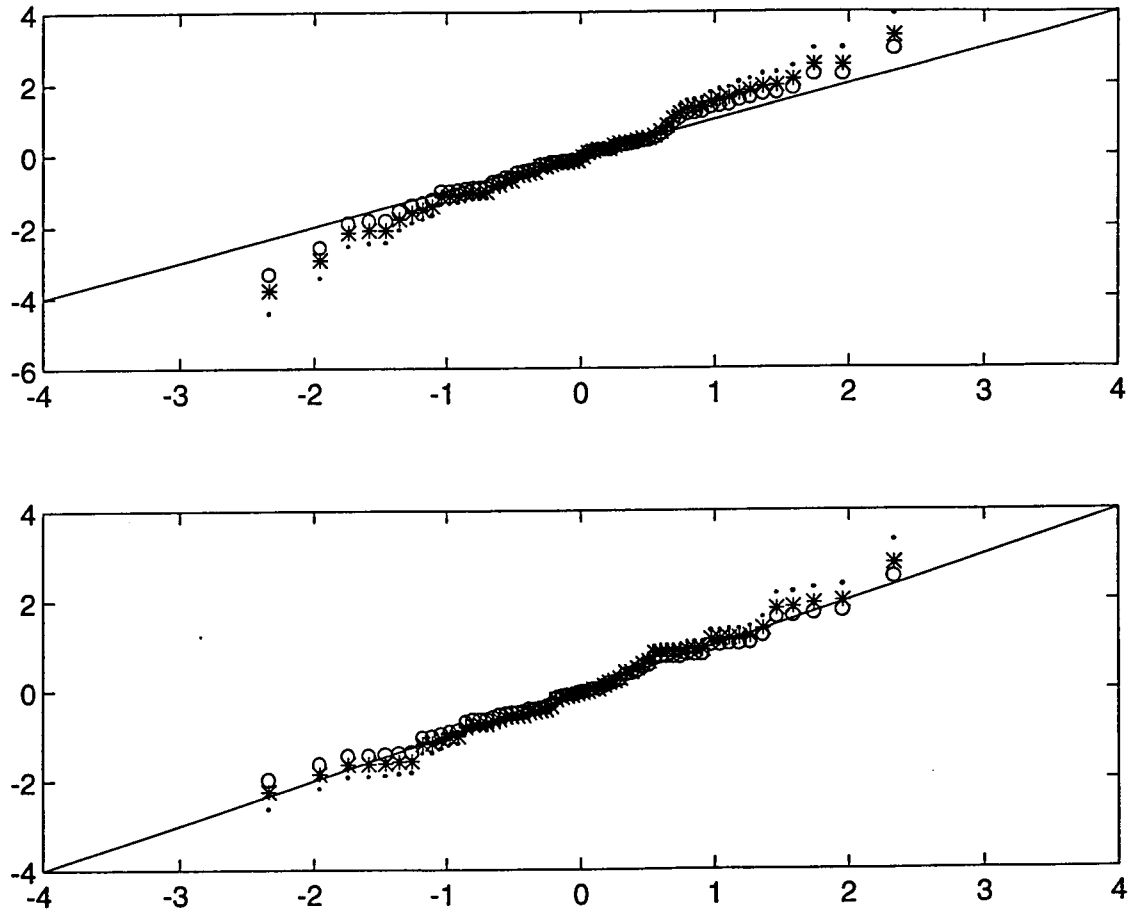


Figure 3.1 (Continued)

factors, $\sqrt{tr(\boldsymbol{K}^*\boldsymbol{K}^{*'}\boldsymbol{V})/n}$. The adjustment factor was equal to 1 for the specified model, 1.326 for the correct model, and 1.175 for the third model.

Figure 3.1 shows the adjusted probability plots for each realization. In all cases, the plots corresponding to $\boldsymbol{V}^*$ suggest that $\boldsymbol{V}$ has not been correctly specified. The slight "S-shape" of these plots indicates that the residuals are overly dispersed, and therefore that one of the parameters of the covariance function may have been under-estimated. For most realizations, the probability plots suggest that $\boldsymbol{V}$ is more likely to have a form closer to the second (correct specification) or third model. In other words, there is an indication that at least one of the three parameters has been under-estimated, and that a model with $\theta_0 = 0.2$ or 0.4 is likely to be closer to the true model than is the currently specified model ($\theta_0 = 0$).

Example 2

This example illustrates the use of the adjusted normal probability plots in detecting misspecifications of the parameter related to the range (θ_2) in Gaussian covariance models. Six realizations were generated from a Gaussian covariance matrix $\boldsymbol{V}$, with parameters $\theta_0 = 0.0$, $\theta_1 = 1.0$, and $\theta_2 = 0.8$. Standardized PRESS residuals were obtained for a covariance matrix $\boldsymbol{V}^*$, again with Gaussian form, but with parameters $\theta_0 = 0.0$, $\theta_1 = 1.0$, and $\theta_2 = 0.7$, corresponding to under-estimation of θ_2. Adjusted probability plots were then obtained for the specified model, and two alternatives, corresponding to the true model, and a third Gaussian covariance model with parameters $\theta_0 = 0.0$, $\theta_1 = 1.0$, and $\theta_2 = 0.9$ (Figure 3.2). The adjustment factors for the true model, and the third model were 1.902 and 3.063 respectively. The appearance of the plots indicates clearly that for the specified model, variability among the t_i's is greater than would be expected if the model was correctly specified, suggesting under-estimation of at least one of the models parameters. The plots also suggest that a Gaussian model with $\theta_2 = 0.8$ is likely to correspond more closely to the correct specification of $\boldsymbol{V}$ than does the current specification.

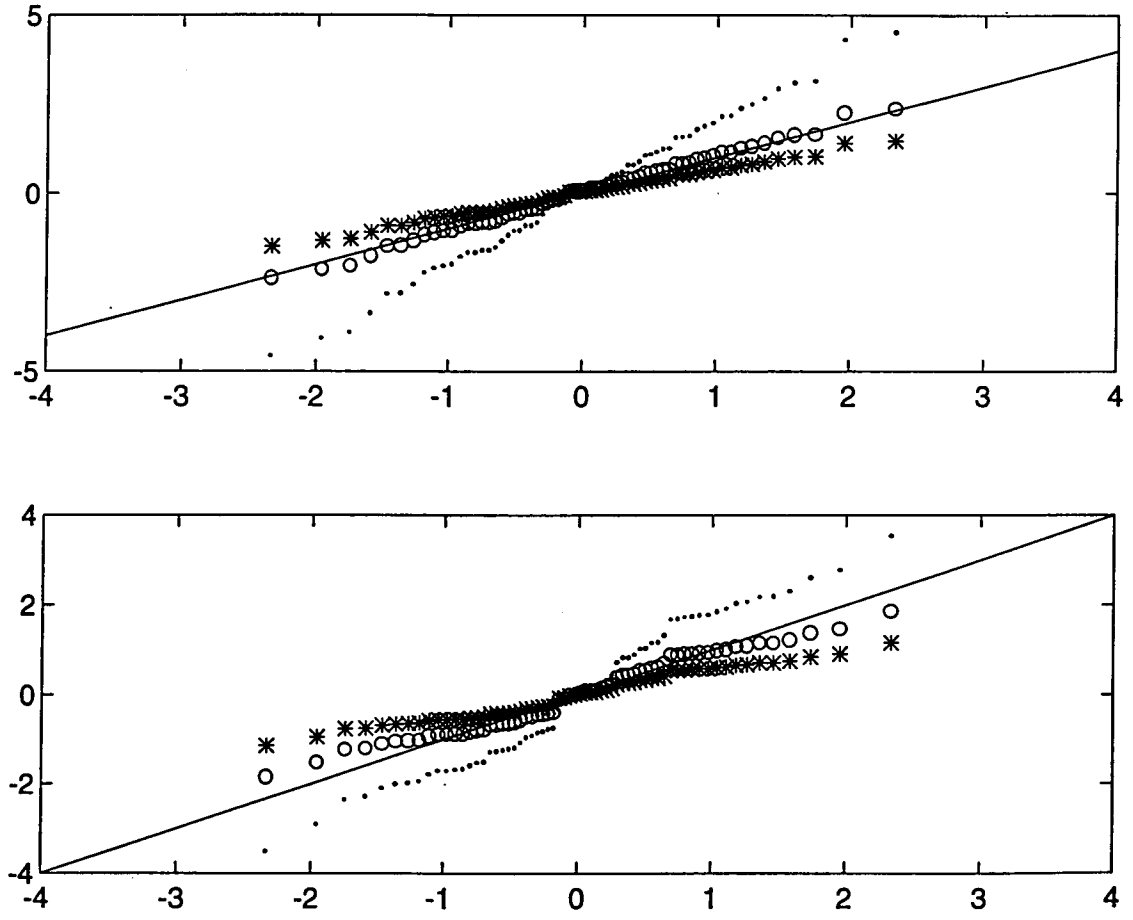


Figure 3.2 Adjusted normal probability plots for the models defined in Example 2. Each of the six individual plots (continued on following pages) represents a realization generated from the true model, corresponding to a Gaussian covariance function with $\theta = [0, 1, 0.8]'$. All three specified models also correspond to Gaussian covariance functions, with parameters $\theta = [\theta_0, \theta_1, \theta_2]' = [0, 1, 0.7]'$ (· · · ·); $\theta = [0, 1, 0.8]'$ (o o o o), and $\theta = [0.2, 1, 0.9]'$ (****).

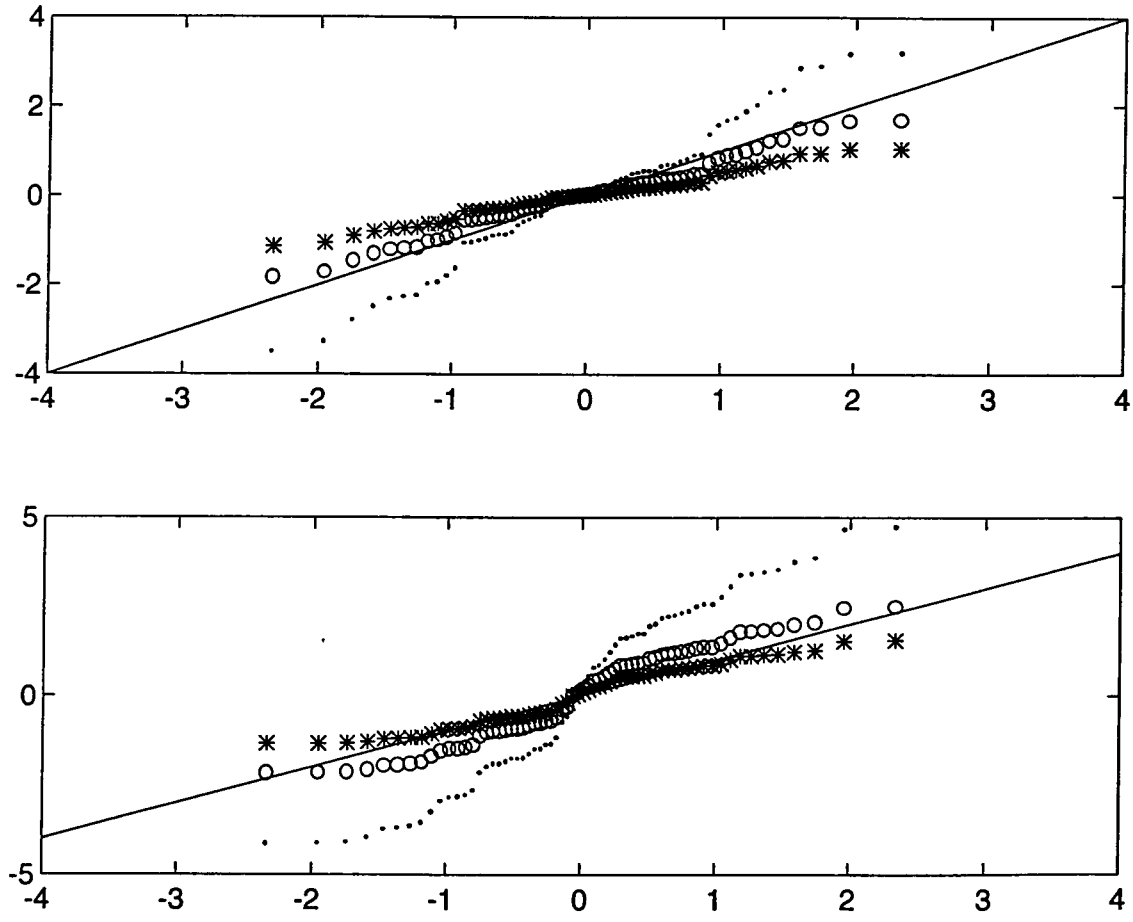


Figure 3.2 (Continued)

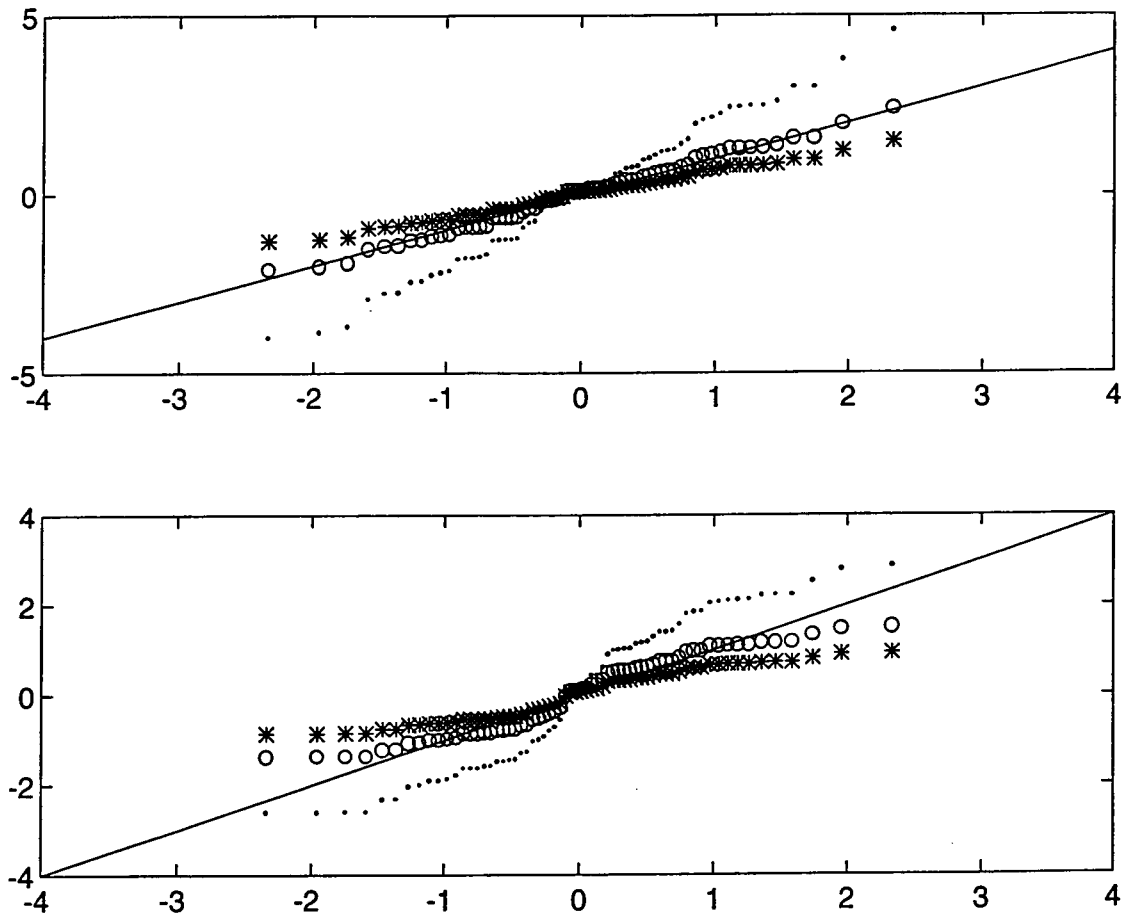


Figure 3.2 (Continued)

Concluding Remarks

The results of the simulation studies demonstrate the behavior of the standardized PRESS residuals and the statistic T_{PR} under various misspecifications of $\boldsymbol{V}$. For the exponential and Gaussian models considered here, under-estimation of any one of the parameters θ_0, θ_1 and θ_2 leads to inflated values of T_{PR}, and thus increased variability among the standardized PRESS residuals. Conversely, over-specification of $\boldsymbol{V}$ leads to small values of T_{PR}, and decreased variability among the standardized PRESS residuals. Thus, T_{PR} may be used to identify models for which the covariance structure has been over- or under-specified. For the exponential covariance model, misspecification of the parameters of the covariance structure is likely to have little effect on the predicted surface, but may have more serious consequences for the prediction error variance. Cressie and Zimmerman (1992) argue that the estimated mean square prediction error tends to be negatively biased for the actual mean square prediction error, and that therefore, the covariance function should be over-estimated in order to obtain an accurate estimator for the mean square prediction error. While asymptotically this bias is negligible, substantial bias in the estimated mean square prediction error may be observed for small samples. Examination of the statistic T_{PR}, and adjusted normal probability plots, could point to over-specified covariance models. However, results of a simulation study conducted by Cressie and Zimmerman (1992) suggest that for small samples, the actual *mspe* tends to be under-estimated when spatial correlation is weak, yet over-estimated when spatial correlation is strong. It is therefore difficult to judge whether over-specification of $\boldsymbol{V}$ is appropriate.

In addition to the situations described in the examples presented here, the adjusted normal probability plots might be used in an exploratory sense to examine models with varying geometric anisotropic structure. For instance, it would be possible to compare the normal probability plot produced for an isotropic specification of the covariance matrix with a number

of anisotropic alternatives, varying the values of the covariance function parameters in perhaps two principal directions.

The major advantage of the adjusted probability plots is that they require computation of the standardized PRESS residuals from a single specified covariance model, but allow the analyst to screen any number of alternative models, without having to perform the fairly intensive computations required to cross-validate each of those models. In this way, the analyst may quickly and easily identify a smaller subset of candidate models to investigate in further detail.

Bibliography

Blom, G. 1958. Statistical Estimates and Transformed Beta-variables. J. Wiley and Sons, New York.

Bradley, R. and Haslett, J. 1992. High-interaction diagnostics for geostatistical models of spatially referenced data. The Statistician 41 : 371 – 380.

Christensen, R. 1990. Linear Models for Multivariate, Time Series, and Spatial Data. Springer-Verlag, New York.

Christensen, R., Johnson, W., and Pearson, L. M. 1992. Prediction diagnostics for spatial linear models. Biometrika 79(3) : 583 – 591.

Christensen, R., Johnson, W. and Pearson, L. M. 1993. Covariance function diagnostics for spatial linear models. Mathematical Geology 21 : 255 – 265.

Cook, R. D. and Weisberg, S. W. 1989. Regression diagnostics with dynamic graphics. Technometrics 31 : 277 – 291.

Cressie, N. 1991. Statistics for Spatial Data. J. Wiley and Sons, New York.

Cressie, N. and Zimmerman, D. L. 1992. On the stability of the geostatistical method. Mathematical Geology 24 : 45 – 59.

Journel, A. G. and Huijbregts, C. J. 1978. Mining Geostatistics. Academic Press, New York.

Stein, M. L. 1988. Asymptotically efficient prediction of a random field with a misspecified covariance function. Annals of Statistics 16 : 55 – 63.

Stein, M. L. and Handcock, M. S. 1989. Some asymptotic properties of kriging when the covariance function is misspecified. Mathematical Geology 21 : 171 – 190.

Thompson, C. M. and Ramsey, F. L. 1995. Use of a modified PRESS statistic for spatial linear model evaluation. Manuscript to be submitted to Mathematical Geology.

Warnes, J. J. 1986. A sensitivity analysis for universal kriging. Mathematical Geology 18 : 653 – 676.

SUMMARY

This chapter summarizes the major findings of the thesis, highlighting the advantages of each of the proposed methods, as well as stressing potential limitations in their application. Suggestions for future research concerning the standardized PRESS residuals and other diagnostics for identifying misspecifications of spatial covariance structure are also discussed.

Conclusions

The research described in the thesis focused on standardized versions of the PRESS residuals. These are similar to the PRESS residuals commonly calculated as part of regression analyses, but are distinct in that each residual is standardized by the quantity $\sigma_{e_{-i}}$, representing the square root of mean square prediction error obtained from the specified model, when $n-1$ data points (excluding the i^{th} observation) are used to predict the value of the process at the i^{th} location. If prediction is the goal of a particular regression analysis, PRESS residuals and the PRESS statistic are typically examined for a set of candidate models, each containing a different subset of available predictor variables. In geostatistics, the usual objective is to obtain a model which provides adequate predictions of a process at unmeasured spatial locations. However, the primary focus in geostatistics is to identify models with good predictive capacity, through adequate specification of the covariance structure for the spatial process under investigation. While predictor variables may be included in universal kriging models, the main purpose in examining the standardized PRESS residuals is to detect outliers and identify possible misspecifications of the covariance function.

Obtaining an adequate estimate of the covariance matrix $\boldsymbol{V}$ may be described as the weak link in the kriging chain. Estimates of the covariance structure are often based on a very limited number of observations. Misspecification of the covariance structure has two potential consequences. The most obvious of these is that poor predictions may be obtained. Secondly,

the mean square prediction errors may be too large or too small if the covariance structure is misspecified (Christensen, 1990; Cressie and Zimmerman, 1992). It is therefore important to determine whether the covariance structure has been adequately specified.

The second chapter of the dissertation deals with a particular case-deletion statistic based on standardized PRESS residuals, which is one of several currently popular with geostatisticians. In attempting to fit or choose a model for the covariance structure, through fitting a model to either the estimated variogram or the estimated covariance function, geostatisticians will typically examine the average of the standardized PRESS residuals, as well as the statistic T_{PR} (or more likely, T_{PR}/n) in order to determine whether an obvious misspecification of the covariance structure has occurred. If the statistic T_{PR}/n is close to 1, then the model is considered to be adequate. In the second chapter, a saddlepoint approximation to the distribution of T_{PR} is presented, as well as estimated tail probabilities associated with the distribution. This provides the analyst with an easily accessible approximate "p-value" for calibrating the statistic T_{PR}. This is beneficial, since the results indicate that the distribution of T_{PR} is highly dependent on the null model specified. The possibility of calculating T_{PR} for regional subsets of the data is also explored.

The development assumes that the spatial process is Gaussian, and that the covariance function $\boldsymbol{V}$ is positive definite. Note that the latter assumption holds only for transitive variogram models (those having a sill, that is, those which are second-order stationary). In fact, most of the more popular geostatistical models for covariance structure (the exponential, spherical, and Gaussian models) are positive definite .

In the third chapter, properties of T_{PR} and other summaries of the standardized PRESS residuals are investigated further. Because of the standardization of these residuals, the t_i's, misspecification of the covariance structure can result in either large or small values of T_{PR}, depending on the nature of the misspecification. This is in contrast to regression analysis,

where small values of PRESS suggest that a model has good predictive capacity. Small values of T_{PR} may indicate that the mean square prediction errors obtained from the model are too high. Therefore, it is important to identify values of T_{PR} in either extreme.

The results show that T_{PR} has little power to detect certain misspecifications of $\boldsymbol{V}$. However, several authors, including Diamond and Armstrong (1984), and Cressie and Zimmerman (1992), have suggested that kriging is a fairly stable procedure. It is likely that T_{PR} has sufficient power to detect those misspecifications of $\boldsymbol{V}$ which could result in a poorly predicted surface, or seriously biased estimates of the mean square prediction error.

The adjusted normal probability plots presented in the third chapter, while not ideal, emphasize the fact that certain patterns in the standardized PRESS residuals suggest a particular type of misspecification of the covariance structure. In particular, high variability among the standardized PRESS residuals suggests that one or more parameters of the covariance function has been under-specified, while under-dispersion points to over-specification of the covariance structure. Thus, patterns in the normal probability plots suggesting high or low dispersion of the t_i's can be used to identify particular misspecifications of the covariance function. Using the method described in the third chapter, any candidate model can be screened using the cross-validation calculations from a single specified model.

Suggestions for Further Research

Given that the statistic T_{PR} was observed to have little power for detecting certain misspecifications of the covariance function, one suggestion might be to develop a more powerful statistic, and associated testing procedures. However, since the kriging procedure has been shown to be reasonably stable in the presence of at least certain types of misspecification of the covariance function (Cressie and Zimmerman, 1992), it is likely that T_{PR} has adequate power to detect those misspecifications of $\boldsymbol{V}$ which could result in a poorly predicted surface. More

difficult to ascertain is the required power to detect those misspecifications which effect the mean square prediction errors associated with a particular covariance model. Preliminary results of Cressie and Zimmerman (1992) suggest that, at least for sufficiently large samples, over-specified covariance models are advantageous, in that they more adequately reflect the mean square prediction errors associated with the actual predictions. However, the size and direction of any bias in the *mspe* is sample size dependent, and needs further investigation.

Rather than searching for statistics with more power to detect general misspecifications of $\boldsymbol{V}$, it would be more fruitful to concentrate on statistics which detect particular inadequacies of the specified covariance model. For instance, diagnostics to detect anisotropy, or misspecification of a single parameter of the covariance function would be useful additions to the set of standard geostatistical exploratory data analysis techniques.

The procedures described in the second chapter render an already familiar statistic T_{PR} more informative, by providing an easily implemented method for calculating approximate p-values associated with the statistic. One minor problem with the method at present is that a good initial estimate of the saddlepoint w_0 is required in order for the Newton-Raphson algorithm for the calculation of approximate p-values to converge to a solution in the interval [0,1]. Guidelines for the selection of starting values of w_0 would improve the practicality of the method.

The adjusted probability plots of the third chapter are a useful screening technique for distinguishing among various candidate models for covariance structure. However, development of these plots was based on a very general observation concerning variability of the standardized PRESS residuals under certain types of misspecification of the covariance structure. Potential improvements to these plots could come through incorporation of the compatibility ideas discussed by Stein (1988) and Stein and Handcock (1989). Inclusion of envelopes in probability or quantile-quantile plots representing compatible covariance functions would be an important

advance. However, it is important to note that results of Stein (1988) are based on infill asymptotics, and therefore their applicability in the small sample setting is uncertain. Diamond and Armstrong (1984) also discuss the definition of neighborhoods of similar variogram or covariance function models, in terms of the predictions and mean square prediction errors they produce.

BIBLIOGRAPHY

Allen, D. M. 1971. The prediction sum of squares as a criterion for selecting predictor variables. Technical Report No. 23, Department of Statistics University of Kentucky, Lexington, KY.

Atkinson, A. C. 1985. Plots, Transformations, and Regression. Clarendon Press, Oxford.

Belsley, D. A. 1991. Conditioning Diagnostics : Collinearity and Weak Data in Regression. J. Wiley and Sons, New York.

Belsley, D. A., Kuh, E., and Welsch, R. E. 1980. Regression Diagnostics : Identifying Influential Data and Sources of Collinearity. J. Wiley and Sons, New York.

Blom, G. 1958. Statistical Estimates and Transformed Beta-variables. J. Wiley and Sons, New York.

Bradley, R. and Haslett, J. 1992. High-interaction diagnostics for geostatistical models of spatially referenced data. The Statistician 41 : 371 – 380.

Carr, J. R. and Roberts, K. P. 1989. Application of universal kriging for estimation of earthquake ground motion : Statistical significance of results. Mathematical Geology 21 : 255 – 265.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. 1983. Graphical Methods for Data Analysis. Duxbury Press, Boston.

Chauvet, P. 1982. The variogram cloud. Internal Note N-725, Centre de Geostatistique, Fontainebleau, France.

Christensen, R. 1990. Linear Models for Multivariate, Time Series, and Spatial Data. Springer-Verlag, New York.

Christensen, R., Johnson, W. and Pearson, L. M. 1992. Prediction diagnostics for spatial linear models. Biometrika 79(3) : 583 – 591.

Christensen, R., Johnson, W. and Pearson, L. M. 1993. Covariance function diagnostics for spatial linear models. Mathematical Geology 25(2) : 145 – 160.

Cook, R. D. 1977. Detection of influential observations in linear regression. Technometrics 19 : 15 – 18.

Cook, R. D. and Weisberg, S. 1982. Residuals and Influence in Regression. Chapman and Hall, New York.

Cook R. D. and Weisberg, S. 1989. Regression diagnostics with dynamic graphics. Technometrics 31 : 277 – 291.

Cressie, N. 1984. Towards resistant geostatistics. In Geostatistics for Natural Resource Estimation, Part 1, G. Verly, M. David, A. G. Journal, and A. Marachel, eds. Reidel, Dordrecht, 21 – 44.

Cressie, N. 1991. Statistics for Spatial Data. J. Wiley and Sons, New York.

Cressie, N. and Hawkins, D. M. 1980. Robust estimation of the variogram, I. Journal of the International Association for Mathemtical Geology 12 : 115 – 125.

Cressie, N. and Zimmerman, D. L. 1992. On the stability of the geostatistical method. Mathematical Geology 24 : 45 – 59.

Daniels, H. E. 1954. Saddlepoint approximations in statistics. Annals of Mathematical Statistics 25 : 631 — 649.

Daniels, H. E. 1987. Tail probability approximations. International Statistical Review 55 : 37 — 48.

Davis, B. M. 1987. Uses and abuses of cross-validation in geostatistics. Mathematical Geology 19 : 241 — 248.

Diamond, P. and Armstrong, M. 1984. Robustness of variograms and conditioning of kriging matrices. Mathematical Geology 16 : 809 – 822.

Gnedenko, B. 1968. The Theory of Probability. Chelsea Publishing Co., New York.

Gordon, N. H. and Ramig, P. F. 1983. Cumulative distribution function of the sum of correlated chi-squared random variables. Journal of Statistical Computing and Simulation 17 : 1 — 9.

Hamlett, J. M., Horton, R., and Cressie, N. 1986. Resistant and exploratory techniques for use in semivariogram analyses. Soil Science Society of America Journal 50 : 868 – 875.

Haslett, J., Bradley, R., Craig, P., Unwin, A., and Wills, G. 1991. Dynamic graphics for exploring spatial data with application to locating global and local anomalies. American Statistician 45(3) : 234 – 242.

Isaaks, E. H. and Srivastava, R. M. 1989. Applied Geostatistics. Oxford University Press, New York.

Jones, R. H. 1989. Fitting a stochastic partial differential equation to aquifier data. Stochastic Hydrology and Hydraulics 3 : 85 – 96.

Journel, A. G. and Huijbregts, C. J. 1978. Mining Geostatistics. Academic Press, New York.

Knudsen, H. and Kim, Y. C. 1978. Application of geostatistics to roll front type uranium deposits : 107^{th} AIME Annual Meeting, Denver Colorado. AIME Preprint No. 78-AR-94.

Krishnaiah, P. R., Hagis, P. and Steinberg, L. 1963. A note on the bivariate Chi-square distribution. SIAM Review 5 : 140 – 144.

Lugannani, R. and Rice, R. O. 1980. Saddlepoint approximation for the distribution of the sum of independent random variables. Advances in Applied Probability 12 : 475 – 490.

Lukacs, E. and Laha, R. G. 1964. Applications of the Characteristic Functions. Griffin's Statistical Monograph No. 14, Griffin, London.

Mathai, A. M. and Provost, S. B. 1992. Quadratic Forms in Random Variables : Theory and Applications. Marcel Dekker, New York.

Matheron, G. 1963. Principles of geostatistics. Economic Geology 58 : 1246 – 1266.

The Math Works, Inc. 1992. MATLAB Reference Guide. The Math Works, Inc., Massachusetts.

Reid, N. 1988. Saddlepoint methods and statistical inference. Statistical Science 3 : 213 – 238.

Robinson, J. 1982. Saddlepoint approximation for permutation tests and confidence intervals. Journal of the Royal Statistical Society, Series B 44 : 91 – 101.

Solomon, H. and Stephens, M. A. 1977. Distribution of a sum of weighted chi-square variables. Journal of the American Statistical Association 72 : 881 – 885.

Solow, A. R. 1990. Geostatistical cross-validation : A cautionary note. Mathematical Geology 22 : 637 – 639.

Stein, M. L. 1988. Asymptotically efficient prediction of a random field with a misspecified covariance funtion. Annals of Statistics 16 : 55 – 63.

Stein, M. L. and Handcock, M. S. 1989. Some asymptotic properties fo kriging when the covariance function is misspecified. Mathematical Geology 21 : 171 – 190.

Thompson, C. M. and Ramsey, F. L. 1995. Use of a modified PRESS statistic for spatial linear model evaluation. Manuscript to be submitted to Mathematical Geology.

Warnes, J. J. 1986. A sensitivity analysis for universal kriging. Mathematical Geology 18 : 653 – 676.

Zimmerman, D. L. and Zimmerman, M. B. 1991. A Monte Carlo comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. Technometrics 33 : 77 – 91.

APPENDICES

APPENDIX A

MATLAB program to calculate the saddlepoint approximation to the density of T_{PR}. The program also computes and plots a χ^2_n density, and the histogram of 10,000 simulated values of T_{PR} for comparison. This particular version of the program deals with 7×7, assuming an exponential covariance for $\boldsymbol{V}$.

```
% This program computes the cumulative generating function, as well as its
% first four derivatives. However, only the first two derivatives are
% used in the saddlepoint approximation

l = 7;
n = l*l;
nd = 1256;
nv = n*ones(nd,1);
load grid7e10.dat;
s = grid7e10;

%  Set up variance covariance matrix for Z exponential model

load d77.dat;

r = 10;
r=r/3;
v = zeros(n,n);
i=1;
while i<= n,
   j=i;
   while j <= n,
     lag = d77(i,j);
     v(i,j) = exp(-(lag/r));
     v(j,i) = v(i,j);
   j = j + 1;
   end
i = i+ 1;
end

%  Set up variance covariance matrix for the terms of S

for j = 1:n,
id = eye(n,n);
djm = [id(:,1:j-1) id(:,j+1:n)];
djp = id(:,j);
op = ones(n-1,1);
cvi = inv(djm'*v*djm);
tm1 = djm'*v*djp;
tm2 = (1-op'*cvi*tm1);
tm3 = op'*cvi*op;
```

```
lambda = (tm1 + (tm2/tm3))'*cvi;
sig(j) = sqrt(1 - (lambda*tm1) + (tm2/tm3));

if j ==1,
  lam2 = [1/sig(j) -lambda/sig(j)];
elseif j == n,
  lam2 = [-lambda/sig(j) 1/sig(j)];
else,
  lam2 = [-lambda(1:j-1)/sig(j) 1/sig(j) -lambda(j:n-1)/sig(j)];
end

lam2 = lam2';
if j == 1,
  a = lam2;
else,
  a = [a lam2];
end

a = reshape(a,n,j);
end

vr = a'*v*a;

lambda = abs(eig(vr));

k = zeros(nd,1);
d1k = zeros(nd,1);
d2k =  zeros(nd,1);
dt = zeros(nd,1);
d3k = zeros(nd,1);
d4k = zeros(nd,1);
lam3 = zeros(nd,1);
lam4 = zeros(nd,1);
chi = zeros(nd,1);

t = -1.1:.001:.155;
t = t';

for i = 1:nd,
  k(i) = -0.5*sum(log(1-2*t(i)*lambda));
  d1k(i) = sum(lambda./(1-2*t(i)*lambda));
  d2k(i) = 2*sum((lambda./(1-2*t(i)*lambda)).^2);
  d3k(i) = 8*sum((lambda./(1-2*t(i)*lambda)).^3);
  d4k(i) = 48*sum((lambda./(1-2*t(i)*lambda)).^4);
  lam3(i) = d3k(i)./d2k(i).^1.5;
  lam4(i) = d4k(i)./d2k(i).^2;

  if i >= 2,
    dt(i) = d1k(i) - d1k(i-1);
  end
end

% Saddlepoint approximation
```

```
n1 = 1;

g=sqrt(n1./(2*pi.*d2k)).*exp(k-t.*d1k);
g1=sqrt(n1./(2*pi.*d2k)).*exp(k-t.*d1k).*(1+((lam4/8)-(5*lam3.^2/24)));
chi = 1/(2^(n/2)*gamma(n/2)).*(d1k.^((n/2)-1)).*exp(-d1k/2);

% Plot of saddlepoint approximation, chi-squared n, and histogram for simulated values

step = max(s)/30;
bins = 0:step:max(s);
[nx,ssm] = hist(s,bins);

 plot(d1k,g,'-.',ssm,nx/(step*10000),'-',d1k,chi,':');
```

APPENDIX B

MATLAB function to calculate approximate tail probabilities associated with the saddlepoint approximation to the density of T_{PR}.

```
% Function to compute exact p-values for given critical value of Tpr

function p = psoln(tpr,lambda)

% Initial starting value for the saddlepoint, t0

t0=.20;
j = 1;
for i = 1:100;
  k = -0.5*sum(log(1-2*t0*lambda));
  d1k = sum(lambda./(1-2*t0*lambda)) - tpr;
  d2k = 2*sum((lambda./(1-2*t0*lambda)).^2);
  d3k = 8*sum((lambda./(1-2*t0*lambda)).^3);
  d4k = 48*sum((lambda./(1-2*t0*lambda)).^4);
  lam3 = d3k./d2k.^1.5;
  lam4 = d4k./d2k.^2;
  ts = t0;
  t0 = t0 - (d1k/d2k);
  if abs(t0 - ts) > .00001,
    i = i + 1;
    j = j + 1;
  else
    i= 100;
  end
end

% Calculate tail probabilities using Lugannani-Rice method
d1k=tpr;
z = t0.*sqrt(d2k);
cz = sqrt(2*(t0.*d1k -k)).*sign(t0);
rz = (1-cdfn(z))./pdfn(z);
rcz = (1-cdfn(cz))./pdfn(cz);
cdfz = cdfn(z);

l1 = lam3/6;
l2 = lam4/24;
l3 = (lam3.^2)/72;
term1 = pdfn(cz);

qlr = term1.*(rcz + (1 ./z) - (1 ./cz));

% Output of specific p-values
p = qlr;
```

APPENDIX C

MATLAB program to generate 1000 realizations from a particular covariance model, and then cross-validate each realization under a specified model for $\boldsymbol{V}$. Summary statistics (T_{PR}, $\widehat{p}$, s^2, range($\boldsymbol{t}$), min($\boldsymbol{t}$), max($\boldsymbol{t}$), mean($\boldsymbol{t}$)) are output for each surface.

```
% CVSIM.M
% 15-1-95
% Program to generate surfaces, and subject each to
% Cross Validation, assuming a model of a particular form


%  Enter desired grid size, eg 3 5 will produce a 3 x 5 grid of points
%  Create coodinates for the specified grid

% g1 = input('number of rows in grid  :');
% g2 = input('number of columns in grid  :');

g1 = 8;
g2 = 8;
n = g1*g2;
o1 = ones(g2,1);
s1 = 1:1:g1;
co1 = o1*s1;
c1 = reshape(co1,n,1);
o2 = o1;
s2 = 1:1:g2;
co2 = o2*s2;
co2 = co2';
c2 = reshape(co2,n,1);
x = [c1,c2];

%  Compute distances among all pairs of points

load d64.dat;
d = d64;

%  Select type of covariance model to be used

disp('  1 = exponential');
disp('  2 = spherical ');
disp('  3 = gaussian  ');

mtype = 3;

% a = input('theta2 :');
a = 0.8;
% var = input('theta1 : ');
var = 1.0;
% ne = input('theta0 :');
```

```
ne = 0.2;
covm = zeros(n,n);

if mtype == 1;

%  Compute the covariance matrix, assuming an exponential covariance
%     of the form
%
%      Cov(r) = k*(exp(-ra))
%
%    where k represents the a priori variance of the random function

     covm = var*exp(-a*d);
     covm = covm+(ne*eye(n,n));

elseif mtype == 2;

% Compute the covariance matrix, assuming a spherical covariance of
%        the form
%
%       Cov(r) = k * (1 - (1.5(ra) - 0.5(r^3*a^3))    r <= a
%
%              = 0                          r > a
%     where k represents the a priori variance of the random function

covm = zeros(n,n);
u=1;
while u <= n;
   v = u;
   while v <= n;
    if d(u,v) < a;
    covm(u,v) = var*(1 - (1.5*(d(u,v)/a)) - (0.5*(d(u,v)/a)^3) - ne);
    end
    covm(v,u) = covm(u,v);
    v = v + 1;
  end
u = u + 1;
end

elseif mtype == 3;

%  Compute the covariance matrix, assuming a Gaussian covariance
%        of the form
%
%       Cov(r) = k * exp(-(r*a)^2) ;
%
%     where k represents the a priori variance of the random function
    covm = var*(exp(-(d*a).^2));
    covm = covm + (ne*eye(n,n));

end

tbar1 = zeros(1000,1);
```

```
cvi1 = lami1 - ((li1*li1')./inv1(i,i));
vi1 = dim'*v*dip;
xi = dim'*x;
xti1 = (x(i,:)' - (xi'*cvi1*vi1))';
xt1(i,1) = xti1(1,1);
zi = dim'*z;
zt1(i) = z(i) - zi'*cvi1*vi1;
h1(i) = xt1(i,:)*xinv1x*xt1(i,:)';
s1(i) = v(i,i) - vi1'*cvi1*vi1;
pr1(i) = (zt1(i) - xt1(i,:)*b1)*(s1(i)/(s1(i) - h1(i)));
t1(i) = (zt1(i) - xt1(i,:)*b1) / (sqrt(s1(i) - h1(i)));
ci1 = (dip-(dim*cvi1*vi1))-(xt1(i,:)*xinv1x*x'*inv1)';
if i == 1,
    c1 = ci1;
else,
  c1 = [c1 ci1];

end
end

vt1 = c1'*v*c1 ./ (sqrt(s1-h1)*(sqrt(s1-h1)'));

lambda1=eig(vt1);
tp1 = t1'*t1;
tpr1(it) = tp1;

tbar1(it) = mean(t1);
stdt1(it) = std(t1)^2;
pt1(it) = psoln(tp1,lambda1);
maxt1(it) = max(t1);
mint1(it) = min(t1);
ranget1(it) = range(t1);
press1(it) = pr1'*pr1;

end

outp1 = [tpr1 pt1' tbar1 stdt1 maxt1 mint1 ranget1 press1];
outp2 = [tpr2 pt2' tbar2 stdt2 maxt2 mint2 ranget2 press2];
save sim11.out outp1 /ascii;
save sim12.out outp2 /ascii;
```