Some aspects of Smoothing Techniques in the Modelling of Spatial Data

Angela Magdalena Diblasi

A Dissertation Submitted to the

University of Glasgow

for the degree of

Doctor of Philosophy

Department of Statistics

August 1996

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

The purpose of this thesis is to explore, apply and develop statistical tools in the area which has been called Spatial Statistics. Through all of this work, a key link is the use of smoothers. Specifically, loess and splines are used in the second chapter and kernel smoothers in the rest. Smoothing techniques are now widely used in a variety of modelling problems. It is their application to the specific area of spatial statistics which is the focus of this thesis.

One particular application involves the modelling the mackerel egg density in the eastern Atlantic This led to the proposal of a generalized additive model for these data. Due to lack of distributional theory for estimators and methods of selection of a model, the proposed model is the result of an analysis which is analogous to that used in the context of generalized linear models. To assess and compare this model with others proposed in the literature, from the point of view of the estimation of the total number of mackerel eggs, the technique of the bootstrap is used.

The spatial processes considered in each chapter are of the form:

$$Y(\boldsymbol{s}_i) = f(\boldsymbol{x}(\boldsymbol{s}_i)) + \varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ is a vector of covariates and $\boldsymbol{s}_i, \quad i = 1, \ldots, n$ are the locations where the process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is observed, and $\varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n$ is the process of errors at the observed locations which is assumed gaussian, stationary and isotropic through the whole work.

Checking the covariance structure of this kind of spatial processes led to the development of a test statistic for the null hypothesis of constant variogram. Because of the lack of distributional theory for the residuals of a generalized additive model, the test is proposed for the case where the function f of the covariates is a linear function.

A test for checking homoscedasticity in a linear model is also proposed as a preliminary study of that for the constant variogram. In both cases, reference bands are also proposed as graphical tools to check constant variance and constant variogram, respectively.

The block-bootstrap approach is analyzed to build confidence intervals for the variogram when it is generated from a regular grid in $\mathbb{R}^2$. The percentage coverage of these intervals is compared with a technique proposed in the literature under more restrictive assumptions.

Techniques of resampling and simulation are employed through the whole work, and the available methods in this area are reviewed and compared as a separate exercise. Further suggestions are also made.

To my family

Acknowledgements

- There no words, nor in English neither in Spanish, to thank to Professor Adrian Bowman, my supervisor. I am greatly indebeted to him for his patient and unfailing support and guidance. He was always ready to listen and give advice. He offered to me his house, his family, and friendship. He helped me also to find aditional financial support from the U.K. Thank you very much, Adrian!!!

- I owe another large debt of gratitude to my friends and colleges in Mendoza and the goverment in Argentina for supporting me in this challenge of a Ph.D. Their confidence in me and its financial support have been also crucial for me during these years in Glasgow. I wish to thank in particular to the heads and colleagues of the Universidad Nacional de Cuyo, Universidad Tecnológica Nacional and Centro de Investigaciones Científicas y Tecnologicas. !'Muchas gracias a todos!

- I owe a debt of gratitude to the members of the Statistic Department of Glasgow University. Members of the staff and postgraduate students have shared good and hard times of my life during my stay in Glasgow. In particular, I want to thank to Drs Ben Torsney and Marian Scott. I am further indebted to Dr James Kay for his invaluable support and encouragment. I am grateful for many conversations about my subject with Dr Alexander Ivanov. I also acknowledge the head of the department, Prof. Ian Ford, for the financial support. My acknowledgment also to Dr Alan Dunmur and Ilya Molchanov for their helps with the final typing of this thesis and to Mary Nisbet for her help and kindness.

- I want to thank to Adelchi Azzalini, from Padua, in Italy, to help me to come to the U.K.

- I cannot also forget the times I was working with Kim Anh Do, from Brisbane, in Australia. She shared her enthusiasm and inspiration with me.

- I acknowledge the interesting sets of data provided by Professor Steve Buckland of the University in St Andrews and Dr Pride from Aberdeen, by Brian Miller and Kirsty Dalziel from the Clyde River Purification Board in Glasgow, and by Ralph Ogden and Ross Cunningham of the Australian National University, in Canberra. I also aknowledge the papers provided by Drs Andy Wood and Grace Chan.

- During my times in Glasgow some other new friendships have provided support. In particular I wish to thank to my friends Necla Gunduz, Mredula Singh, Athanase Polymenis, Gilla Ghaffari, Reza Eshraghian, Monica Chiogna, Maura Blaney, Pauline Cairns and Natalia Warms. Friends who I have shared my life with within and outwith the walls of the University.

- My gratitude also to my friend Julia Storey whose help with the English Language and constantly support aided the completion of this thesis.

- Thanks to Prof. Neville Davis from Nottingham Trent University who warmly invited me to give a seminar about one of the subjects of this thesis.

- My gratitude to many other to long to list in here who also gave me tremendous support to make this work possible. In particular my friends in Argentina.

- Finally I wish to thank my mother, my sister, my brother in law and my nephews and nice for their unconditional help and love.

Or like the borealis race
that flit ere you can point their place;
Or like the rainbow's lovely form
Evanishing amid the storm.
Nae man can tether time nor tide;
Robert Burns

Inconmesurable silencio de las noches en Los Andes
que permite escuchar a las estrellas

Contents

Abstract i

Acknowledgements iv

1 Introduction 1

1.1 Spatial data and smoothing techniques 1

1.2 Applied work 3

1.3 An exercise to check homoscedasticity 4

1.4 Techniques for checking covariance structure 6

1.5 Algorithms for simulation and re-sampling 7

1.6 Summary of chapters 8

2 Analysing the distribution of Mackerel egg biomass 10

2.1 Introduction 10

2.2 Previous literature 13

2.3 The data 16

2.3.1 The zero-valued data 18

2.3.2 The sample grid 19

2.4 A model for density of mackerel-egg 21

2.4.1 Introduction 21
2.4.2 Generalised Additive Models (G.A.M.) 22
2.5 Selection of models 31
2.5.1 A model for density of mackerel eggs 36
2.5.2 Difficulties with GAM inference 39
2.6 Bootstrap Analysis of total 40
2.7 Conclusions 50

3 Testing for Constant Variance in a Linear Model 52
3.1 Introduction 52
3.2 A description of the test 62
3.2.1 The test statistic 62
3.2.2 The distribution of T 64
3.2.3 Three examples 68
3.3 Alternative versions 72
3.3.1 Ignoring the correlation of the residuals 72
3.3.2 A bootstrap approach 74
3.4 A power study 76
3.5 A reference band 80
3.6 Some remarks 84

4 Testing for Constant Variogram 89
4.1 Introduction 89
4.2 Previous literature 92
4.3 A test statistic 96

4.3.1 Adapting the test statistic for a process defined on a regular grid 105

4.3.2 Independent residuals 108

4.4 Simulation results 111

4.5 Examples 115

4.5.1 Example 1: Simulated data 118

4.5.2 Constant semivariogram 119

4.5.3 Example 2: In areas nearby the Clyde River 120

4.5.4 Example 2: In some billabongs 124

4.6 Reference bands 127

4.7 Comments 129

5 Simulating and re-sampling methods for spatial data 133

5.1 Introduction 133

5.2 Simulation methods 137

5.2.1 Using a similarity transformation of the covariance matrix 140

5.2.2 LU triangular decomposition of the covariance matrix . 141

5.2.3 Embedding the covariance matrix in a circulant matrix 143

5.2.4 The turning bands method 149

5.3 Incorporating dependence with kernel smoothers 156

5.4 Re-sampling methods for hypothesis testing 162

5.5 Building local confidence intervals for the variogram through block-bootstrap 163

5.5.1 Introduction 163

5.5.2 Previous Literature 165

5.5.3 The methodology 168

5.6 Simulation results 171
5.6.1 Normal processes 171
5.6.2 Non-normal processes 175
5.6.3 Some remarks 179
5.7 Some general comments 180

6 Reflections **186**
6.1 Introduction 186
6.2 Generalized linear models 187
6.3 Linear models 188
6.4 Resampling 190

References **192**

List of Figures

1.1 Different sets of spatial data 2

2.1 The geographical position of mackerel data 17

2.2 Zero-value data 19

2.3 Regular grid 20

2.4 Fitted values of log(density of mackerel egg) to a G.A.M. model with the only one covariate "Bottom depth" 29

2.5 Fitted values of log(density of mackerel egg) to a G.A.M. model with the only one covariate "Distance to 200m contour" 30

2.6 The observed values of log(density of mackerel eggs) at each point on the regularized grid. 31

2.7 The values of log(density of mackerel eggs) fitted to the Generalised Additive model (G.A.M.) $\hat{log}(density) = loess(latitude, longitude)$ at each point on the regularized grid 32

2.8 Box-plots of the bootstrap distributions of the total number of mackerel eggs as calculated from different models 46

2.9 Box-plots of the empirical distributions of the total biomass of mackerel eggs from the selected model and when no model is fitted 48

2.10 The bootstrap frequencies of chosen estimator T_0 and the raw estimator T 48

2.11 A plot of the bootstrapped chosen estimator against the raw estimator 49

2.12 A box-plot of the bootstrap distribution of the difference between the two estimators of the Figure ?? 49

3.1 Data and fitted line for the snow geese data 69

3.2 Data and fitted line for the subset of the snow geese data . . . 70

3.3 Data and fitted line for the male cats data 71

3.4 Simulated data from the model $y_i = 1 + 2x_i + \varepsilon_i$ was used, where x_i is a design point in [0,1] and ε_i has a normal distribution with mean zero and standard deviation $\sigma(x_i) = (0.5 + x_i)0.25, \quad i = 1, \ldots, 50.$ 72

3.5 The 95%-reference bands for the data presented in the examples 83

3.6 Reference bands for a range of smoothing parameters with the simulated Data 85

3.7 Reference bands for simulation data 86

4.1 Models for the semivariograms used in the simulations 112

4.2 Two simulated gaussian processes on a regular grid 119

4.3 Estimated and theoretical variograms - "raw s" and "smooth s" 120

4.4 Scatter plot of the indicators of pollution 121

4.5 Estimated semivariogram, "raw s" and "smooth s" for the indicators of pollution 123

4.6 Scatter plots of phosphorus, nitrogen and turbidity in the billabongs 125

4.7 Estimated semivariogram, "raw s" and "smooth s" for the variables of the billabongs 126

4.8 Reference bands for the simulated processes 129

4.9 Reference bands for the indicators of pollution 130

4.10 Reference bands for the processes of residuals in the models for the billabongs 131

5.1 Two simulated processes with different variograms 148

5.2 The turning bands corresponding to one hyper-plane 153

5.3 Uniformly distributed hyper-planes 154

5.4 Building a dependent process with kernel smoothers 158

5.5 Kernel smoothers with compact supports 160

5.6 A stationary process and its estimated covariance function . . 161

5.7 Block-bootstrap distributions of the variogram estimators . . . 174

5.8 Block-bootstrap and log-normal distribution of the variogram estimator 175

5.9 Block-bootstrap and log-normal local confidence intervals for the variogram 176

List of Tables

2.1 Analysis of deviances for different generalised additive models (G.A.M.) 37

2.2 Analysis of deviances of several G.A.M for the mackerel data . 38

3.1 Size and power of the test 78

3.2 Power for the three different versions of the test 88

4.1 Size and power of the test for constant variogram 114

4.2 Size and power of the test for constant variogram 115

4.3 Size and power for the different versions of the test for constant variogram 116

4.4 Size and power for the different versions of the test for constant variogram 117

4.5 Size for the three versions of the test for constant variogram . 117

5.1 Coverage percentages for block-bootstrap confidence intervals . 182

5.2 Coverage percentages for block-bootstrap confidence intervals . 183

5.3 Coverage percentage for a non-Gaussian process 184

5.4 Coverage percentage for a non-Gaussian process 185

Chapter 1

Introduction

Aquí me pongo a cantar
al compás de la vigüela....
José Hernandez

1.1 Spatial data and smoothing techniques

Spatial data arise when observations on random variables are related to locations in a set of two or more dimensions and their properties referred to this spatial set. In other words, spatial data are realizations of a stochastic process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ where D is a subset of $\mathbb{R}^m$ with $m \geq 2$ and $Y(\boldsymbol{s})$ can be a vector or a single random variable. Through this work it will be a single variable. Natural examples of spatial data are those dealing with resource assessment, environmental monitoring, medical imaging, etc.

The set D can be assumed to have different properties. Throughout this work it will be considered a fixed subset of $\Re^m$ which contains an m-dimensional rectangle of positive volume. Nevertheless, most of the results obtained here can be extended to other kind of sets, such as regular or irregular lattices in $\mathbb{R}^m$. The observed set of data in the examples and applied work will display different characteristics such as regular grids in $\mathbb{R}^2$ in the simulations, regular but not rectangular set as in mackerel data, and irregular and not rectangular as in the examples dealing with spatial monitoring in chapter 4. All these different set of data are shown in Figure 1.1.

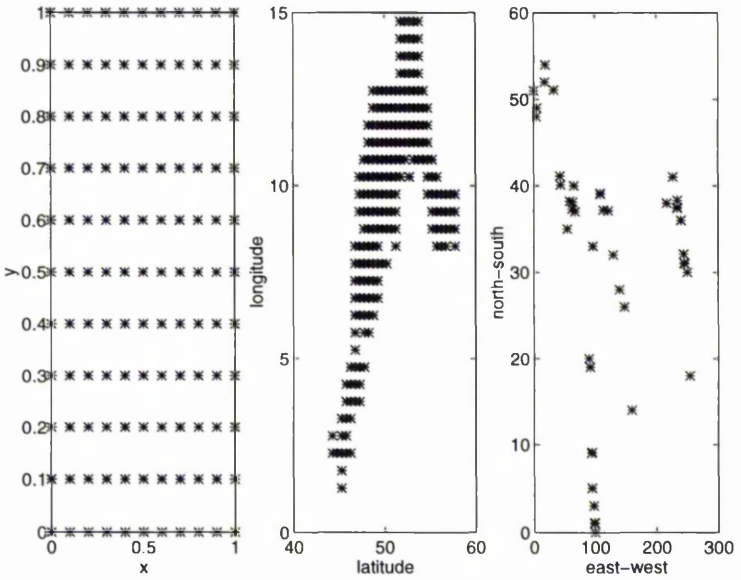

Figure 1.1. On the left a set of spatial data on a regular grid in $[0,1]^2$ is shown. In the middle, a regular set of data in the modelling mackerel data. On the right, the irregular locations of billabongs as discussed in chapter 4.

A preliminary question to modelling a spatial process is dealing with its mean and variance-covariance structure. When the parametric tools, so few

in this context, are not suitable to give an answer to these questions, the non-parametric ones can offer a flexible alternative to them. Both problems, modelling the mean and the variance-covariance matrix are tackled in this work. These questions are important because the presence and type of spatial correlation present in data can markedly affect the type of analysis which it is necessary adopt.

Modelling the mean is studied through an exercise in which smoother splines and loess (locally-weighted running smoothers) are used to fit a generalised additive model to a set of data as explained in the next section. In the second problem, the variance-covariance structure is modelled and here kernel smoothers are used. Also these tools are suggested as a way to simulate a isotropic stationary process.

As is widely known, the price paid for the flexibility of the non-parametric techniques is the difficulty to find the distributions and properties of the statistics involved in the analysis.

1.2 Applied work

One particular applied statistical problem dealing with spatial data has been tackled. A brief summary of it is given now.

The estimation of the total number of mackerel eggs from a survey as described in chapter 2 has been carried out. In this case, the original set of data was manipulated in order to make it more homogeneous, given the different criteria

under which parts of the survey were conducted.

A generalised additive model was fitted to the data. The response variable is the logarithm of the density of mackerel egg and the covariates are latitude, longitude, bottom depth and distance to the two hundred meters contour, as suggested in previous studies. Different nested generalised additive models were considered and a model with all of these covariates was chosen, following an examination of deviances as suggested by Hastie and Tibshirani (1990). These models were also compared with the model of Borchers et al.(1994) from the point of view of their behaviour in estimating the total number of mackerel eggs.

The main aim in this approach was to propose a model to estimate the total number of mackerel eggs which also can take into account the nature and meaning of the way in which these variables are introduced into the model.

This exercise was also a very useful introduction to the manipulation of spatial data. The exploration of techniques of re-sampling such as the bootstrap, and the awareness of the necessity to have a tool to analyse the structure of dependence or covariance was observed.

1.3 An exercise to check homoscedasticity

With the idea of searching for tools for checking for a constant variogram, a preliminary exercise has been carried out. The model considered was a simple linear model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$ where ε_i are normal random

variables with zero-mean and variance σ_i^2, $i = 1, \ldots, n$ and the statistical hypotheses to be considered were

$$\begin{aligned} H_0 : \sigma_i^2 &= \sigma_0^2, \quad i = 1, \ldots, n \\ H_1 : \sigma_i^2 &= \text{smooth function of } x_i, \quad i = 1, \ldots, n \end{aligned}$$

To check these hypotheses, tools from the context of smoothing techniques were used to compare the behaviour of the variables $s_i = |r_i|^{1/2}$ $i = 1, \ldots, n$, where the r_i's are the least squares residuals from fitting the data to the linear model under the null hypothesis. A kernel smoothed version of the variables s_i, $i = 1, \ldots, n$, was considered, in order to check their behaviour under the null and the alternative hypotheses.

A quadratic form in the variables s_i, $i = 1, \ldots, n$ arose as a natural test statistic. The distribution of this quadratic form was approximated by the distribution of a shifted and scaled χ^2 random variable by matching their moments.

A simulation study was performed in order to check the power and size of the test proposed. The approximate distribution of the test statistic was also compared with the empirical bootstrap distribution following a simulation study.

This approach provided useful preparation from the point of view of the knowledge of methods of approximation of the distribution of quadratic forms and of investigating smooth versions of the square root of residuals, which have the attractive property of having approximately normal distribution.

1.4 Techniques for checking covariance structure

With the knowledge and training provided by the previous exercise, a search for tools to check for a constant variogram, or independence of the variables in a spatial process, was begun.

Under the assumption of second order stationarity and isotropy, the assumption of constant variogram leads naturally to the assumption of independence of the variables in the process.

The idea of considering least squares residuals if the process is modelled by a linear model was again considered. In this case the "pilot" variables to check constant variogram were not precisely the least squares residuals but their differences for each value of the distance between two points in the set of data. These variables are also normal and, consequently, the square root of their absolute values are again approximately normal. This property has been used in the literature also in the context of spatial estimation (see, for instance, Hawkins & Cressie (1984)).

With these variables and the resource of the kernel smoothers, a test statistic for the hypothesis:

$$\begin{aligned} H_0 : 2\gamma(h_i) &= o_0^2, \quad i = 1, \ldots, n \\ H_1 : 2\gamma(h_i) &= \text{smooth function of } h_i, \quad i = 1, \ldots, n \end{aligned}$$

where $2\gamma(h) = var(Y(\boldsymbol{s}) - Y(\boldsymbol{t})$ for $(\boldsymbol{s}, \boldsymbol{t})$ such that $\| \boldsymbol{s} - \boldsymbol{t} \| = h$ was built

by using the the variables and smoothers described before. The test statistic here is again a quadratic form and despite the singular covariance matrix of the variables, its distribution can also be approximated by a χ^2 distribution.

A modification of this statistic is also proposed for the case of a data set on a regular grid with the aim of doing more efficient calculations.

Bootstrap and permutation techniques are also employed in order to study the power and size of the test. Simulated examples, and the application of the test to data dealing with indicators of pollution in the area nearby the river Clyde in Glasgow, and indicators of the quality of water of some billabongs in Australia, are also considered.

As a graphical tool to check the structure of a covariance or variogram, reference bands are built and displayed for the examples mentioned.

1.5 Algorithms for simulation and re-sampling

In the context of spatial statistics, the algorithm for simulating a process efficiently is an indisputable tool. A review of the different approaches proposed in the literature is given in chapter 5.

An idea for simulating a dependent Gaussian process, starting from an independent one by using smoothing techniques is also suggested.

Different approaches to re-sampling are also considered and the method of moving block-bootstrap is used to build confidence intervals for the variogram

of a Gaussian process. A simulation study was carried out to calculate the coverage of these intervals. A comparison of the coverage percentages of this method and of the confidence intervals from a distribution for the classical estimator of the semi-variogram as calculated under a more restrictive assumption (Baickowsky & Mardia, (1994)), is performed.

The behaviour of this block-bootstrap method under an assumption of non-normality is considered.

1.6 Summary of chapters

1. **Chapter 2** is an exercise of applied work whose theoretical background is in the context of generalized additive models and the ordinary bootstrap.

2. **Chapter 3** investigates a proposal of a test statistic for homoscedasticity in a linear model. Kernel smoothers and the distribution of quadratic forms are used in this test. Graphical tools are also proposed.

3. **Chapter 4** investigates a proposal for a test for constant variogram for a spatial process under the assumption of second order stationarity and isotropy. Different versions for the test are considered. A simulation study to calculate the power and size of the test is performed under different options for the calculation of the distribution of the test statistic. Reference bands as graphical tools are proposed and applications to examples are given.

4. **Chapter 5** reviews methods of simulation and re-sampling for spatial

data and makes new proposals. A particular case of the block-bootstrap method in its version of "moving windows" is used to build confidence bands for the variogram.

5. **Chapter 6** is a chapter of reflections about further work and extensions of the methods proposed in this study.

Chapter 2

Analysing the distribution of Mackerel egg biomass

2.1 Introduction

The necessity of modelling the biomass of different species of fish has been a preoccupation of several scientists around the world. The data considered in this chapter refer to the biomass of mackerel found near the costs of Spain, France, Ireland and the United Kingdom. Different models have been studied and compared in order to estimate this biomass.

The parameter of interest here is the "total number of mackerel eggs" which can be defined as:

$$\tau = a \sum_{i=1}^{n} E(D_i) \tag{2.1}$$

where D_i is the density of mackerel-eggs (as above) at the point i and a is constant related to the volume of the set where the total is required to be estimated.

Previous work has identified a small number of variables which are related to the variation in egg density. For a spatial approach it seems to be convenient to model the response variable (density of eggs or a transformation of it) as an additive function of some relevant covariates. Parametric models do not seem to be flexible enough to capture the relationships with some covariates such as bottom depth. In this context, the generalised additive models considered by Hastie and Tibshirani (1990) offer an attractive tool. Some of these previous works are described in section 2.2.

The data were collected at one particular time, which led to this study being carried out with a simple spatial model. It was also necessary to carry out some initial manipulation of the data. Large numbers of zero-values for the density of eggs is not unusual in samples of this type. The geographical position of the zero-valued points can easily change if the sample is collected at different times, as is explained later. Also, the different countries involved used different methodologies. This results in very different densities of points in some sub-areas of the region. For these reasons the original sampled points were transformed to a regular grid and the points corresponding to zero-values for egg density of mackerel were omitted. The way in which the sample is manipulated before fitting a model to them is explained in section 2.3.

The use of generalised additive models (GAM's) is not new in the literature on estimation of fish-biomass. However, some previous works have tackled the problem with models in which the functional dependence of the response variable on covariates such as latitude and longitude is rather complex. Some general background and the notation is outlined in section 2.4.

A natural model which uses the variables latitude and longitude pooled together in a relationship which is thought of as a smooth combination of both is explored. Because of the lack of inference theory for comparing different nested plausible models, a descriptive methodology has been used, by analogy with the tools from generalised linear models theory. The selection of an adequate generalised linear model for the variable log density of eggs is analysed in section 2.5.

The descriptive approach for model selection is then checked with tools provide by the bootstrap environment. In fact, the distributional theory for estimators in a G.A.M has still not been developed even when there are assumptions of normality on the variables involved. Hence, an estimated value for the total biomass of mackerel in the sampled area has been calculated by bootstrapping the distributions of this total. This has been carried out for a variety of generalised additive models and the results compared. These models are compared with a previous model due to Borchers et al. (1994). The results are shown in section 2.6 .

Some conclusions and comparisons with a previous model are outlined in section 2.7.

2.2 Previous literature

A large proportion of zero-values is a common situation in fish surveys and especially in multi-species fish surveys. In fact, sometimes the area in which a specific species is to be found generally is not well known and the sample may overlap a considerable part of unoccupied "habitat". However, dealing with the zero-values separately may lead to inefficient estimators of abundance because a suitable habitat might change from time to time for different reasons or an area may be unoccupied simply because of a low population level. These ideas motived the work of Pennington (1983) on the the problem of the biomass estimation of Mackerel. It is assumed that X is a random variable so that, $Pr(X \neq 0) = p$, $E(X|X \neq 0) = \mu$, $var(X|X \neq 0) = \sigma^2$ and $E(X) = \alpha$, $var(X) = \beta$.

Let $x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n$ be a sample from X where the first m values are non-zero. It was shown by Aitchison (1955) that if for $m > 0$, $a_{(m)}$, $e_{(m)}$ and $f_{(m)}$ are unbiased estimators for μ, μ^2 and σ^2, respectively, then

$$c = \frac{m}{n} a_{(m)} \quad \text{if} \quad m > 0 \quad \text{and} \quad c = 0 \quad \text{if} \quad m = 0$$

$$\text{and} \quad d = \frac{m}{n} f_{(m)} + \frac{m}{n}\left(\frac{n-m}{n-1}\right) e_{(m)} \quad \text{if} \quad m > 0 \quad \text{and} \quad d = 0 \quad \text{if} \quad m = 0$$

are unbiased estimators of α and β respectively. One parameter of interest in the applications is, in fact, var(c) and, hence, a good estimator of this parameter is very useful. Using an unbiased estimator of the sample mean for the non-zero values, a minimum variance unbiased estimator of var(c) is given in this paper. Pennington applied these ideas to the case in which the distribution of the random variable X conditioned on the non-zero values is

log-normal. In this context the distribution of X is called the Δ distribution and expressions for c, d, and a minimum variance unbiased estimator for var(c) were calculated and applied to estimate the total egg production of Atlantic mackerel. The total estimate of spawning stock size based in the total egg production compared favourably with other estimates using other methodologies. This methodology was adopted by the 1993 Mackerel/Horse Mackerel Egg Production Workshop.

Another approach to biomass estimation is to model the egg density as a function of some covariates. Some of the ideas underlying this approach are that at least some of the variation in the egg density is due to the variation in covariates (such as latitude, longitude, time, bottom depth, etc) and these models provide an objective mean of interpolating into UN-sampled areas. In this context, the works by Pope and Woolner (1985) and Bochers et al. (1994) should be mentioned.

Pope and Woolner (1985) fitted quadratic response surfaces to latitude, longitude and time. They combined a non-parametric estimate of mean egg abundance with a variance estimator based on the assumption of a log-normal distribution for egg density.

Borchers et al. (1994) considered the egg density (for Mackerel) as a function of latitude, longitude, bottom-depth and distance from 200m contour. In this work, the emphasis is on the selection of a good model which focuses on its efficiency as a predictor model and its simplicity. They used the tool of the generalised additive model as proposed by Hastie and Tibshirani (1990). The log of density is expressed as an additive function of these covariates. This

model is:

$$log(D) = \beta_0 + S(lat) + S(lon) + S(BDp) + S(D200) + \beta_1(lat.lon) + \beta_2(lat.BDp) + \varepsilon$$

where lat is the latitude, lon is the longitude, D200 is the distance from 200 m contour and BDp is the bottom depth. S is a spline smoother and β_0 and β_1 are constants. An analysis of the deviance for the estimated model was performed. The distribution of errors ε is assumed to be over-dispersed Poisson. The variability of the total estimated is studied via the parametric bootstrap. It was assumed that for each point i in the sample grid, $Y_i = log(D_i)$ is an over-dispersed Poisson random variable with mean equal to the modelled estimated abundance in this point. The residuals r_i are also over-dispersed Poisson random variables with cumulative distribution function F_i. Then, the transformed residuals $R_i = F_i(r_i)$ are identically distributed (uniform if there is no over-dispersion) random variables and a bootstrap methodology can be applied in order to estimate the variance of the estimate. The steps in the paper can be written as follows:

1. Calculate the residuals $r_i = y_i - \hat{y}_i$ for each point i in the sampled area.
2. Transform the residuals by using their cumulative distributions to obtain identically distributed (uniform if there is no over-dispersion) random variables $R_i = F_i(r_i)$.
3. Generate values R_i^* from the distribution F_i (a permutation of the values R_i and a corresponding assignment to points in the sampled area is enough).
4. Calculate back transformed values by using the inverse function of F_i,

e.g., $r_i^* = F_i^{-1}(R_i^*)$

5. Obtain a bootstrap estimation $\hat{y}_i{}^*$ by adding r_i^* to the estimated value $\hat{y}_i$, $\hat{y}_i{}^* = \hat{y}_i + r_i^*$

6. Repeat steps 3 to 5 a reasonable number of times.

7. Calculate a bootstrap estimation of the variance of $\hat{y}_i$ with the bootstrap values $\hat{y}_i{}^*$.

The bootstrap methodology as described above was adjusted because of numerical problems with the behaviour of the distributions of residuals and, as a consequence, it led to a high and non-trivial computational effort. The zero-values for density were not considered in this model.

Even though this model is useful as a tool of exploratory analysis, the way in which the covariates are introduced is rather complex. As was mentioned in the introduction, tools provided by G.A.M.'s are considered here in order to build a more suitable model. The lack of the corresponding distributional theory for G.A.M. is also addressed by an appropriate bootstrap technique. A confidence interval for the biomass of egg mackerel is calculated from the empirical distributions of the total biomass as estimated with different models.

2.3 The data

The data considered for the study here were collected by Dr I.G. Priede Dept. of Zoology, University of Aberdeen, U.K.. They were obtained with the

assistance of Professor Steve Buckland of the Wildlife Population Assessment Research Group, University of St Andrews,U.K. The observed sample is a set of five variables: density of eggs, longitude, latitude, bottom depth and distance from 200m contour observed in 634 sample units.

The geographical extension of the sample is on a band between 44.25 and 57.75 northern latitude and 1.25 and 14.75 western longitude. This band covers part of the coasts of Spain, France, Ireland and Great Britain as is shown in Figure 2.1.

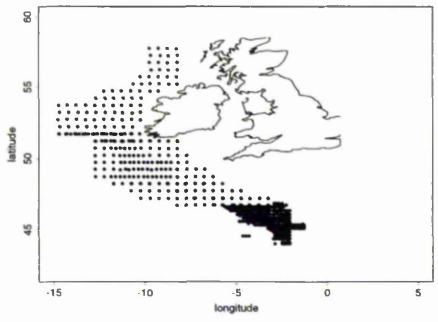

Figure 2.1. The points indicates places where values for mackerel egg density and values for others covariates as bottom depth and distance from 200m contour were taken.

The variable "distance from 200m contour" is the shortest distance, in metres, from the sampled points to a line drawn to follow the depth of 200m. It is negative if the sampled point is between the coast and the 200m contour line and positive if it is on the opposite side. Its values are between -173km and 165.8km with an average of 48km, a median of 5.15km and a standard deviation of 44.27km.

Bottom Depth is the depth (in meters) at the sampled point. Its values are in the range of 42m to 4450m with an average of 836.378, a median of 171.5m and a standard deviation of 128.6m for this sample.

The response variable is the density of mackerel eggs. Its values were obtained as a conversion of observed eggs at a specific stage into daily egg production. Its values are in the range of 0 and 601.708 with an average of 38.22, a median of 5.125 and a standard deviation of 2.8. A transformation to the logarithm of this variable is considered for the statistical analysis to normalise this variable as advised in previous studies.

2.3.1 The zero-valued data

A suitable habitat for mobile populations may change from time to time due to many factors including the timing of the survey. Also an area may be unoccupied simply because of a low population level (Pennington (1983)).

In this study the zero-valued data were concentrated mostly in the areas corresponding to the lowest latitudes. As shown in the Figure 2.2, they were in areas at the border of non-zero values. From 634 points in the original sample, 265 had a zero-value for the density of eggs. However, these values were mostly concentrated in a small area compared with the whole sampled area (Figure 2.2).

This geographical distribution, the considerably large number of points with this feature, and the snapshot time characteristic of the sample led to the decision of ignoring the zero-values data at least in a preliminary approach.

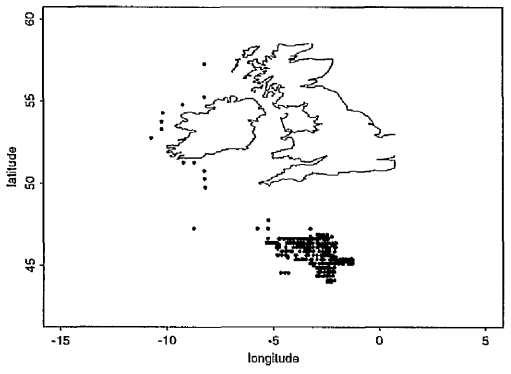

Figure 2.2. The points correspond to places with zero value for mackerel egg density

2.3.2 The sample grid

The original sample grid (Figure 2.1) has been made from samples taken under different criteria depending on the country which designed it. In the areas near to Spain the density of points is considerable higher. This is also the area with the highest density of zero-values for the density of eggs and it is, jointly with other smaller areas, on the border of the whole region under consideration.

These reasons led to the construction of a regular grid in order to simplify the treatment and analysis of the data.

In this context "regular grid" means a grid in which the absolute difference

between two consecutive values of latitude is 0.5 degrees and the same rule is used for longitude (Figure 2.3). The average values of the covariates and the response variable were considered in each cell of the regularized grid. The new extreme values for the response variable were 0.6 and 365.88 for the minimum and maximum respectively. This transformation of the data removes the out-lier of 601.708 for this variable.

The distribution of the variable "density of eggs" is approximately log-normal. This led to the log-transformation of this variable.

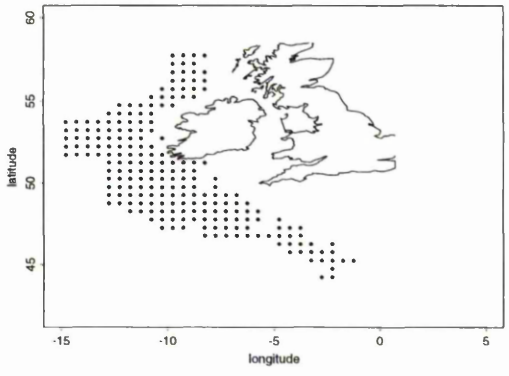

Figure 2.3. The regular grid built from equally spaced values for latitude and longitude

2.4 A model for density of mackerel-egg

2.4.1 Introduction

To search for an estimator of the parameter "total number of egg mackerel" as defined in 2.1, the model for the random variable "density of mackerel-eggs" was assumed as:

$$D_i = d(\boldsymbol{x}_i)\eta_i, \quad i = 1, \ldots, n \tag{2.2}$$

where d is a function of the covariates latitude (lat), longitude (long), bottom depth (BDp) and distance to 200m contour (D200), $\boldsymbol{x}_i = (lat_i, long_i, BDp_i, D200_i)$, $i = 1, \ldots, n$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)$ is a vector of independent and identically distributed log-normal variables η_i with a constant mean μ and constant variance ν. This model can be written as:

$$y_i = log(D_i) = log(d(\boldsymbol{x}_i)) + log(\eta_i), \quad i = 1, \ldots, n \tag{2.3}$$

or $y_i = f(\boldsymbol{x}_i) + \varepsilon_i$ where $y_i = log(D_i)$, $f(\boldsymbol{x}_i) = log(d(\boldsymbol{x}_i))$ and $\varepsilon_i = log(\eta_i)$, $\imath = 1, \ldots, n$ is a zero-mean normal variable with variance $\sigma^2 = 2ln(\mu)$. For the function f an additive model of the form

$$f(lat, long, BDp, D200) = f_1(lat, long) + f_2(BDp) + f_3(D200)$$

seems to be adequate according to biological and historical reasons (see section

2.2).

Under these assumptions, the parameter τ to be estimated, or more precisely $\tau_0 = \frac{\tau}{a}$ (given that a is known), can be expressed as:

$$\tau_0 = \sum_{i=1}^{n} E(D_i) = \sum_{i=1}^{n} d(\boldsymbol{x}_i) E(\eta_i) = \sum_{i=1}^{n} d(\boldsymbol{x}_i)\mu \tag{2.4}$$

With this approach in view, tools from the Generalised Additive Models will be employed.

A natural estimator for τ_0 is:

$$T_0 = \hat{\tau_0} = \sum_{i=1}^{n} exp(\hat{Y}_i) = \sum_{i=1}^{n} exp(\hat{log}(D_i)) \tag{2.5}$$

where $\hat{Y}_i$ is the fitted value from the model to be chosen.

Consequently, some general ideas about Generalised Additive Models will be considered now and then the challenge to build an appropriate generalised additive model will be faced.

2.4.2 Generalised Additive Models (G.A.M.)

Biological reasons lead naturally to the consideration of the density of eggs as a function of covariates such longitude and latitude, bottom depth and distance from a line in the contour map. In this context, the generalised additive model as developed by Hastie and Tibshirani (1989) offers an attractive tool because

it is possible to build a very general model under very general assumptions. Nevertheless, there are some different ways to cope with the idea of building a generalised additive model.

The Generalised Additive Model as proposed by Hastie and Tibshirani (1989) can be written as:

$$y_j = \sum_{i=1}^{k} f_i(x_{ij}) + \varepsilon_j, \quad j = 1, \ldots, n$$

where μ is an unknown parameter, f_i is a smooth function of the covariate (vector of covariates) X_i, x_{ij} is the value of the variable X_i at the point j $i = 1, \ldots, k$, and ε_j is a random variable with zero-mean, $j = 1, \ldots, n$. Therefore, in terms of mean values, and under the assumptions that the covariates are non-random (otherwise, expectation should be replaced by conditioned expectations) this model can be written as:

$$E(Y_j) = \sum_{i=1}^{k} f_i(x_{ij})$$

Within this general form of the model, estimated values for the functions f_i can be calculated. One of these numerical methodologies is the widely known back-fitting algorithm. Different approaches to the underlying theory for this algorithm can be considered. One of these deals with the idea of minimisation of the expression:

$$E(Y - g(\boldsymbol{X}))^2$$

over the space of the functions $g(\boldsymbol{X})$, where $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$ so that

$g(\boldsymbol{X}) = \sum_{i=1}^{m} f_i(X_i)$ belongs to the space $H = H_1 + H_2 + \cdots + H_m$ and f_i to $H_i \quad i = 1, \ldots, m$. Each set H_i is the space of the functions $\Phi_i(X_i)$ with expected value zero, finite second-order moment and an inner product defined for each pair of functions as the expected value of its product. The minimum exists and is unique because H is a closed space, but the terms $f_i(X_i)$ may not be uniquely determined. As in the case of ordinary linear regression, one way to find a solution is through a characterisation of residuals $Y_j - \sum_{i=1}^{m} f(x_{ij}), \quad j = 1, \ldots, n$. In fact, the observed vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ does not belong to H because of the error ε_j $(Y_j = \sum_{i=1}^{k} f_i(x_{ij}) + \varepsilon_j), \quad , j = 1, \ldots, n$ but it is possible to search for a solution $\hat{g}(\boldsymbol{X}) = \sum_{i=1}^{m} \hat{f}_i(X_i)$ in H, the orthogonal projection of Y. Because H is generated by H_j, $j = 1, \ldots, m$, $Y - \hat{g}(\boldsymbol{X})$ should be orthogonal to H_j, for $j = 1, \ldots, m$. It implies that the projection P_j of the residual vector $\boldsymbol{Y} - \hat{g}(\boldsymbol{X})$ for $j = 1, \ldots, m$ should be the null vector. These results can be written as:

$$P_j(\boldsymbol{Y} - \hat{g}(\boldsymbol{X})) = P_j(\boldsymbol{Y} - \sum_{i=1}^{m} \hat{f}_i(x_i)) = \boldsymbol{o_j}, \quad \text{for} \quad j = 1, \ldots, m$$

where $\boldsymbol{o_j}$ is the null vector in H_j. and, the equality $P_j(\hat{f}_j(x_j)) = \hat{f}_j(x_j)$ implies

$$\hat{f}_j(x_j) = P_j(\boldsymbol{Y} - \sum_{i \neq j} \hat{f}_i(x_i)), \quad \text{for} \quad j = 1, \ldots, m$$

At this point it is necessary to define a projection $P_j, \quad j = 1, \ldots, m$. One method suggested by Hastie and Tibshirani (1990) is to represent a projection

P_j by a linear smoother S_j. This leads to a system:

$$\begin{bmatrix} I & S_1 & S_1 & \dots & S_1 \\ S_2 & I & S_2 & \dots & S_2 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ S_m & S_m & S_m & \dots & I \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{f}}_1 \\ \hat{\boldsymbol{f}}_2 \\ . \\ . \\ \hat{\boldsymbol{f}}_m \end{bmatrix} = \begin{bmatrix} S_1\boldsymbol{Y} \\ S_2\boldsymbol{Y} \\ . \\ . \\ S_m\boldsymbol{Y} \end{bmatrix}$$

where S_j, for $j = 1, \dots, m$ is the matrix for the selected linear smoother. Therefore, if the sample size is n, the matrix of the system above is an (nmxnm)-matrix. This system could be thought as a generalisation of the system which leads to the normal equations in a linear regression model (Buja, Hastie and Tibshirani (1989)). Moreover, this system has the same shape corresponding to the widely known method of Gauss-Seidel for an algebraic system of equations and this idea has been used in order to find a "good" numerical solution for the said system. The algorithm can be formalised as follows:

1. Consider initial vectors $\hat{\boldsymbol{f}}_j^{(0)} = (f_j(x_{1j}), f_j(x_{2j}), \dots, f_j(x_{nj}))^T$, for $j = 1, \dots m$.
2. Calculate the vector $\hat{\boldsymbol{f}}_j^{(l)} = S_j(\boldsymbol{Y} - \sum_{i \neq j} \hat{\boldsymbol{f}}_i^{(l-1)})$ for $j = 1, \dots, m$.
3. Repeat 2 from l=1 to the value of l for which the numerical vectors converge.

It is possible to prove that the system above is consistent under quite wide assumptions for the smoothers S_i, $i = 1, \dots, m$ (see Buja, Hastie and Tibshirani (1989)).

The linear smoothers used to model the data here are smoothing splines and locally-weighted running-line smoothers (loess). A linear smoother can be defined as follow:

Consider data of the form (x_i, y_i), $i = 1, \ldots, n$ if $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ and suppose the aim is to estimate locally, at a point x_0, the dependency of $\boldsymbol{y}$ on $\boldsymbol{x}$. This dependency can be expressed, for a fixed point x_0 and a fixed vector $\boldsymbol{x}$ as a function $L_{x_0,\boldsymbol{x}}$ which is defined for $\boldsymbol{y}$ in some set. When this function is linear, the smoother is called a "linear smoother" and its values can be expressed as: $L_{x_0,\boldsymbol{x}}(\boldsymbol{y}) = S\boldsymbol{y}$ for a matrix S which does not depend on $\boldsymbol{y}$.

1. Smoothing splines

   These splines are usually characterised as the solution of the minimisation of the functional:

$$\wp(g) = \sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int_{-\infty}^{+\infty} [g"(u)]^2 du$$

   Hence a spline smoother, for a fixed value of the constant λ, is the function L in the Sobolev space W of functions g with absolutely continuous derivative and squared-integrable second derivative, such that $\wp(L) = \min_{g \epsilon W} \wp(g)$. The parameter λ is concerned with the "amount of smoothing" for the spline. For $\lambda = 0$, the solution is any interpolating line and for $\lambda = \infty$ the solution is the least squares regression line. These kind of splines are also linear and it can be proved that the associated matrix S is $S = (I + \lambda K)^{-1}$ where $K = A^T C^{-1} A$. A and C are tridiagonal matrices. If $h_i = x_{i+1} - x_i$ for $i = 1, \ldots, n$, A is a

(n-2)xn matrix with entries $a_{ii} = 1/h_i$, $a_{i,i+1} = -(1/h_i + 1/h_{i+1})$, and $a_{i,i+2} = 1/h_{i+1}$, and C is a (n-2)x(n-2) symmetric matrix with entries $c_{i-1,i} = c_{i,i-1} = h_i/6$ and $c_{ii} = (h_i + h_{i+1})/3$.

2. Locally-weighted running-line smoothers (loess)

These smoothers can be considered as a particular case of those which are usually referred in the literature as "kernel smoothers". The kernel smoothers are smoothers such that, the ij-entry, say s_{ij} in the matrix S are simple functions of weights $w(x_0; \boldsymbol{x})$ which decrease as a function of $|x_0 - x_i|$, where $x_i, i = 1, \ldots, n$ are the components of the vector $\boldsymbol{x}$. These weights are generally called "kernel functions".

Several kinds of kernel functions can be found in the specific literature and a good choice of them is usually concerned with the context. An example considered several times in this thesis is that concerning with kernel functions of the form

$$w_i(x_0; \boldsymbol{x}) = \frac{\exp\left(-0.5\left(\frac{x_0 - x_i}{b}\right)^2\right)}{\sum_{j=1}^{n} \exp\left(-0.5\left(\frac{x_0 - x_j}{b}\right)^2\right)}$$

The locally-weighted running smoothers (loess) combine the idea of dealing with the density of points close to the target point and the smoothness features of the kernel smoothers. Let the target point be x_0, k a positive integer number to define the number of nearest neighbours of x_0, $N(x_0)$ the set of these k nearest neighbours and $\Delta(x_0)$ the maximum distance from each point in $N(x_0)$ to x_0. The weight w_i is

assigned to each point x_i using the function:

$$w_i(x_0; \boldsymbol{x}) = \begin{cases} \frac{\left[1-\left(\frac{\|x_0-x_i\|}{\Delta(x_0)}\right)^3\right]^3}{c} & \text{if } \|x_0 - x_i\| > \Delta(x_0) \\ 0 & \text{otherwise} \end{cases}$$

where c is a normalising constant.

These weights then define the smoothing matrix as described in 1 above. The loess smoothers are easily applied to more than one variable problems as in the case studied here.

If a smoother can be expressed as $L_{x_0,\boldsymbol{x}}(\boldsymbol{y}) = S\boldsymbol{y}$ (linear smoother), its **degrees of freedom** is defined as:

$$df = tr(S) = \sum_{i=1}^{n} \lambda_i \tag{2.6}$$

The G.A.M. is implemented in the Splus software environment with smoothing splines and locally-weighted running-line smoothers, and with the possibility of extension to other smoothers. Because of its flexibility to extend to more than one dimension, the loess smoothers offer a very useful tool in this work. Nevertheless models which involve other smoothers were also built.

One of the difficulties with the selection of an appropriate generalised additive model involving smoothers is dealing with their degrees of freedom. Figure 2.4 shows plots of smoothing splines with different degrees of freedom for bottom depth (BDp). For smoothing splines the higher the value of the degrees of freedom the more flexible the result. The notation $s(x, l)$ will be used to indicate "smoothing spline on the variable x with l degrees of freedom" For

loess smoothers the degrees of freedom are more complicated to interpret but they are quite flexible and have the very important possibility to be used for more than one dimensional covariates.

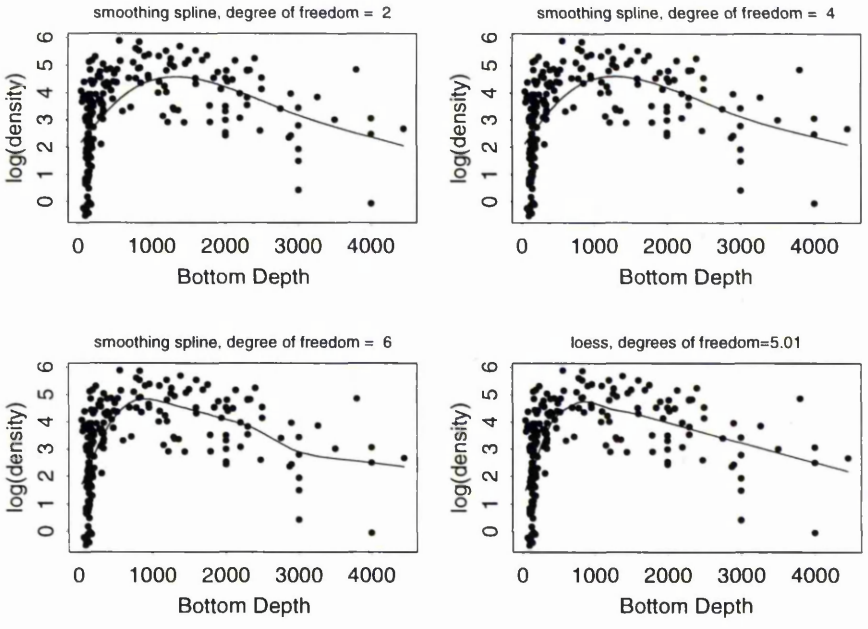

Figure 2.4. The fitted values of log(density of mackerel) to G.A.M. models where the only covariate is bottom depth are plotted against this covariate. Different smoother was used in each panel.

Figure 2.5 shows similar plots to those in Figure 2.4 but for the variable distance from 200m contour. The plot on the right bottom corner of the Figure 2.4 seems to be very similar to other alternative spline smoothers. This graphical exploration supports the decision of considering a loess for a term in the model for the covariate bottom depth. An analogous conclusion can be drawn from the Figure 2.5 for the variable distance from 200m contour.

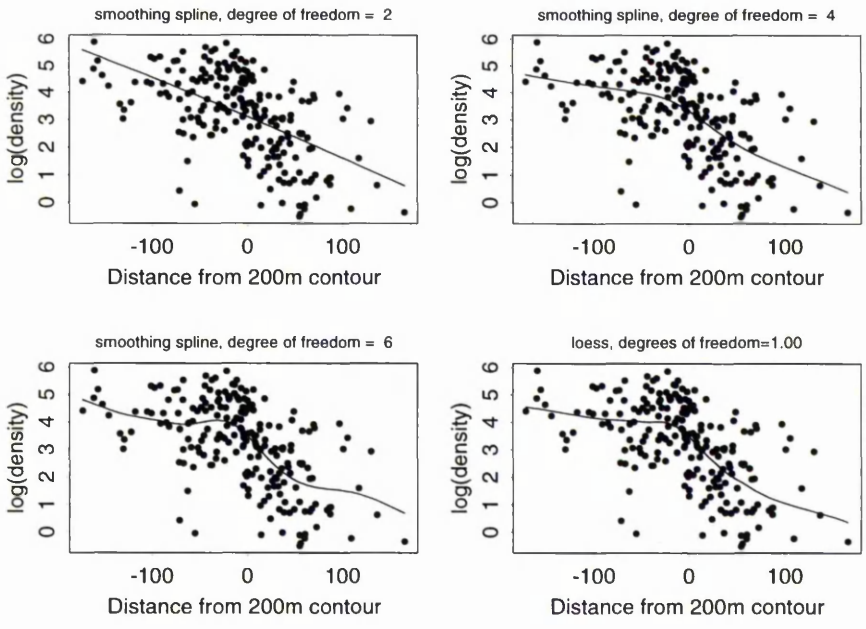

Figure 2.5. The fitted values of log(density of mackerel egg) to G.A.M. models where the only covariate is distance from 200m contour are plotted against this covariate. Different smoother was used in each panel.

The variables latitude and longitude are considered jointly below in one term of an appropriate additive model. Hence, in this case the most convenient smoother to resort to is loess. The selection of loess as the smoother for each term in the model will also simplify the interpretation of the selected additive model.

2.5 Selection of models

In the context of generalised additive models, the response variable of interest here is density of mackerel eggs. A plot of the logarithm of this variable is shown in Figure 2.6 for each spatial location on the regularized grid. This

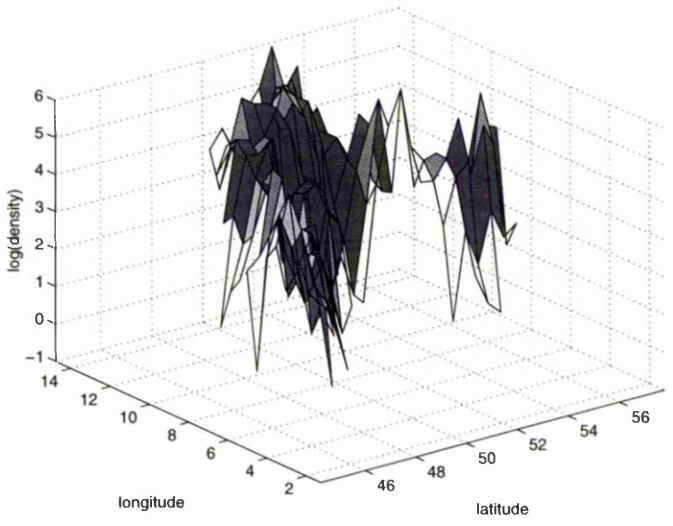

Figure 2.6. The observed values of log(density of mackerel eggs) at each point on the regularized grid.

study was started by exploring graphically the dependency of the response variable on each of the mentioned covariates under the assumption that the mackerel data come from a population in which the log(density) of eggs can be modelled as a sum of functions of the covariates latitude, longitude, bottom-depth and distance from 200 m. contour, plus an error.

Figure 2.7 shows the values of log(density) replaced by those from the loess. In fact, it is necessary to consider the dependence of log(density) on the covariates latitude and longitude jointly rather than separately. The loess

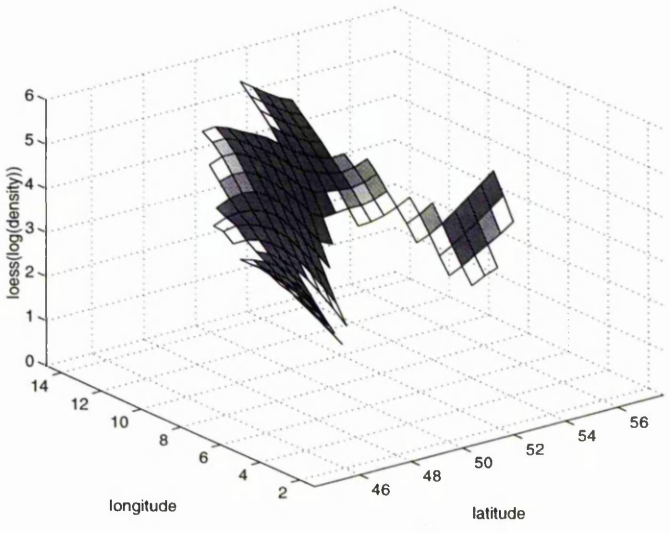

Figure 2.7. The values of log(density of mackerel eggs) fitted to the Generalised Additive model (G.A.M.) $\widehat{log}(density) = loess(latitude, longitude)$ at each point on the regularized grid .

smoother offers possible means of doing this, allowing the use of a term in a generalised additive model.

The selection of suitable smooth terms for the model is carried out by analysis of deviance. This, as suggested by Hastie and Tibshirani (1990), is based on the analogous approach for the generalised linear models. The most general ideas are introduced briefly now. In this work we have tackled this problem by using the analysis of the deviances as an approximation to compare nested models and a simulation methodology in order to check different models is used as suggested by Hastie and Tibshirani (1990) by analogy with generalised linear models.

In fact, it is widely known that the discrepancy or goodness of fit of a model can be assessed by the deviance. This statistic is based on the comparison

between the estimated model whose goodness of fit is under investigation and the saturated model which is, intuitively speaking, the model with the maximum number of parameters allowed by the data. The deviance for the estimated model η is defined as:

$$D(\boldsymbol{y}, \hat{\eta}) = 2\left(l(\eta_{max}, \boldsymbol{y}) - l(\hat{\eta}, \boldsymbol{y})\right)$$

where $l(\eta_{max}, \boldsymbol{y})$ is the maximum log-likelihood achievable for the given data, and $l(\hat{\eta}, \boldsymbol{y})$ is the value of the log-likelihood for the estimated model.

The deviance for normal linear models has an interesting feature which other measures of goodness of fit do not have. This feature is its additivity under maximum likelihood nested models because of the orthogonality of the terms in the models. Under generalised additive models this feature is not necessarily valid but the difference of deviances for two nested models can still be used in order to perform informal tests to compare two nested models with some heuristic justification.

Let η_1 and η_2 now be two nested generalised additive models with η_1 nested within η_2. Under the null hypothesis that the model η_1 is correct the statistic:

$$D(\hat{\eta_2}, \hat{\eta_1}) = D(\boldsymbol{y}, \hat{\eta_1}) - D(\boldsymbol{y}, \hat{\eta_2})$$

would have an asymptotic χ^2 distribution with degrees of freedom equal to the difference in the dimensions of the two models if the generalised linear model theory applied. Although the asymptotic distribution is not χ^2 in the context of G.A.M., it has been shown by Hastie and Tibshirani that this can still be used as an approximation for screening generalised additive nested

models.

The degrees of freedom for the χ^2 distribution here is related to the degrees of freedom for the smoothers involved in the additive model. For the model $y = f + \varepsilon$ if $f = f_1 + f_2 + \ldots + f_k$ and S is the matrix for the smoothers corresponding to the sum f so that $\hat{f} = Sy$, the degrees of freedom for the distribution of a statistic which could be associated with the deviance is:

$$df(error) = n - tr(2S - SS^T)$$

A more useful statistic to compare models is the difference of deviances of two nested models. The distribution of this difference could also be approximated by a χ^2 distribution. If $S_{(j)}$ is the matrix corresponding to S when the j^{th} term is removed from the model so that $\hat{f} = S_{(j)}y$, the degrees of freedom corresponding to the χ^2 distribution for the difference of deviances is:

$$\Delta f_j(error) = tr(2S - SS^T) - tr(2S_{(j)} - S_{(j)}S_{(j)}^T)$$

Therefore, this quantity could be considered as the expected value of the increase in the residual sum of squares (up to a scale factor) if the j^{th} predictor is excluded from the model, assuming its exclusion does not increase the bias. Then, under the assumption that η_1 is nested within η_2, and if the null hypothesis of η_1 correct is true, the increase in the deviance $D(\hat{\eta_2}, \hat{\eta_1})$ when one (or more) variable(s) are cancelled from the model η_2 in order to obtain the model η_1 should be "small" e.g. the p-value should be greater than 0.05. Therefore, the rule to reject the null hypothesis that the model η_1 is correct could be expressed as: Reject the null hypothesis if the observed value of $D(\hat{\eta_2}, \hat{\eta_1})$ is greater than the corresponding quantile of the χ^2 distribution

with the degrees of freedom identified above. In other words, there not enough evidence to remove the predictor f_j from the model if the p-value is greater than 0.05 or some other appropriately chosen value.

In this work, different models involving the covariates latitude, longitude, bottom depth and distance from the two hundred meters contour have been considered and the different possibilities for nested models have been analysed under the criterion described above. Some of these models were investigated in order to explore the effect of the degrees of freedom parameter which is shown as a second argument of each term. The results are shown in Table 2.1. This table shows "significant" differences between the nested models fitted with different smoothing splines. In fact, in all of these models there is not enough evidence to remove one or any of the covariates. For example, between the models,

$$log(density) = s(lat, 6) + s(long, 6) + s(BDp, 6) + s(D200, 6) + \varepsilon$$

and

$$log(density) = s(lat, 6) + s(long, 6) + s(BDp, 6) + \varepsilon$$

the difference in deviances is 32.82 which can be compared with a quantile of the χ^2 distribution with 5 degree of freedom. Indeed, 32.82 is considerably greater than the 95% corresponding quantile of the χ^2 distribution.Then, from the point of view of this methodology by analogy to that built in the context of the generalised linear models, there is not enough evidence to ignore the variable "distance to 200m contour" (D200).

The same conclusion can be extracted from the model (1) above and the

model

$$log(density) = s(lat, 6) + s(long, 6) + s(D200, 6) + \varepsilon$$

In fact, the difference of deviances in this case is 21.51 with 5 degrees of freedom. The last model in the first part of the table can also be compared with models (1), (2) and (3) given that it is sub-model of them. Also in this case the differences of deviances are too big to have any reasons (even by analogy) to accept the last model as the more adequate one.

The five first principal models, and all their possible nested sub-models in table 1 have been built by using spline smoothers with identical degrees of freedom. However, terms in a generalised additive model need not have this feature. To illustrate this, the last model in the table has been constructed with different degrees of freedom for the smoothers. This produces the same results as in the other cases.

2.5.1 A model for density of mackerel eggs

On the question of the selection of a smoother for the variables BDp and D200, figure 2.4 and figure 2.5 show that there is no evidence to prefer a smoothing spline instead of a loess smoother. Since the latter is the only convenient choice for the case of two-dimensional covariates, it is particularly convenient to use loess in the one dimensional case too.

ANALYSIS OF GENERALISED ADDITIVE MODELS												
model									deviance	df	D(dev)	D(df)
$\widehat{log}$(density)	=	s(lat,6)	+	s(long,6)	+	s(BDp,6)	+	s(D200,6)	176.82	20		
df		5.00		5.00		5.00		5.00				
chisq		29.71		48.38		39.65		59.01				
$\widehat{log}$(density)	=	s(lat,6)	+	s(long,6)	+	s(BDp,6)			209.65	15	32.82	5
df		5.00		5.00		5.00						
chisq		28.09		47.49		118.28						
$\widehat{log}$(density)	=	s(lat,6)	+	s(long,6)	+	s(D200,6)			198.34	15	21.51	5
df		5.00		5.00		5.00						
chisq		43.58		80.43		117.84						
$\widehat{log}$(density)	=	s(lat,6)	+	s(long,6)					295.83	10	119.00	10
df		5.00		5.00								
chisq		60.80		72.39								
$\widehat{log}$(density)	=	s(lat,5)	+	s(long,5)	+	s(BDp,5)	+	s(D200,5)	187.96	16		
df		4.00		4.00		4.00		4.00				
chisq		22.88		41.04		42.42		51.85				
$\widehat{log}$(density)	=	s(lat,5)	+	s(long,5)	+	s(BDp,5)			221.30	12	33.34	4
df		4.00		4.00		4.00						
chisq		20.44		40.23		107.50						
$\widehat{log}$(density)	=	s(lat,5)	+	s(long,5)	+	s(D200,5)			211.52	12	23.56	4
df		4.00		4.00		4.00						
chisq		36.54		68.85		104.45						
$\widehat{log}$(density)	=	s(lat,5)	+	s(long,5)					303.01	8	115.05	8
df		4.00		4.00								
chisq		53.19		67.32								
$\widehat{log}$(density)	=	s(lat,4)	+	s(long,4)	+	s(BDp,4)	+	s(D200,4)	202.94	12		
df		3.00		3.00		3.00		3.00				
chisq		14.45		29.42		46.55		40.53				
$\widehat{log}$(density)	=	s(lat,4)	+	s(long,4)	+	s(BDp,4)			234.64	9	31.70	3
df		3.00		3.00		3.00						
chisq		13.71		32.55		97.18						
$\widehat{log}$(density)	=	s(lat,4)	+	s(long,4)	+	s(D200,4)			231.55	9	28.61	3
df		3.00		3.00		3.00						
chisq		25.80		49.72		80.98						
$\widehat{log}$(density)	=	s(lat,4)	+	s(long,4)					311.84	6	108.90	6
df		3.00		3.00								
chisq		44.41		57.77								
$\widehat{log}$(density)	=	s(lat,3)	+	s(long,3)	+	s(BDp,3)	+	s(D200,3)	223.67	8		
df		2.00		2.00		2.00		2.00				
chisq		6.21		16.21		49.60		28.00				
$\widehat{log}$(density)	=	s(lat,3)	+	s(long,3)	+	s(BDp,3)			250.55	6	26.88	2
df		2.00		2.00		2.00						
chisq		8.91		25.26		88.85						
$\widehat{log}$(density)	=	s(lat,3)	+	s(long,3)	+	s(D200,3)			260.54	6	36.87	2
df		2.00		2.00		2.00						
chisq		12.48		28.32		52.00						
$\widehat{log}$(density)	=	s(lat,3)	+	s(long,3)					323.94	4	100.27	4
df		2.00		2.00								
chisq		32.25		43.69								
$\widehat{log}$(density)	=	s(lat,2)	+	s(long,2)	+	s(BDp,2)	+	s(D200,2)	242.29	4.9		
df		1.00		1.00		1.9		1.00				
chisq		1.31		6.63		58.96		14.98				
$\widehat{log}$(density)	=	s(lat,2)	+	s(long,2)	+	s(BDp,2)			261.13	3.9	18.84	1
df		1.00		1.00		1.90						
chisq		5.29		16.28		97.00						
$\widehat{log}$(density)	=	s(lat,2)	+	s(long,2)	+	s(D200,2)			294.23	3	51.94	1.9
df		1.00		1.00		1.00						
chisq		2.82		11.99		26.65						
$\widehat{log}$(density)	=	s(lat,2)	+	s(long,2)					345.92	2	103.63	2
df		1.00		1.00								
chisq		24.05		31.15								
$\widehat{log}$(density)	=	lo(lat,long)	+	s(BDp,6)	+	s(D200,2)			207.95	11.7		
df		5.68		5.00		1.00						
chisq		29.13		68.24		9.66						
$\widehat{log}$(density)	=	lo(lat,long)	+	s(D200,2)					256.06	6.7	49.01	5
df		5.68		1.00								
chisq		53.36		23.97								
$\widehat{log}$(density)	=	lo(lat,long)	+	s(BDp,6)					214.59	10.7	41.47	1
df		5.68		5.01								
chisq		45.67		91.76								
$\widehat{log}$(density)	=	lo(lat,long)							279.30	5.7	71.8	6
df		5.68										
chisq		126.30										

Table 2.1. Analysis of deviances for different generalised additive models (G.A.M.) with the common response variable log(density of mackerel eggs) (log(density) and some or all the covariances latitude (lat), longitude (long), bottom depth (BDp) and distance from 200m contour (D200).

In choosing an appropriate model it is important to consider the context of the data and associated non-statistical criteria. In fact, there are very clear intuitive reasons for building a model in which the variables latitude and longitude are pooled together. On the other hand, there are no reasons not to consider the variables BDp and D200 as covariates for an appropriate model. Following this idea, the model involving the biggest number of covariates considered here was:

$$log(D) = \beta_0 + lo(lat, lon) + lo(BDp) + lo(D200) + \varepsilon$$

where lo is the locally-weighted running-line smoother (loess).

Results about deviances, degrees of freedom, difference of deviances and differences of degree of freedom are shown for different nested models in table 2.2.

ANALYSIS OF GENERALISED ADDITIVE NESTED POSSIBLE MODELS									
model							deviance	df	D(dev)	D(df)
$\hat{log}$(density)	=	lo(lat,long)	+	lo(BDp)	+	lo(D200)	193.65	13.4		
df		5.68		4.00		3.70				
chisq		35.48		34.67		45.73				
$\hat{log}$(density)	=	lo(lat,long)	+	lo(D200)			210.75	9.1	17.10	4.01
df		5.68		3.70						
chisq		61.74		95.66						
$\hat{log}$(density)	=	lo(lat,long)	+	lo(BDp)			212.88	9.7	19.23	3.70
df		5.68		4.00						
chisq		45.68		96.99						
$\hat{log}$(density)	=	lo(lat,long)					279.01	5.7	85.36	7.7
df		5.68								
chisq		126.36								

Table 2.2. Analysis of deviances of the generalised additive model log(density of mackerel eggs)=loess(latitude,longitude)+loess(bottom depth)+loess(distance form 200m contour)+ε, and all of its possible nested models.

From Table 2.2 there is not enough evidence to remove covariates from the initial model. Between the models in the first and second rows in this table, the difference of deviances is 19.23 with 3.7 degrees of freedom which is indeed

too high to drop the term lo(D200) from the first model even in an approximate approach. Similar conclusions can be derived from the term involving the covariate bottom depth for which the difference of deviances is 17.10 with approximated 4 degrees of freedom for a χ^2 distribution.

Following the analogous methodology for generalised linear models, there is not enough evidence to remove variables from the model in the first row of Table 2. It could be argued that the variables bottom depth (BDp) and distance to two hundred meters contour (D200) are themselves functionally dependent on longitude and latitude and therefore these variables may not contribute much additional information. However, these variables do clearly contain highly relevant information on egg density and so they should be included.

2.5.2 Difficulties with GAM inference

The difficulty concerned with the selection of the degrees of freedom for each smoother in the model was simplified here with the use of loess for each covariate. As explained in section 2.4, the selection of this smoother contributes to the homogeneity of the terms of the model. The previous analysis performed for different possible nested generalised additive models is indeed based mostly on analogous methodologies for generalised linear models. In fact, the distribution of the deviance may not be χ^2 even asymptotically.

The lack of inferential tools for specific estimators and their distributions leads to a search for alternative approaches to check the validity of this kind of model. In this case the aim is the estimation of the total number of mackerel

eggs. Consequently, an appropriate methodology should take into account the efficiency of the model regarding this objective. A very attractive tool to resort to, because of its generality, is the bootstrap.

2.6 Bootstrap Analysis of total

The analysis of the estimated value for the total number of mackerel eggs was carried out by using a bootstrap technique. In fact, the lack of distributional theory for estimators in a context of generalised additive models, even under normality assumptions, leads to a search for more general tools (such as the bootstrap) in order to estimate the number of eggs in the area. The selected model and all their possible nested models are checked from this point of view.

Under the assumptions considered here, the parameter τ to be estimated, or more precisely, τ_0 is,

$$\tau_0 = \sum_{i=1}^{n} E(D_i) = \sum_{i=1}^{n} d(\boldsymbol{x}_i)E(\eta_i) = \sum_{i=1}^{n} d(\boldsymbol{x}_i)\mu$$

using the model (2) $D_i = d(\boldsymbol{x}_i)\eta_i \; i = 1, \ldots, n$.

The estimator proposed in this approach is:

$$T_0 = \hat{\tau}_0 = \sum_{i=1}^{n} exp(\hat{Y}_i) = \sum_{i=1}^{n} exp(\hat{log}(D_i))$$

where $\hat{Y}_i$ is the fitted value from the G.A.M.:

$$\hat{Y}_i = lo(lat_i, long_i) + lo(BDp_i) + lo(D200_i)$$

or, simply: $\hat{D}_i = \exp(\hat{Y}_i) = \exp(\hat{f}(\boldsymbol{x}_i))$

It will be assumed that these estimators are invariant under logarithmic and exponential transformations, i.e.,

$$\hat{f}(\boldsymbol{x}_i) = \hat{log}(d(\boldsymbol{x}_i)) = log(\hat{d}(\boldsymbol{x}_i))$$

and

$$E(\hat{f}(\boldsymbol{x}_i)) = E(log(\hat{d}(\boldsymbol{x}_i))) = E(d(\boldsymbol{x}_i))$$

This property was not demonstrated in the context of G.A.M. estimators, but it is enjoyed by smoothers because of their approximately linearity when a Taylor series is considered. Another estimator, the simplest one is the statistic T defined simply as:

$$T = \sum_{i=1}^{n} D_i = \sum_{i=1}^{n} exp(log(D_i)) = \sum_{i=1}^{n} exp(Y_i)$$

A factor for scale correction has been introduced. In fact:

$$E(T_0) = \sum_{i=1}^{n} E(exp(\hat{log(D)}_i)) = \sum_{i=1}^{n} E(exp(log(\hat{d}(\boldsymbol{x}_i)))) = \sum_{i=1}^{n} E(\hat{d}(\boldsymbol{x}_i)) = \sum_{i=1}^{n} d(\boldsymbol{x}_i)$$

assuming that the estimators for the means are unbiased and invariant as explained above.

$$E(T) = \sum_{i=1}^{n} E(exp(Y_i)) = \sum_{i=1}^{n} E(d(\boldsymbol{x}_i)\eta_i)$$

$$= \mu \sum_{i=1}^{n} d(\boldsymbol{x}_i) = \exp(\frac{1}{2}\sigma^2) \sum_{i=1}^{n} d(\boldsymbol{x}_i)$$

because of standard properties of the log-normal distribution. Then, in order to compare the estimators T_0 and T, they will be re-scaled through the constant $\frac{1}{2}\hat{\sigma}^2$.

The distribution of the variable T_0 is then bootstrapped and its empirical distribution is calculated. The observed total number of mackerel eggs t_0 is, indeed, the observed value of the random variable T_0 for the sampled grid (the observed value of T_0 is 8995.38 and it is marked in figure 9).

After the data are fitted to the selected model,

$$log(D_i) = lo(lat_i, long_i) + lo(BDp_i) + lo(D200_i)) + \varepsilon_i, \quad i = 1, \ldots, n \quad (2.7)$$

new values from this model are obtained to calculate the empirical distribution of the estimator T_0 total number of mackerel eggs, T_0 via bootstrap.

This distribution is obtained by adding independent normal variables with zero-mean and an estimated variance to the fitted values from the model. Because of the lack of theoretical results in the context of generalised additive models, this variance is estimated as: $\hat{\sigma}^2 = \sum(y_i - \hat{y}_i)^2/(n - df)$ and $\hat{y}_i$ is the fitted value from the model 2.7, $i = 1, \ldots, n$. The degrees of freedom, df, in the denominator of the estimated value of the variance of errors is equal to the total degrees of freedom for the fitted model, for analogy with the linear models with constant variance.

If the model for the population were the chosen model and the total is estimated from other nested models, empirical distributions for the corresponding versions of T_0 can be calculated via the bootstrap. If one of these models whose

corresponding estimator is, say, T_1, is "close" to the chosen model, from the point of view of its capacity to produce similar estimators for the parameter τ then, the distribution of T_1 is expected to be similar to the distribution of T_0. The bootstrap methodology can be, consequently, used to compare the distributions of the brother estimators of T_0 as coming from different G.A.M models (T_1, T_2, and T_3 for the models in the second, third and forth rows in table 2, respectively, and T_4, for the model of Borchers et al) under the assumption of validity of the selected model. This criterion can be also applied to compare the distributions of T_0 and the raw estimator T. Therefore, this procedure could be used as alternative way to compare the models. The corresponding random variables for the total as estimated from the different models (in the order of the Table 2) are called T_1, T_2 and T_3 respectively.

This bootstrap methodology can be summarised in the following steps:

1. Consider that the model for the population of mackerel eggs is 2.7:

$$y_i = lo(lat_i, long_i) + lo(BDp_i) + lo(D200_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

   where y_i is the observed value for log(density) in the population and $\varepsilon \sim N(0, \sigma^2)$.

2. Fit the model 2.7 to the values y_i. Let these fitted values be y_{0i}, $i = 1, \ldots, n$. The parameter τ_0 can be calculated from this model: $\tau_0 = \sum_{i=1}^{n} \exp(y_{0i})$

3. Generate values from the model

$$y_i^* = y_{0i} + \varepsilon_i, \quad i = 1, \ldots, n$$

where $\varepsilon \sim N(0, \hat{\sigma}^2)$, $\hat{\sigma}^2 = \sum(y_i - \hat{y}_i)^2/(n - df)$ and $\hat{y}_i$ is the fitted value from the model (2.7), $i = 1, \ldots, n$. The degrees of freedom, df, in the denominator of the estimated value of the variance of errors is equal to the total degrees of freedom for the fitted model, by analogy with the linear models with constant variance.

4. Fit the model (2.7) to the values y_i^*. Let these fitted values be $\hat{y}_{0i}$, $i = 1, \ldots, n$ and calculate values $T_0^* = \sum_{i=1}^n \exp \hat{y}_{0i}$ of the estimator T_0 from the chosen model.

5. Fit the model:

$$y_i = lo(lat_i, long_i) + lo(D200_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{2.8}$$

to the values y_i^*.

Let these fitted values be $\hat{y}_{1i}$, $i = 1, \ldots, n$ and calculate an estimated value of the total number of mackerel eggs from this model: $T_1^* = \sum_{i=1}^n \exp \hat{y}_{1i}$

6. Fit the model:

$$y_i = lo(lat_i, long_i) + lo(BDp_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{2.9}$$

to the values y_i^*.

Let these fitted values be $\hat{y}_{2i}$, $i = 1, \ldots, n$ and calculate an estimated value of the total number of mackerel eggs from this model: $T_2^* = \sum_{i=1}^n \exp \hat{y}_{2i}$

7. Fit the model:

$$y_i = lo(lat_i, long_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{2.10}$$

to the values y_i^*. Let these fitted values be $\hat{y}_{3i}$, $i = 1, \ldots, n$ and calculate an estimated value of the total number of mackerel eggs from this model: $T_3^* = \sum_{i=1}^{n} \exp \hat{y}_{3i}$

8. Finally, and in order to compare with the model from Borchers et al., fit the model

$$y_i = \beta_0 + S(lat_i) + S(lon_i) + S(BDp_i) + S(D200_i) + \beta_1(lat_i.lon_i) + \beta_2(lat_i.BDp_i) + \varepsilon_i \tag{2.11}$$

to the values y_i^* and calculate an estimated value of the total number of mackerel eggs from this model: $T_4^* = \sum_{i=1}^{n} \exp \hat{y}_{4i}$

9. Repeat 1) to 8) a reasonable number of times in order to calculate the empirical distribution of the estimators of the total number of eggs.

10. Compare some quantiles of these distributions.

Figure 2.8 shows the box-plots for the empirical bootstraps distributions of T_0^*, T_1^*, T_2^*, T_3^*, and T_4^*.

The corresponding 95% confidence intervals for the parameter τ, as calculated from these empirical distributions, are: [7319.80,10.645.05], [6827,9671.23], [7320,10669.84], [6433, 9556.48], and [7388.24,10719.68].

All the empirical distributions can be considered approximately symmetric. Their means are 8916.35, 8165.15, 8887.57, 7756.52, and 8878.41 and their

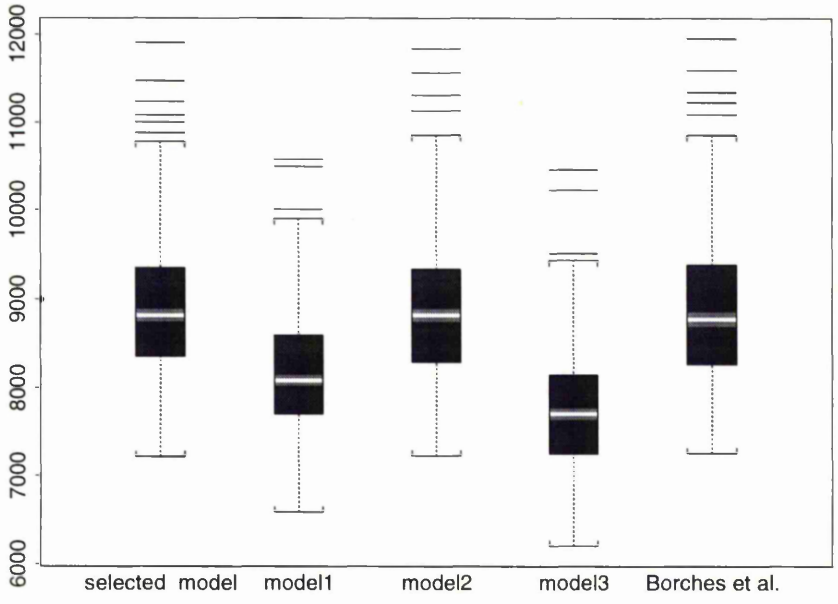

Figure 2.8. The box-plots for the empirical distributions (via bootstrap) of the estimators T_0^* (selected model), T_1^*, (model 1), T_2^* (model2), T_3^* (model 3) and T_4^* (Borchers et al.'s model) for the parameter τ are shown in this figure.

medians are 8900.23, 8138.80, 8864.91, 7748.06, and 8838.41 respectively. Their interquartile ranges are also not significantly different. In fact, they are: 2119.85, 1782.37, 2076.07, 1737.61 and 2002.63 respectively.

From Figure 2.8 it is also clear that the empirical distributions of T_0^* and T_2^* for the total number of mackerel eggs are not substantially different. This result agrees with the stronger functional dependency of the response variable on the covariate bottom depth than the variable distance from 200m contour, even though there was not enough evidence from the study of the deviances to eliminate it.

Otherwise, the corresponding distributions for T_1^* and T_3^* have their quantiles below the corresponding quantiles for T_0^*, T_2^* and T_4^*. Particularly, the model based only on the covariates latitude and longitude produces quantiles much lower than those calculated from the other models.

The distribution of T_4^* corresponding to the model from Borchers et al. (1994) is very similar to that of the selected model here, T_0^*. This characteristic also confirms the fact that the selected model offers a very good option to model the density of mackerel.

Figure 2.9 shows a box-plot of the bootstrapped estimator T_o^* corrected by $exp(0.5\hat{\sigma}^2)$ and the bootstrapped raw estimator T^*. This figure shows that even though the mean of these two distributions are very similar (14607.27 and 14711.8, respectively), the dispersion (standard deviation) of the values when they are fitted to a model (under the assumption of validity of it) decreases considerably (from 1829.98 to 1340.66).

Figure 2.10 is a plot of the frequencies of the bootstrap estimators T_0^* and T^*. The sharper shape of the curve corresponding to the first estimator is also highlighted in this figure. The value of the parameter τ_0 (corrected by $\exp(0.5\sigma^2)$) , under the assumption that this is the model of the population, is marked on the horizontal axis and a dotted vertical line through this point is drawn to show the symmetry of these curves respect to this line.

In figure 2.9 a comparison of T_0^* and T^* is performed via a scatter plot. This graph shows the larger dispersion of the values of T^* than those of T_0^*. This characteristic is highlighted in Figure 2.12 where the distribution of $T^* - T_0$ is represented. The mean of this distribution is 104.5356 and the standard

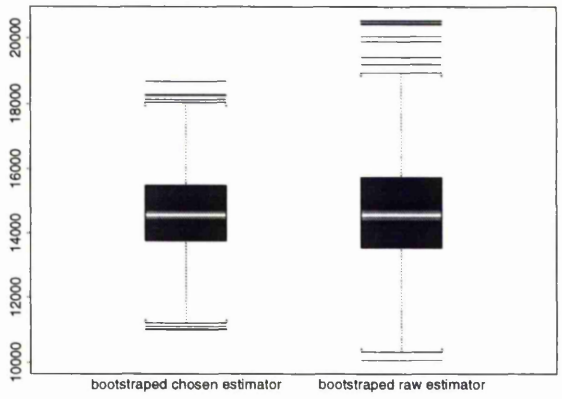

Figure 2.9. The box-plots for the empirical distributions (via bootstrap) of the total biomass of mackerel T_0^* (selected model) and the same variable when no fitted model is considered are shown in this figure.

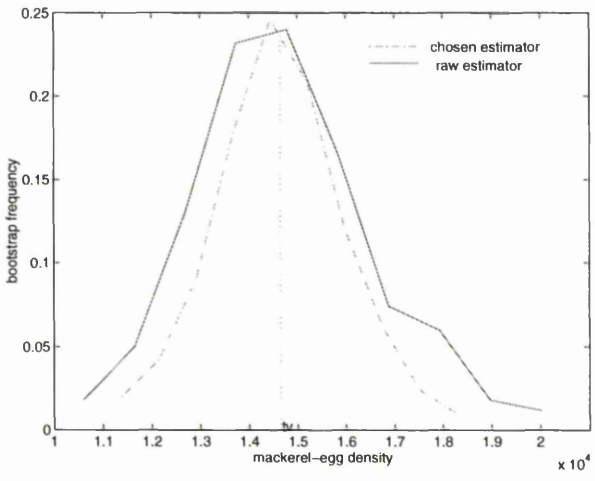

Figure 2.10. The bootstrap frequencies of chosen estimator T_0 and the raw estimator T are shown in this plot. The point indicated as "tv" is the value of the parameter τ_0 if the model for the population is the chosen model. The vertical line through tv highlights the symmetry of both curves respect to this line.

deviation 1180.017. The number of outliers visible in this plot is not surprising because of the more variable behaviour of the raw estimator T.

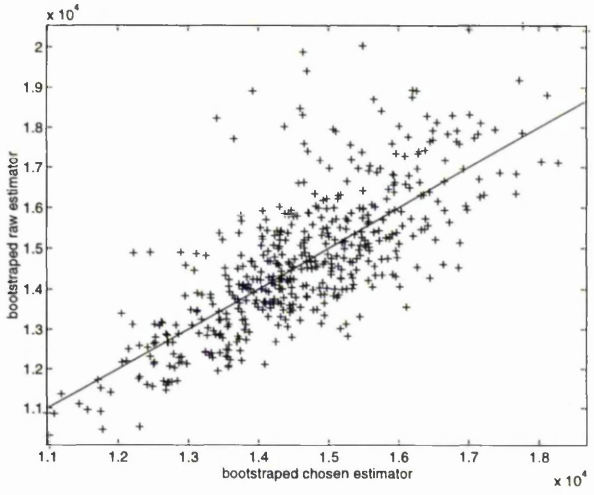

Figure 2.11. A plot of the bootstrapped chosen estimator,T_o^*, against the bootstrapped raw estimator, T, is shown in this figure. The line "y=x" is also displayed.

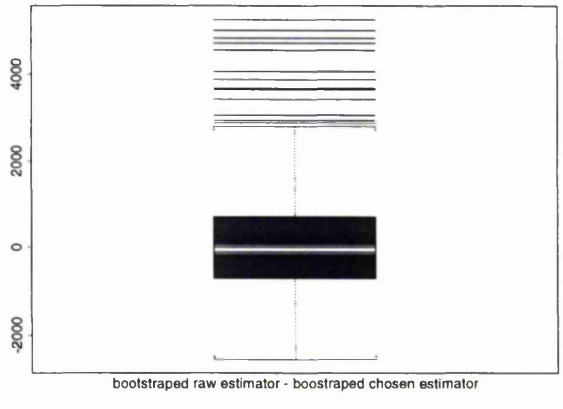

Figure 2.12. A box-plot of the bootstrap distribution of the difference $T_o^* - T$ is shown in this figure.

2.7 Conclusions

The tools provided by generalised additive models offer an attractive way to model density of mackerel eggs as a function of covariates such as latitude, longitude, bottom depth (BDp) and distance from the two hundred meters contour (D200). Even though formal inference tools have not been developed yet, an analysis analogous to that constructed for generalised linear models can be used as a first step to select an appropriate model. Probabilistic investigations such as distributions for the estimates of total through different models can be carried out with the bootstrap.

A first non-statistical selection of adequate covariates should take into account the nature and meaning of these variables in relation with the phenomenon they try to explain. This chapter has attempted to extend the model of Borchers, Buckland and Ahmadi (1993) by considering the variables latitude and longitude in a single joint term of the model. This was done in an attempt to construct natural combinations of covariates. Both models, the model from Borchers et al. (1993) and the model selected in this work, have produced similar results from the point of view of the aim for which these models were built,i.e.,an estimation of the density of egg mackerel. These results would justify the use of the model proposed here given its simplicity and more straightforward interpretation.

Modelling the density of egg mackerel as it was done here can help to detect which covariates have more influence and which can be discarded in the case where there are restrictions on resources or time. The variable "distance from two hundred meters contour" is an example in this case. In fact, even when

there is not enough evidence to withdraw this variable from the model from the point of view of the analysis of deviances, the bootstrap study shows little difference in the estimated total number of mackerel eggs from models with and without this variable. On the other hand, the variable bottom depth is shown to be highly relevant to this estimate.

Chapter 3

Testing for Constant Variance in a Linear Model

3.1 Introduction

It is a very common assumption in linear regression models that the variance of the error term is constant. It is also very common for this assumption to be checked informally by using an appropiate graphical method, such as a residual plot. However, plots of this kind do not always allow clear conclusions to be reached. It is the objective of this chapter to explore more formal tests of the assumption of constant variance.

The research literature contains a variety of work on heteroscedasticity taking different approachs and using different tools. This work can be classified into papers where the main aim is to estimate the linear parameters under

an assumption of heteroscedasticity, and those whose interest is focussed on checking this assumption.

Several works dealing with the estimation of the variance function, and linear regression parameters under the assumption of heteroscedasticity appeared during the last two decades. Fuller and Rao (1977) proposed an estimator for a model in which the variances of errors are assumed to have different constant values, i.e. $var(\varepsilon_i) = \sigma_i^2$ where σ_i^2 is a positive real number for $i = 1, \ldots, n$. The estimators are calculated in two steps in an iterative procedure.

Carroll (1982) considered a model in which the variances of errors are smooth functions of the design points. For these models, he proved also that these estimators are equivalent (asymptotically) to the weighted least squares estimators with known variances. Also, in some of these papers, the problem of finding "good" estimators has been tackled by using nonparametric smoothing techniques and specifically those concerned with kernel smoothers. A work of this type is the paper by Müller and Stadtmüller (1987) who have used kernel smoothers to obtain estimators of the variance function in the general regression model.

Work concerned with checking homoscedasticity using both parametric and nonparametric tools can be found in the recent statistical literature. Since this problem is the main focus of this chapter detailed references are considered here.

In the context of nonparametric regression models, Müller and Zhao (1995)

proposed a methodology to estimate parametric and nonparametric components of the model. They considered the model:

$$y_i = g(t_i) + \delta_i, \quad for \quad 0 \leq t_i \leq 1 \quad and \quad i = 1, ..., n$$

where δ_i are independent errors with $E(\delta_i) = 0$, $E(\delta_i^2) = var(y_i) = \sigma^2(t_i)$, for $1 \leq i \leq n$. No other distributional assumptions are made; g is assumed to be smooth, while g and σ^2 follow the generalized linear model:

$$G(\sigma^2(t)) = \theta_0 + \sum_{j=1}^{p-1} \theta_j H_j(g(t))$$

where G and H_i, for $1 \leq i \leq (p-1)$ are known link functions and θ_i, $0 \leq i \leq (p-1)$, are unknown parameters.

In this context, the authors have considered estimators for the nonparametric parts of the model, g and σ^2, and the parametric ones, namely the vector $\beta = (\theta_0, \theta_1, ..., \theta_{p-1})^T$. The estimators proposed for g and σ^2 are:

$$\hat{g}(t) = \sum_{i=1}^{n} w_i(t) y_i$$

$$\hat{\sigma^2}(t) = \sum_{i=1}^{n} w_i(t) y_i^2 - (\sum_{i=1}^{n} w_i(t) y_i)^2 = \sum_{i=1}^{n} w_i(t) y_i^2 - \hat{g}^2(t)$$

where the functions w_i are defined by using appropriate kernel functions. An estimator for the parameter vector $\beta = (\theta_0, \theta_1, ..., \theta_{p-1})^T$ is obtained through weighted least-squares as

$$\hat{\beta} = (\hat{\theta}_0, \hat{\theta}_1, ..., \hat{\theta}_{p-1}) = arg\min_{(\theta_0, \theta_1, ..., \theta_{p-1})} \sum_{i=1}^{n} q(t_i) \left[G(\hat{\sigma}^2(t_i)) - \sum_{l=0}^{p-1} \theta_l H_l(\hat{g}(t_i)) \right]^2$$

where q is a Lipchitz-continuous weight function. The parameter estimate can be expressed as:

$$\hat{\beta} = (\hat{X}^T Q^{-1} \hat{X})^{-1} \hat{X}^T Q^{-1} \hat{Z}$$

where $X = (x_{ij}),\ 1 \leq i \leq n,\ 1 \leq j \leq p,\quad \hat{X} = (\hat{x}_{ij}),\ 1 \leq i \leq n,\ 1 \leq j \leq p;$

$$x_{ij} = H_{j-1}(g(t_i)), \quad \hat{x}_{ij} = H_{j-1}(\hat{g}(t_i))$$

$$Z = (G(\sigma^2(t_1), G(\sigma^2(t_2), ..., G(\sigma^2(t_n))^T,\ \hat{Z} = (G(\hat{\sigma}^2(t_1), G(\hat{\sigma}^2(t_2), ..., G(\hat{\sigma}^2(t_n))^T$$

$$Q^{-1} = diag(q(t_1), q(t_2), ..., q(t_n))$$

The authors also suggested an iterative method of simultaneous estimation of parametric and nonparametric components under some assumptions on the representation of G, g, and β. In this scenario, the estimators are obtained as the convergence of the corresponding sequences. The asymptotic distribution of $\hat{\beta}$ is particularly interesting from the point of view of testing homoscedasticity. The null hypothesis corresponding to homoscedasticity is:

$$H_0 : \theta_1 = \theta_2 = ... = \theta_{p-1} = 0$$

or, equivalently, $H_0 : \Lambda\beta = O$, where $\Lambda = (\lambda_{ij}),\ \ 0 \leq i \leq (p-2); 0 \leq j \leq (p-1)$ is defined as:

$$\lambda_{ij} = \begin{cases} 1 & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Under H_0, and under some assumptions of regularity for the functions G and H_i, i=1,..,p-1, and for the kernel K and sequence of bandwiths b_n, where b_n is the bandwith corresponding to a sample size n used to estimate g and σ^2, it has been proved that: $\sqrt{n}(\Lambda\hat{\beta}) \xrightarrow{\ell} N(O, \Sigma)$ where $\Sigma = \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$,

$\Sigma_0 = [\rho_{kl}]$ $\Sigma_1 = [\tau_{kl}]$, $0 \leq k,\ \ l \leq p-1$, and

$$\rho_{kl} = \int_0^1 f(t) H_k(g(t)) H_l(g(t)) q(t) dt$$

$$\begin{aligned} \tau_{kl} \ = \ & \int_0^1 f(t) H_k(g(t)) H_l(g(t)) q^2(t) \{G'(\sigma^2(t))\}^2 \{\mu_4(t) - \sigma^4(t)\} \\ & + 2G'(\sigma^2(t)) \left\{ \sum_{j=0}^{p-1} \theta_j H_j'(g(t)) \right\} \mu_3(t) + \left\{ \sum_{j=0}^{p-1} \theta_j H_j'(g(t) \right\}^2 \sigma^2(t) dt \end{aligned}$$

$$\text{where} \ \ \mu_3(t) = \sum_{i=1}^{n} w_i(t) y_i^3 - \{\hat{g}(t)\}^3 - 3\hat{g}(t)\hat{\sigma}^2(t)$$

,

$$\mu_4(t) = \sum_{i=1}^{n} w_i(t) y_i^4 - \{\hat{g}(t)\}^4 - 4\hat{g}(t)\hat{\mu}_3(t) - 6\hat{g}(t)\hat{\sigma}^2(t)$$

This hypothesis is a particular case of a general one which can be expressed as: $\Xi\beta = \xi_0$, where Ξ is a (mxp)-matrix of rank m ($m \leq p$) and ξ_0 is a m-vector. The test statistic proposed is:

$$T_n = n\left[\Xi\hat{\beta} - \xi_0\right]^T \left[\Xi\hat{\Sigma}\Xi\right] \left[\Xi\hat{\beta} - \xi_0\right]$$

where $\hat{\Sigma} = \hat{\Sigma_0}^{-1}\hat{\Sigma_1}\hat{\Sigma_0}^{-1}$, $\hat{\Sigma_0} = [\hat{\rho}_{kl}]$ $\hat{\Sigma_1} = [\hat{\tau}_{kl}]$, $0 \leq k,\ \ l \leq p-1$,

$$\begin{aligned} \hat{\tau}_{kl} \ = \ & (1/n) \sum_{i=1}^{n} f(t_i) H_k(\hat{g}(t_i)) H_l(\hat{g}(t_i)) q^2(t_i) \{G'(\sigma^2(t_i))\}^2 \{\mu_4(t_i) - \sigma^4(t_i)\} \\ & + 2G'(\sigma^2(t_i)) \left\{ \sum_{j=0}^{p-1} \theta_j H_j'(g(t_i)) \right\} \mu_3(t_i) + \left\{ \sum_{j=0}^{p-1} \theta_j H_j'(g(t_i)) \right\}^2 \sigma^2(t_i) \end{aligned}$$

Under H_0 and some assumptions for the functions G and H_i i=1,..,p-1; the

kernel K and sequence of bandwiths (b_n) it has been also showed that:

$$T_n \xrightarrow{\ell} \chi^2(m)$$

For the null hypothesis of homoscedasticity, $\Xi = \Lambda$, the test statistic becomes $T_n = n\left[\Lambda\hat{\beta}\right]^T \left[\Lambda\hat{\Sigma}\Lambda^T\right] \left[\Lambda\hat{\beta}\right]$, and the test for homoscedasticity can be defined as: reject H_0 if $T_n > \chi^2_{m;\alpha}$ where $\chi^2_{m;\alpha}$ is the $100(1-\alpha)\%$ quantile of the χ^2-distribution with m degrees of freedom.

The idea of using nonparametric smoothing techniques in order to check assumptions about the form of a parametric model has produced various works in the statistical literature. One of these works is the paper by Azzalini and Bowman (1993) who tackled the problem of checking linear trend in a regression model. In this work, they started from the informal checking of linearity through the analysis of residual plots and they built a formal test statistic based on nonparametric kernel estimation to identify patterns in the residuals. They considered the null hypothesis of linear trend against the alternative one of a smooth non-linear trend given by a smooth function of the independent variable. The test statistic proposed for this hypothesis is to comapre the residual sums of squares of the raw residuals and a smoothed version. This is defined initially as $F = (y'M_0y - y'M_1y)/y'M_1y$, where y is the response variable, $M_0 = I - X(X'X)^{-1}X'$ with X the design matrix, and $M_1 = (I-W)'(I-W)$ with W the (nxn)- matrix of the weights w_{ij} in the estimation of the regression function g as $\hat{g}(x_i) = \sum_{j=1}^{n} w_{ij}y_j$. The form of F follows the methodology of a pseudo-likelihood ratio for the considered hypotheses of interest. However, the distribution of this test statistic depends on the unknown linear regression parameters under the null hypothesis. This

problem was sorted out by reformulating the hypothesis in terms of residuals. Hence, the least squares residuals are compared with a smooth version of them through the test statistic $F = (e'e - e'M_1e)/e'M_1e$. The idea of analyzing the residuals for checking these hypotheses led to a test statistic whose distribution can be approximated straightfowardly.

Plots of residuals are one of the most traditional informal methods for checking homoscedasticity in the context of linear regression models. This approach is also the base of the test of Cook and Weisgberg (1983). They have considered a linear model such as $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where ϵ_i, i=1,.,n, are independent and normally distributed with variance $var(\epsilon_i) = \sigma^2(exp(\lambda^T \boldsymbol{z}_i))$, where λ is an unknown vector of parameters and z_i is a known vector that may be different for each i. For example, $\boldsymbol{z}_i$ may be the response variable y_i (or a function of it), the vector of predictors $\boldsymbol{x}_i$, etc. The variance of errors is, therefore, assumed to be a monotonic function in each component of z_i. The hypotheses in this case are:

$$H_0 : \lambda = 0 \quad \text{against} \quad H_1 : \lambda \neq 0$$

This test statistic is also based on the idea of analyzing the behaviour of the residuals $\hat{\varepsilon}_i$, $i = 1, \ldots, n$ in the regression of y on X. In this approach the squares of the least squares residuals (divided by $\hat{\sigma}^2 = (\sum_i \hat{\varepsilon}_i^2)/n$) are regressed on $\boldsymbol{z}_i$, $i = 1, \ldots, n$. The sum of squares due to the latter regression SSreg (divided by two) is the statistic for this test, i.e., S=SSreg/2 where

$$SSreg = (U - \bar{U})^T(U - \bar{U}) - (U - Z\hat{\gamma})^T(U - Z\hat{\gamma})$$

and $(U - \bar{U})^T = (u_1, u_2, .., u_n) - ((1/n)\sum_{i=1}^n u_i)(1, 1, .., 1)$, $u_i = \hat{\varepsilon}_i^2/\hat{\sigma}^2$, where

$\hat{\varepsilon}_i$ is the $i^{\underline{th}}$ residual in the regression of y on X, i=1,...,n; Z is a nx(q+1)-matrix whose $i^{\underline{th}}$ row is $(1, z_i)$, and $\hat{\gamma}$ is the estimated vector of the parameter vector $\gamma^T = (\gamma_0, \gamma_1, ..., \gamma_q)$ in the regression of U on $z_1, .., z_n$. The test statistic S has an asymptotic χ^2 distribution under the null hypothesis and the test is: Reject H_0 if the corresponding observed value of S is greater than the $(1-\alpha)$-quantile of the χ^2 -distribution with q degree of freedom.

The test statistic of Cook and Weisberg (1983) for the hypothesis of homoscedasticity in a linear model belongs to the class of statistics which have been called "score statistics" (see for example Simonoff and Tsai (1994)). The common assumption for this kind of test is that the errors are normal and independent with variance-covariance matrix $\sigma^2 W$ where W has diagonal entries $w_{ii} = w(\boldsymbol{z}_i, \boldsymbol{\delta}), \quad i = 1, \ldots, n$, $\boldsymbol{z}_i'$ is the ith row of the nxq matrix $\boldsymbol{Z}$ of variance predicting variables and $\boldsymbol{\delta}$ is a qx1 vector of unknown parameters. The score statistic was proposed originally by Rao (1947). It has the form $S = \boldsymbol{V}_0 I_0 \boldsymbol{V}_0$, where $\boldsymbol{V}_0 = \partial/\partial\theta$ is the first-derivative (score) vector, $\boldsymbol{\theta}$ is the vector of parameters and $I_0 = E(-\partial^2 l/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')$, where l is the likelihood function is the expected information matrix, both evaluated at the null hypothesis. Also, S is a first-order approximation to the likelihood ratio statistic. This statistic was not only proposed by Rao (1947) and Cook & Weisberg but also by Godfrey (1978) and Breusch & Pagan (1979) apparently independently.

An effort to improve the robustness of the score statistics when the distribution of errors is not normal was done by Koenker (1981). He proposed to "studentize" the statistic S when the variance is hypothsesized to be a function of the fitted values. The resultant statistic is $S^* = 2\hat{\sigma}^4 S/\hat{\phi}$ where $\hat{\phi} = \sum_i (\hat{e}_i{}^2 - \hat{\sigma}^4)^2/n$. Unlike S, S^* is asymptotically χ^2_q for a large class of error distributions.

Other versions of the score statistics S and S^* are their derivations from the modified profile likelihood ratio statistic as suggested by Simonoff and Tsai (1994). They are, respectively,

$$S_m = S + \sum_{a=1}^{q} \left(\sum_{i=1}^{n} h_{ii} t_{ia} \right) \tau_a$$

and

$$S_m^* = S^* + \sum_{a=1}^{q} \left(\sum_{i=1}^{n} h_{ii} t_{ia} \right) \tau_a$$

where h_{ii} is the ith diagonal element of the matrix $H = X(X'X)^{-1}X'$,

$$t_{ia} = \frac{\partial w(\boldsymbol{z}_i, \boldsymbol{\delta})}{\partial \delta_a} - \sum_j \frac{\partial w(\boldsymbol{z}_j, \boldsymbol{\delta})}{n \partial \delta_a} / n$$

evaluated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$, $\tau_a \quad a = 1, \ldots, q$ are the components of the vector $\boldsymbol{\tau} = (\overline{D}'\overline{D})^{-1}\overline{D}'\boldsymbol{u}$ where $\overline{D} = (I - \mathbf{1}\mathbf{1}'/n)D$ and D is the nxq matrix with entries $d_{ij} = \partial w(\boldsymbol{z}_i, \boldsymbol{\lambda})/\partial \lambda_j$ and $\boldsymbol{u}$ is the vector with components $u_i = \hat{e}_i/\sigma^2, \quad i = 1, \ldots, n$.

Another approach for test statistics for homoscedasticity under the assumption of the same structure for the variance-covariance matrix of errors is the likelihood ratio statistic. Rutemiller and Bowers (1968) seem to be the first people associated with the derivation of an expression for this statistic under the asssumption of a specific form for the function w, $w(\boldsymbol{z}, \boldsymbol{\lambda}) = \boldsymbol{z}'\boldsymbol{\lambda}$. However this expression may be negative and therefore inappropriate for a variance. This idea led Harvey (1976) to consider a positive expression for the variance, as assumed by Cook & Weisberg (1983), and to derive a corresponding likelihood ratio statistic for the test under this assumption. The corresponding

test statistic is: $L = nlog(\hat{\sigma}^2/\hat{\sigma}_\lambda^2) - \sum_i z_i'\hat{\boldsymbol{\lambda}}$ where $\sigma^2 = \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}/n$ and $\hat{\boldsymbol{e}}$ is the vector of least square residuals. Simonoff and Tsai (1994) derived a modification of this statisitic by using the profile log-likelihood statistic which is:

$$L_m = \frac{n-p-2}{n}L + log\left(\frac{det(X'X)}{det(\hat{X}_m'\hat{X}_m)}\right)$$

where $\hat{X}_m = \hat{G}^{-1/2}X$ and $\hat{G}$ is the diagonal matrix with ith entry

$$g_{ii} = w(\boldsymbol{z}_i, \hat{\boldsymbol{\lambda}}) / \left(\prod_j w(\boldsymbol{z}_i, \hat{\boldsymbol{\lambda}})\right)^{1/n}.$$

Various of the test statistics for heteroscedasticity in a linear model under the assumption of independent errors with variance-covariance matrix $\sigma^2 W$ where W has diagonal entries $w_{ii} = w(\boldsymbol{z}_i, \boldsymbol{\delta}), \quad i = 1, \ldots, n$, were compared by Simonoff and Tsai (1994) through Monte Carlo simulations. Even though these test statistics deal with a specific form for the variance-covariance matrix of errors, this form is quite general and matrices of this type occur, for example, in the area of time series analysis with its applications in bussiness and economics.

In the remainder of this chapter, a nonparametric approach, based on nonparametric smoothing techniques from the area of kernel estimation, is used to build a formal test of constant variance. As shown in the simulations studies, one of the most important advantages of this approach is its capacity to detect heteroscedasticity under very general assumptions for the shape of the variance functions. An approximation to the distribution of the test statistic is created by matching the moments of a quadratic form to those of a shifted χ^2 distribution. A simulation study of the power for the test is carried out. A bootstrap study of the empirical distribution of the test statistic is also

performed. A graphical follow-up to the global test is also proposed. This is refered to as "reference band". Several examples are used as illustrations.

3.2 A description of the test

3.2.1 The test statistic

As a first approach we consider the simple linear model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{i=1,...,n} \tag{3.1}$$

where ε_i has a normal distribution with mean 0 and variance σ_i^2 for i=1,..,n, and $cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. The ideas to be developed can be easily extended to the general linear model. The hypothesis to be tested can be written as:

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 \\ H_1 : \sigma^2 &= \sigma_i^2 = \textit{smooth} \text{ function of } x_i, \quad \text{i=1,...,n} \end{aligned}$$

Using least squares, the fitted regression model $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ produces the residuals $r_i = y_i - \hat{y}_i$. Under the assumption of Normally distributed errors, these residuals are also Normally distributed with mean zero. It would be natural to examine the behaviour of the variables $|r_i|$ or r_i^2 in order to check scale changes in the errors. However, these variables have skewed distributions and nonparametric smoothing techniques are more stable when the underlying distributions are approximately Normal. Cleveland (1993) observes

that the transformation $t_i = |r_i|^{1/2}$ induces approximately Normality. This transformation has the same basic shape as the function $\Phi^{-1}F$ which provides an exact means of creating a Normal random variable, with distribution function Φ, from a Chi-squared random variable, with distribution function F. Since the r_i all have slightly different variances, even when the variance of ε_i is constant, it is more appropriate to deal with the adjusted variables $s_i = |r_i|^{1/2} - E_0(|r_i|^{1/2})$, where the subscript 0 denotes that the calculation is carried out under the null hypothesis. The expectation $E_0(|r_i|^{1/2})$ can be calculated easily, and an explicit expression is given in Section 2.2 below.

Under H_0 the values of s_i will lie close to their average $\bar{s}$, whereas under H_1 local variations in the scatter about the average are to be expected. If it is reasonable to assume that these local variations change in a smooth manner, then it becomes natural to employ nonparametric smoothing to identify the trends in a more powerful way, without making any assumptions about the shape of these trends. The kernel method of nonparametric regression provides a simple means of doing this. Wand & Jones (1995) give an introduction to this technique. In its simplest form, a smooth curve is defined across the design space as

$$\tilde{s}(x) = \sum_{j=1}^{n} w_j(x) s_j \tag{3.2}$$

where the weights $w_j(x)$ are defined by the kernel function and sum to 1 to provide a weighted average. In this paper a Normal kernel function is used, giving weights

$$w_j(x) = \frac{exp(-(\frac{x-x_j}{h})^2)}{\sum_{k=1}^{n} exp(-(\frac{x-x_k}{h})^2)}$$

for $j = 1, \ldots, n$ and h the bandwidth. If the values $\tilde{s}(x_i)$ of the smooth curve at each design point x_i are denoted simply by $\tilde{s}_i$ then a suitable test statistic

to assess the degree of scale variation in the data is

$$T = \frac{\sum_{i=1}^{n}(s_i - \bar{s})^2 - \sum_{i=1}^{n}(s_i - \tilde{s}_i)^2}{\sum_{i=1}^{n}(s_i - \tilde{s}_i)^2} \tag{3.3}$$

Under the null hypothesis, there will be little difference in the size of the terms $\sum_{i=1}^{n}(s_i - \bar{s})^2$ and $\sum_{i=1}^{n}(s_i - \tilde{s}_i)^2$. Under the alternative, the first of these terms should become systematically larger than the second and so the test statistic will tend towards large positive values. A formal test will take the form

$$\text{Reject } H_0 \quad \text{iff} \quad T > t_0$$

where t_0 is determined by the size of the test:

$$\alpha = P(reject\ \ H_0|\ H_0\ \ is\ \ true) = P(T > t_0|\sigma^2 = \sigma_0^2)$$

3.2.2 The distribution of T

The smoothing parameter h controls the degree of smoothing which is applied to the data. For the approximation of the T-statistic distribution, it is important to realize that expression 3.2 is linear in the variables $s_1, s_2, \ldots, s_n$. Then, if $s = (s_1, s_2, \ldots, s_n)^T$, $\tilde{s} = (\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n)^T$, and W is the nxn matrix with entries w_{ij}, we can write: $\tilde{s} = Ws$ and the quadratic form $\sum_{i=1}^{n}(s_i - \tilde{s}_i)^2$ can be expressed as $\sum_{i=1}^{n}(s_i - \tilde{s}_i)^2 = (s - Ws)^T(s - Ws) = s^T Bs$, where B is the matrix $(I - W)^T(I - W)$, with the nxn identity matrix I. The proposed statistic 3.3 can therefore be written as:

$$T = \frac{s^T As - s^T Bs}{s^T Bs} \tag{3.4}$$

where A is the matrix $I - \frac{L}{n}$ and L is the nxn matrix with all of its entries equal to one. This shows that T is the ratio of two quadratic forms, e.g., $T = \frac{s^T Cs}{s^T Bs}$, with C=A-B.

The distribution of the ratio of two quadratic forms in normal variables has been widely discussed in the literature (see for example Mathai and Provost, 1992). In this case, the random variables in the quadratic forms are approximately normal and so, following the ideas of Azzalini and Bowman (1993), the distribution of T can be accurately approximated. In fact, if t_1 is an observed value of T, the corresponding p-value for the test can be written as:

$$p = P(\ T > t_1 \mid H_0 \text{ true}) = P(\ \frac{s^T Cs}{s^T Bs} > t_1 \mid \sigma^2 = {\sigma_0}^2) \tag{3.5}$$

$$= P(s^T(C - t_1 B)s > 0 \mid \sigma^2 = {\sigma_0}^2) \tag{3.6}$$

Hence, in order to calculate the p-value, we can calculate the distribution of $Q_{t_1}(s) = s^T(C - t_1 B)s$, which is again a quadratic form in the approximately normal variables s_i, i=1,...,n.

The quadratic form $Q_{t_1}(s)$ can be expressed as a quadratic form in independent approximately normal variables by:

$$Q_{t_1}(s) = s^T(C - t_1 B)s = z^T V^T(C - t_1 B)Vz = z^T \Lambda z = Q(z)$$

where V is the matrix of the eigenvectors of $C - t_1 B$, Λ the diagonal matrix of its eigenvalues, $\lambda_1, \ldots, \lambda_n$, and $z = V^T s$. Then, the quadratic form $Q_{t_1}(s)$

can be expressed as $Q(z) = \sum_{j=1}^{n} \lambda_j z_j^2$ where the variables $z_i, i = 1, .., n$ are approximately normal and independent. The distributions of these quadratic forms can be approximated by the distribution of a linear combination of χ^2 random variables, by matching the first cumulant of both random variables. The cumulant generating functions, $K_{Q(z)}$ and K_U of Q(z) and $U = a + bU_1(c)$ where a,b, and c are constants and $U_1(c)$ is a χ^2 random variable with c degree of freedom, respectively, are:

$$K_{Q(z)}(\theta) = (-1/2) \sum_{j=1}^{n} log(1 - 2\theta\lambda_j) = \sum_{j=1}^{\infty} a_j \frac{\theta^j}{j!}$$

where $a_j = 2^{j-1}(j-1)! \sum_{i=1}^{n} \lambda_i^j = 2^{j-1}(j-1)! tr(\Sigma(C - t_1 B))^j$

$$K_U(\theta) = \theta a - (c/2) log(1 - 2\theta b) = \sum_{j=1}^{\infty} b_j \frac{\theta^j}{j!}$$

where $b_1 = a + cb, \quad and \quad b_j = 2cb^j(j-1)!, \quad for \quad j > 1$, see, for instance, Jonston & Kotz (vol.2), 1975.

Then, from the equation system $a_j = b_j, \quad j = 1, 2, 3$, and, calling $d_j = tr(\Sigma(C - t_1 B))^j, \quad j = 1, 2, 3$, the p-value for an observed value t_1 of the test statistic T can be calculated as:

$$\begin{aligned} p &= P(s_t(C - t_1 B)s > 0 \mid \sigma^2 = {\sigma_0}^2) \\ &\approx P(U > 0 \mid \sigma^2 = {\sigma_0}^2) \\ &= P(U_1(c) > |a/b| \mid \sigma^2 = {\sigma_0}^2) \\ &= P((U_1(c) > x_1 \mid \sigma^2 = {\sigma_0}^2) \end{aligned}$$

where U_1 is a χ^2 random variable with $c = d_2^3/(4d_3^2)$ and $x_1 = (2d_1 d_2 d_3 - d_2^2)/(4d_3^2)$

The constants c and x_1 depend on the variance-covariance matrix of s. The entries of this matrix can be calculated from the exact distribution of each variable $t_i = \sqrt{|r_i|}$, $i = 1, \ldots, n$ and the joint-distribution of each pair of residuals (r_i, r_j), $i = 1, \ldots, n$, $j = 1, \ldots, n$, and $i \neq j$.

The explicit expression for the density function of $t_i = \sqrt{|r_i|}$ is:

$$f_{t_i}(x) = \frac{4x}{\sqrt{2\pi\sigma_{r_i}^2}} \exp(-\frac{x^4}{2\sigma_{r_i}^2}) \; I_{[0,\infty)}(x)$$

where $\sigma_{r_i}^2$ is the variance of the residual r_i for i=1,..,n.
The expected values and variances are:

$$E(t_i) = \frac{\sqrt{\sqrt{2}\sigma_{r_i}}}{\sqrt{\pi}} \; \Gamma(3/4) \quad var(t_i) = \frac{\sqrt{2\sigma_{r_i}^2}}{\pi} \; (\sqrt{\pi} - \Gamma^2(3/4))$$

Then, the entries σ_{ij} of Σ are: $\sigma_{ii} = var(s_i) = var(t_i)$, for i=1,...,n,, and $\sigma_{ij} = cov(s_i, s_j) = E(t_i t_j) - \frac{\sqrt{2\sigma_{r_i}\sigma_{r_j}}}{\pi}\Gamma^2(3/4)$ $i \neq j$, i=1,..,n; j=1,...,n where $E(t_i t_j)$ can be calculated by numerical integration by using the joint distribution of (r_i, r_j). In fact, under the null hypothesis, the vector **r** of residuals has a normal distribution with null mean and variance-covariance matrix $\sigma_0^2 H$, where $H = I - X(X^T X)^{-1} X^T$. Hence, a pair (r_i, r_j) has a bivariate normal distribution with null mean and variance-covariance matrix $\sigma_0^2 H_{ij}$ where H_{ij} is the corresponding 2x2 submatrix of H. Then,

$$\begin{aligned} E(t_i t_j) &= \int_{-\infty}^{\infty}\int_{0}^{\infty} \sqrt{|xy|} f_{r_i,r_j}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{|xy|}(1/(2\pi\sigma_0^2\sqrt{|H_{ij}|})) \exp\left(-(1/(2\sigma_0^2))\right)(h_{ii}^{(-1)}x^2 + 2h_{ij}^{(-1)}xy + h_{jj}^{(-1)}y^2))dxdy \\ &= \sqrt{2}\sigma_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{|xy|}(1/(\pi\sqrt{|H_{ij}|})) \exp\left(-(h_{ii}^{(-1)}x^2 + 2h_{ij}^{(-1)}xy + h_{jj}^{(-1)}y^2)dxdy\right. \end{aligned}$$

$$= \sqrt{2}\sigma_0 d_{ij}$$

where $h_{kl}^{(-1)}$,k=i,j;l=i,j are entries of the inverse matrix of H_{ij} and d_{ij} is the value of the last integral. Certainly, the constant σ_0^2 is unknown but it can be estimated by $\hat{\sigma_0^2} = (\sum_{i=1}^{n} r_i^2)/(n-2)$. Although, this estimation is not necessary because $cov(s_i, s_j) = \sqrt{2}\sigma_0(d_{ij} - \frac{\sqrt{h_{ii}h_{jj}}}{\pi}\Gamma^2(3/4)), \quad i,j = 1,\ldots,n; \quad var(s_i) = \sqrt{2}\sigma_0\frac{\sqrt{h_{ii}}}{\pi}(\sqrt{\pi} - \Gamma^2(3/4)), \ i = 1,\ldots,n$, where $h_{ij}, \quad i,j = 1,\ldots,n$ are the entries of the projection matrix H and, consequently, the constant σ_0 is canceled in the calculations of the degree of freedom c and the (1-p) quantile x_1. This integral was calculated numerically by using Nag subroutines.

3.2.3 Three examples

In order to illustrate the practical implementation of the test, three examples are considered. The two first examples have been taken from the literature in this area and the third has been generated with an appropriate dependence for the variances of the errors. In the first, the functional dependence from the design points is quite marked, in the second and third a visual inspection does not identify clearly whether or not the variance is constant.

Example 1: Snow geese

This example is taken from Cook & Jacobsen (1978). It is also described by Weisberg (1987; page 102). In the experiment, the aim is to estimate the number of snow geese in their summer range areas west of Hudson Bay in Canada. An observer estimates the number x_i of geese in a flock spotted

from a small aircraft. Simultanously a photo of the flock is taken and the number y_i of geese is accurately assessed. This procedure was repeated for 48 flocks. The plot of the points and the fitted line is shown in Figure 3.1.

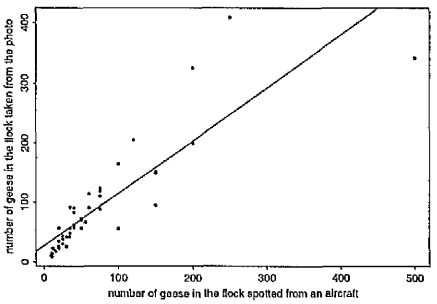

Figure 3.1. Data and fitted line for the snow geese data

A test of non-constant variance for these data was proposed by Cook and Weisberg (1983). A specific parametric alternative, where the variance is assumed to be of exponential form $\sigma^2 exp(\lambda x_i)$ was used. The null hypothesis can then be expressed as $\lambda = 0$. The test statistic has an asymptotic χ^2 distribution. The p-value produced by this test is very small, providing clear evidence that the variance is not constant. In the nonparametric test proposed in this paper, where the alternative hypothesis is expressed simply as a smooth function of x, the observed value of the test statistic is $t_0 = 81.72$ and the value of the approximate χ^2 quantile with 48 degree of freedom was 79.804. Therefore, the p-value is less than 0.00025 and the null hypothesis is convincingly rejected.

As an example where the change in variance is not so extreme, a subset of the data corresponding to cases where $x \leq 100$ was also analysed. These data are

plotted in Figure 3.2 where the change in variance with x is less extreme. The

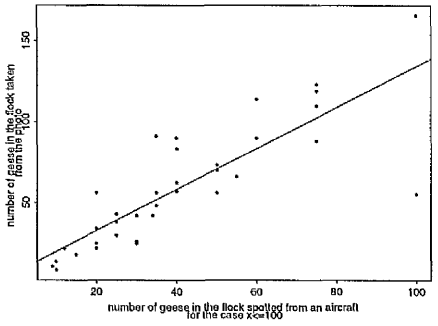

Figure 3.2. Data and fitted line for the subset of the snow geese data

nonparametric test produces a p-value of 0.035 and so the change in variance is still detected.

In this analysis the smoothing parameter was chosen to be one eighth of the range of the x-values. Since this value refers to the standard deviation of the Normal kernel function, each kernel covers approximately half the observations from tail to tail.

Example 2: Body weight and heart weight of cats

Fisher (1947) describes a dataset which includes the heart weights and body weights of a group of cats. Venables & Ripley (1994) also analyse these data. Figure 3.3 shows a plot of these variables for the male cats and the fitted line.

Aitchison (1986) suggested that a log transformation of both variables is appropriate. Such a transformation is often necessary in comparing weights of

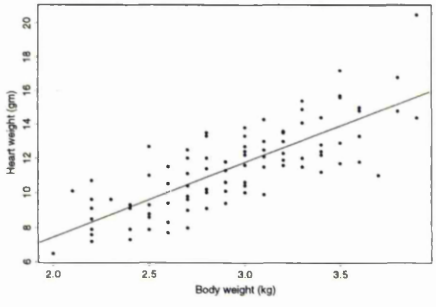

Figure 3.3. Data and fitted line for the male cats data

this kind. The nonparametric test applied to these data, again with a smoothing parameter equal to one eighth of the range of the data, produces a p-value of 0.084. In this case, the mild change in variance for large body weights does not, of itself, provide convincing evidence of changing variance. A decision on whether to adopt log scales for these data will depend on other factors.

Example 3: Simulated linear variance dependence

For the simulated example, the model $y_i = 1 + 2x_i + \varepsilon_i$ was used, where x_i is a design point in [0,1] and ε_i has a normal distribution with mean zero and standard deviation $\sigma(x_i) = (0.5 + x_i)0.25, \quad i = 1, \ldots, 50$. The $x_i, \quad i = 1, \ldots, n$ are equally spaced. The estimated model was $\hat{y} = 0.9888741 + 1.954983x$. A plot of the generated points, and the estimated regression line is shown in Figure 3.4. The values of the square root of the absolute value of the residuals do not offer clear evidence of nonconstant variance for errors. The p-value calculated by using a bandwidth of one fifth was 0.047.

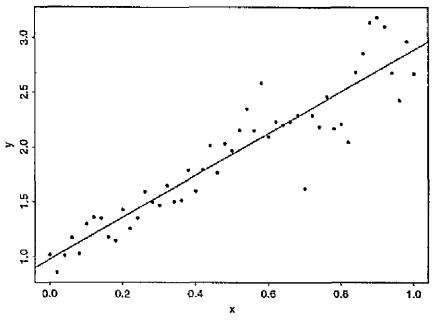

Figure 3.4. Simulated data from the model $y_i = 1 + 2x_i + \varepsilon_i$ was used, where x_i is a design point in [0,1] and ε_i has a normal distribution with mean zero and standard deviation $\sigma(x_i) = (0.5 + x_i)0.25, \quad i = 1, \ldots, 50$.

3.3 Alternative versions

3.3.1 Ignoring the correlation of the residuals

It is widely known that, under the assumptions of the proposed linear model, the variance-covariance matrix of the residuals has a dominant diagonal. This behaviour leads to the possibility of considering this matrix to be approximately diagonal, and hence considering the residuals to be approximately independent. The assumption of approximate independence of residuals implies the independence of the variables $t_i = \sqrt{|r_i|}$, i=1...,n, and so the calculation of their variance-covariance matrix, Σ, does not require numerical approximations. The centred variable $s = (s_1, s_2, \ldots, s_n)^T$, where $s_i = t_i - E(t_i)$, then have a diagonal variance-covariance matrix Σ with the $i^{\underline{th}}$ diagonal entry given by the expression for $\text{var}(t_i)$. Under the null hypothesis these values

are $\sigma_{ii} = \sigma_o\sqrt{h_{ii}}\sqrt{2}(\sqrt{\pi} - \Gamma^2(0.75))/\pi = \sigma_0 k\sqrt{h_{ii}}$. Then, regarding the residuals as independent residuals, and under the null hypothesis, the variance-covariance matrix of $s = (s_1, s_2, \ldots, s_n)^T$ is $\Sigma = \sigma_0 kD$, where k is a constant and D is the diagonal matrix whose diagonal is equal to the square root of the diagonal of the projection matrix H. In this case:

$$d_1 = tr((C - t_1B)\Sigma) = \sigma_0 k tr((C - t_1B)D) = \sigma_0 k d_1^\star$$

$$d_2 = tr((C - t_1B)\Sigma)^2 = \sigma_0^2 k^2 tr((C - t_1B)D)^2 = \sigma_0^2 k^2 d_2^\star$$

$$d_3 = tr((C - t_1B)\Sigma)^3 = \sigma_0^3 k^3 tr((C - t_1B)D)^3 = \sigma_0^3 k^3 d_3^\star$$

where $d_i^\star = tr(AD)^i$, for i=1,2,3. Then,

$$d_1 = \sigma_0 k \sum_{i=1}^{n} a_{ii}\sqrt{h_{ii}}$$

$$d_2 = \sigma_0^2 k^2 \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2\sqrt{h_{ii}}\sqrt{h_{jj}}$$

$$d_3 = \sigma_0^3 k^3 \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ik}a_{kj}\sqrt{h_{jj}}\sqrt{h_{kk}}$$

where $A = (a_{ij}) = I - \mathbf{1}/n - (1 + t_1)(I - W)^T(I - W)$, $H = (h_{ij}) = I - X(X^TX)^{-1}X^T$ and t_1 is the observed value of the statistic T. Therefore, the expressions for the degree of freedom $c = d_2^3/(4d_3^2) = d_2^{\star 3}/(4d_3^{\star 2})$ of the χ^2 distribution used as approximation and the $1 - p$ quantile $x_1 = (2d_1d_2d_3 - d_2^2)/(4d_3^2) = (2d_1^\star d_2^\star d_3^\star - d_2^{\star 2})/(4d_3^{\star 2})$ depend only on the residuals (through t_1), the matrix W of weights and the matrix X (through H) and, consequentely, their calculations are straightforward. Hence, the price payed for the adoption of independence is a further approximation in the calculation of the p-values.

In order to illustrate that the approximation of the variance-covariance matrix of the residuals by a diagonal-variance matrix does not pay too high a cost in the calculations of p-values for the test, a small simulation study was carried out. The results are tabulated later.

3.3.2 A bootstrap approach

Another possible tool, which has been widely used in the statistical literature during the last decade, is the bootstrap. Even though the idea underlying the bootstrap principle is very old and simple, it was Efron (1979) who made the statistical world aware of its promising features. In fact, this tool is based on the following simple idea. If F is the distribution function of a population, a parameter (the word "parameter" is used because of a lack of any other more appropriate) θ is a function of F, say, $\theta = \theta(F)$, an estimator of θ is a function of $\hat{F}$, or, equivalently, a function of χ where $\hat{F}$ is a distribution function calculated from a sample χ drawn from the population, e.g., $\hat{\theta} = \hat{\theta}(F) = \hat{\theta}(\chi)$. If a sample χ^* is drawn randomly from χ with replacement, a new version for θ could be calculated by regarding the new sample, say, $\hat{\theta}^* = \hat{\theta}(\hat{F}^*) = \hat{\theta}(\chi^*)$, where $\hat{F}^*$ is a distribution function calculated from the sample χ^*. The usefulness of the bootstrap is based on the idea that $\hat{\theta}^\star$ is to $\hat{\theta}$ as the latter is to θ

Since the first work by Efron (1979), many efforts have been made to develop the theory dealing with this methodology and its applications. Some of these results can be found in Bickel & Freedman (1981), Freedman (1981), Singh (1981) among others as well as in the books from Efron and Tibshirani (1993),

and Hall (1993). In some cases, the distribution function of the population is completly unknown whereas in other cases it is known up to a (vector of) parameter(s). In the former situation, $\hat{F}$ is the empirical distribution function of the sample χ, and $\hat{F}^*$ is the empirical distribution function of the sample (or *re*sample) χ^*. When the distribution function of the population F is known up to a (vector of) parameter(s) λ, say, $F = F_{(\lambda)}$, then $\hat{F} = F(\hat{\lambda})$, where $\hat{\lambda}$ was obtained from χ, and $\hat{F}^* = F(\hat{\lambda}^*)$, where $\hat{\lambda}^*$ was obtained from χ^*.

In the setting of the problem presented here, the bootstrap methodology is used to calculate the p-values of the test statistic for a simulated linear model. Under the null hypothesis the underlying distribution function is considered known up to a parameter. In fact, with the notation above, F is the cumulative normal distribution function with mean $X\beta$, λ is the common variance of errors, and θ is a value of the test statistic T.

One advantage of this methodology is that the empirical distribution of the test statistic can be generated directly, without the use of moment approximations. The numerically intensive approach of the bootstrap is likely to lead to slower execution of the test. It does however provide a helpful means of checking on the accuracy of the proposed test.

An algorithm to implement the bootstrap in the present setting is as follows:

1. Simulate of a set of observed values $y_1^*, y_2^*, ..., y_n^*$ from the fitted model 3.1.

2. Obtain the corresponding least squares fitted values $\hat{y_1^*}, \hat{y_2^*}, ..., \hat{y_n^*}$ and the corresponding residuals $r_1^*, r_2^*, ..., r_n^*$.

3. Calculate the observed value t^* of the test statistic T.

4. Repeat steps 1 to 3 a large number of times.

5. Calculate the value of p from the empirical distribution of T.

The performance of the bootstrap test is analyzed through simulation and the results are described in Section 3.4 below.

3.4 A power study

A small simulation study was carried out in order to analyze the performance of the test. Different values of the sample size n as well as different functions for the variance of the errors were considered. A design based on equally-spaced values of the explanatory variable in the interval [a,b]=[0,1] was considered. The values for the regression parameters were β_0=1, β_1=2 and the simulated model was $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, i=1,...,n where ϵ_i, i=1,..,n were independent normal with zero mean. In each case 500 samples were generated by using NAG subrutines and the number of times that the observed significance was below 0.05 was counted.

The values of the bandwith to estimate the variance function as a smooth curve were h=0.08(b-a), 0.16(b-a) and 0.32(b-a). The functions considered for the variances were $\sigma_1(x) = 1$, $\sigma_2(x) = (0.25+x)$, $\sigma_3(x) = 0.25+(x-0.5)^2\sigma_0/2$, and $\sigma_4(x) = 0.25exp(xln(5))$. All of these functions have the same minimum and maximum to make comparison easier. The simulation was also performed for different values of sample size.

In table 3.1 and table 3.2, results when an independent structure for the variance-covariance matrix of residuals is adopted, and the test of Cook & Weisberg, are also shown.

		$\sigma_1(x) = 1$				$\sigma_2(x) = 0.25 + x$			
		non-ind. resid.	indep. resid.	boots-trap	C. & W.	non-ind. resid.	indep. resid.	boots-trap	C. & W.
n	h	%	%	%	%	%	%	%	%
30	0.08	5.6	5.8	6.0	3.8	41.0	41.2	47.4	63.2
	0.16	5.2	6.0	5.2		52.8	53.8	59.4	
	0.32	3.6	4.2	5.2		61.4	62.4	70.2	
35	0.08	5.4	5.2	6.8	4.0	48.6	47.6	66.6	78.0
	0.16	5.0	4.4	6.4		62.2	64.2	70.2	
	0.32	3.0	3.2	8.2		72.0	71.6	80.4	
40	0.08	2.8	3.4	8.2	5.2	51.2	54.8	62.6	86.4
	0.16	4.2	4.6	8.2		69.0	72.0	78.0	
	0.32	4.4	5.0	9.8		78.4	79.4	87.0	
45	0.08	6.2	6.2	6.8	5.4	60.0	60.4	72.6	88.0
	0.16	5.8	6.2	6.0		75.8	75.8	83.0	
	0.32	5.6	5.6	6.0		84.0	85.2	90.4	
50	0.08	3.6	3.6	7.0	5.0	71.4	72.8	75.6	92.1
	0.16	4.8	3.6	4.6		85.6	86.0	89.6	
	0.32	4.0	4.6	6.4		89.6	90.0	93.8	
55	0.08	4.6	4.4	6.4	5.4	76.6	76.8	79.4	94.6
	0.16	5.0	4.4	7.4		89.0	88.4	91.4	
	0.32	4.6	4.6	6.8		92.8	92.4	95.6	
60	0.08	5.4	5.4	6.8	4.4	75.8	76.6	83.0	96.8
	0.16	4.6	5.6	8.4		88.0	88.6	92.4	
	0.32	3.6	3.8	8.8		95.2	95.4	98.0	
65	0.08	5.0	5.8	6.8	4.8	84.4	86.0	87.6	98.8
	0.16	5.0	6.0	7.4		92.6	93.6	94.6	
	0.32	5.6	5.2	9.2		95.4	95.8	97.2	
70	0.08	6.4	6.0	7.2	5.8	87.8	88.0	88.0	97.6
	0.16	5.6	5.4	6.2		94.2	94.4	96.4	
	0.32	6.4	6.0	7.0		96.2	96.0	98.2	
75	0.08	4.4	4.4	7.0	3.6	90.6	90.6	91.8	99.2
	0.16	4.2	4.6	7.2		96.8	96.6	96.8	
	0.32	4.8	5.6	7.0		98.2	98.4	99.0	
80	0.08	5.6	5.2	6.8	4.4	93.2	93.0	94.4	99.8
	0.16	5.4	5.0	6.8		98.0	97.4	97.8	
	0.32	6.6	6.2	8.8		98.6	99.0	98.4	
85	0.08	4.8	6.0	6.6	5.6	91.4	92.6	94.8	99.8
	0.16	4.4	5.0	6.6		97.4	97.0	94.4	
	0.32	5.4	5.2	7.8		98.4	99.0	94.4	

Table 3.1. Size and power for the three versions of the nonparametric test of Cook & Weisberg (1983), using simulated data from a linear model $y_i = 1+2x_i+\varepsilon_i$, where ε_i is normal with zero-mean and standard deviation $\sigma_1(x_i)$ and $\sigma_2(x_i)$, $\imath = 1, \ldots, n$, with a variety of sample size n and smoothing parameters h.

From the simulation studies, some of whose results are shown in Table 3.1 and Table 3.2 is possible to conclude:

1. In all cases, the size of the test, indicated by the results for $\sigma_1(x)$, is close to the target value of 5%. In addition, there is very little difference between the performance of the bootstrap and the other two tests. Since the bootstrap provides a direct means of generating the empirical distribution of the test statistic T, these results therefore provide confirmation that the approximation considered in this paper for the distribution of T is effective.

2. The approximation of the variance-covariance matrix of the residuals by a diagonal matrix can be considered an adequate approximation since its effect on the performance of the test is very slight. Even in the case of non-independent residuals the computer time necessary for the calculation of the p-value is not an obstacle, but the approximation in the independence case avoid the calculation of non-diagonal covariance matrix entries by numerical integration.

3. The boostrap technique to obtain the empirical distribution of the test statistic T, as it was used here, confirms that the approximation used for the distribution of the test statistic is reasonably good.

4. The most relevant parameters related with the power of the test are the sample size n and the shape of the functional dependency of the variance on the design points. In fact, for the same functional dependence for the variance in Table 3.1 and Table 3.2, the power increases considerably for increasing values of n and, for the same value of n, the power of the test increases from the function $\sigma_2(x) = (0.25 + x)$ to the corresponding

one for the function $\sigma_4(x) = 0.25exp(xln(5))$ which are both monotonic in the interval of design.

5. For the monotonic functions σ_3 and σ_4 the power of the test is quite good even though it is lower than the power for the Cook & Weisberg test. However for the non-monotonic function σ_4 this power is better than the power of the latter test.

6. The range of smoothing parameters considered is very large, changing by a factor of four. Despite this, the change in power is relatively small with a small increase as h increases for σ_2 and a small decrease for σ_3.

3.5 A reference band

The idea of smooth kernel estimation can also be used as a tool in order to check the local behaviour of the variance in a linear regression model. When the null hypothesis is rejected, it could be interesting to check how different the variance at a particular fixed point x is from a constant variance. The variable $s(x) = \sqrt{|r(x)|} - E(\sqrt{|r(x)|})$, where $r(x)$ is the least squares residual at the design point x, can be used to build reference bands. These are very useful tools in graphical exploration and they allow comparison of the variances of errors at different design points. In fact, because of the approximate 0-mean normality of $s(x)$, the random variable

$$Q(x) = \frac{\hat{s}(x) - \bar{s}}{\sqrt{var(\hat{s}(x) - \bar{s})}} \tag{3.7}$$

has a standard normal distribution under H_0. Then, for each value of x in the interval and under the null hypothesis of homoscedasticity, it is expected with an approximate probability of 95% that

$$-1.96\sqrt{var(\hat{s}(x)-\bar{s})} \leq \hat{s}(x)-\bar{s} \leq 1.96\sqrt{var(\hat{s}(x)-\bar{s}} \tag{3.8}$$

Therefore, if $q_1(x) = -1.96\sqrt{var(\hat{(s)}(x)-\bar{s})}$ and $q_2(x) = 1.96\sqrt{var(\hat{s}(x)-\bar{s})}$, the 95%-reference band is defined as:

$$Rb_{0.95} = \{(x,y)|a \leq x \leq b; q_1(x) \leq y \leq q_2(x)\} \tag{3.9}$$

Under the null hypothesis it is expected that each point (x,y), where $a \leq x \leq b$ and y= $\hat{s}(x)-\bar{s}$, must belong to the $Rb_{0.95}$ with a probability of 95% and, therefore, this methodology can be used as an approximate graphical tool for checking homoscedasticity.

Explicit expressions for q_1 and q_2 can be calculated because their values depend only on the variances and covariances of the variables $s_i, \quad i = 1,\ldots,n$. In fact,

$$\begin{aligned} var(\hat{s}(x)-\bar{s}) &= \sum_{k=1}^{n}(w_k(x)-(1/n))^2 var(s_k(x)) \\ &+ \sum_{l=1}^{n}\sum_{m=1}^{n}(w_l(x)-(1/n))(w_m(x)-(1/n))cov(s_l,s_m) \end{aligned}$$

and, in terms of the moments of the variances of residuals,

$$\begin{aligned} var(\hat{s}(x)-\bar{s}) &= \sum_{k=1}^{n} a_k(x)^2 0.12\sqrt{var(r_k(x))} \\ &+ \sum_{l=1}^{n}\sum_{m=1}^{n} a_l(x)a_m(x)(E(s_l s_m) - 0.68\sqrt{var(r_l)var(r_m)}) \end{aligned}$$

where $a_k = w_k(x) - (1/n)$, and $w_k(x)$ is the kernel normal function evaluated at x.

Under the null hypothesis, $\text{var}(r_k) = h_{kk}$ is the k^{th} entry in the diagonal of the matrix $(I - X(X^TX)^{-1}X^T)$ multiplied by σ_0^2, $k = 1, \ldots, n$; and $cov(s_l, s_m)$ in the second term of the expression below can be calculated through the joint distribution of (r_l^2, r_m^2), l,m=1,..,n by numerical integration. If the residuals are considered approximately independent, the second term in the expression for the variance below is zero and the reference band can be easily calculated.

Figure 3.5 shows the 95%-reference bands for the different examples given before.

Plots (a) and (b) in this figure are the reference bands for the snowgeese data and the snowgeese data with $x \leq 100$, as explained in the first example. In (a), the p-value was less than 0.003 and, consequently, the points on the smooth curve are far outside of the reference band for almost every value of the independent variable. In (b) the p-value for the test was considerably greater than that for the whole data set as considered in (a). However, it was still small enough to identify significant heteroscedasticity in the linear model. Again the reference band contains only a small arch of the observed curve. Plot (c) presents the 95% reference band for the male cats data from the second example. In this case the p-value for the test is 0.084 and the evidence of heterocedasticity is not conclusive. The smooth curve strays outside the band only for a small proportion of x-values.

In (d), the 95% reference band for the simulated data in the example 3 is shown. In this example, the null hypothesis of homoscedasticity is rejected

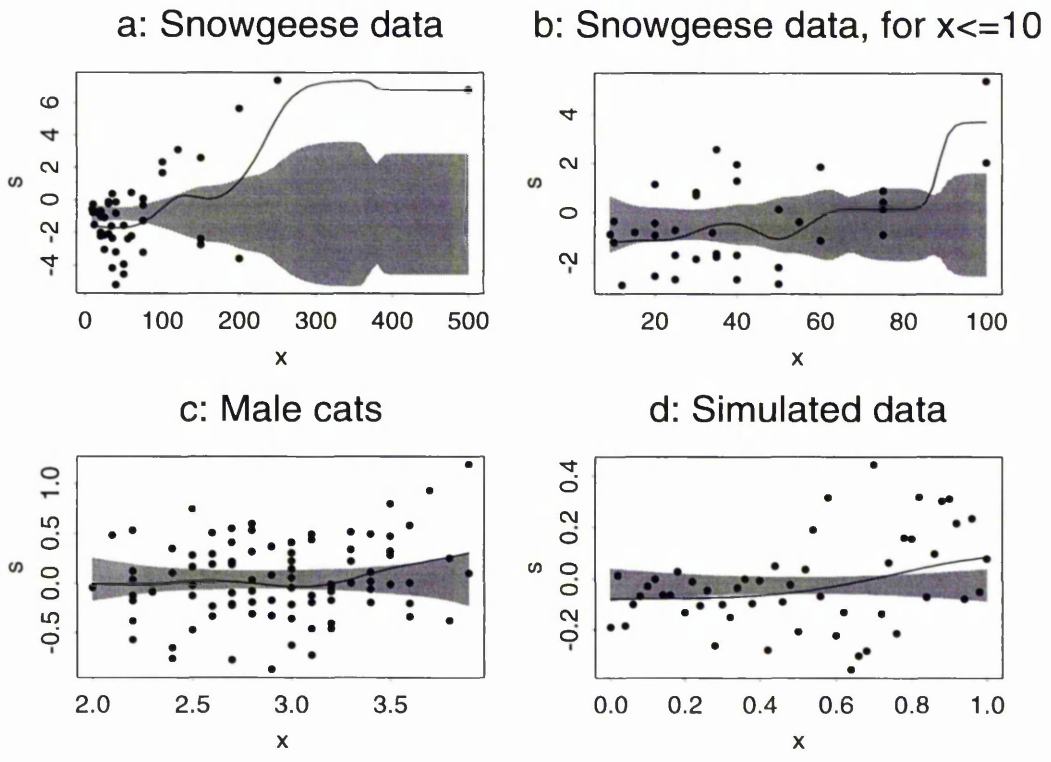

Figure 3.5. This figure shows the 95%-reference bands, the smooth curve $\tilde{s}$, and the values of the variable $s = \sqrt{|r|} - E(\sqrt{|r|})$, where r is the corresponding residual, for each example in the text. The shadow areas are the 95% reference bands. The curve in each panel is the smooth curve $\tilde{s}$, and the dots are the values of the variables s, for each residual r.

with a p-value of 0.048.

Figure 3.5 reflects once again the relationship between the observed curve and the reference band.

3.6 Some remarks

One of the most important advantages of using nonparametric techniques to check constant variance is that no particular shape of variance pattern is assumed. This widens the scope of the available tools. However this generalization will necessarily lead to some reduction in power over parametric methods when the parametric assumptions provide a good description of the true pattern. The results of Section 3.4 show that the power of the nonparametric test can reach very reasonable levels even in cases where changes in variance are not marked, and not always easily identified visually.

Reference bands are very useful graphical resources for checking homoscedasticity in a linear model. As was shown in the examples, they may be a very good illustrative complement of the formal test statistic performed in this approach.

The bandwith h is an important parameter in smoothing techniques. Bowman & Young (1996) reviewed a number of nonparametric tests. In the present case the bandwidth seems to have little influence on the power of the test for reasonable large sample sizes ($n = 70$ in table 3.1). For smaller values of n, the power of the test may be affected by the shape of the variance function under the alternative hypothesis as illustrated in 3.1. Large h gives the highest power for σ_2 and σ_4 whereas $h = 0.16$ seems to be best for σ_3.

Reference bands for the simulated data, corresponding to a wide range of smoothing parameters are shown in the Figure 3.6. The values used are h =0.16, 0.32, 0.48 and 0.64. In each case the information conveyed by the

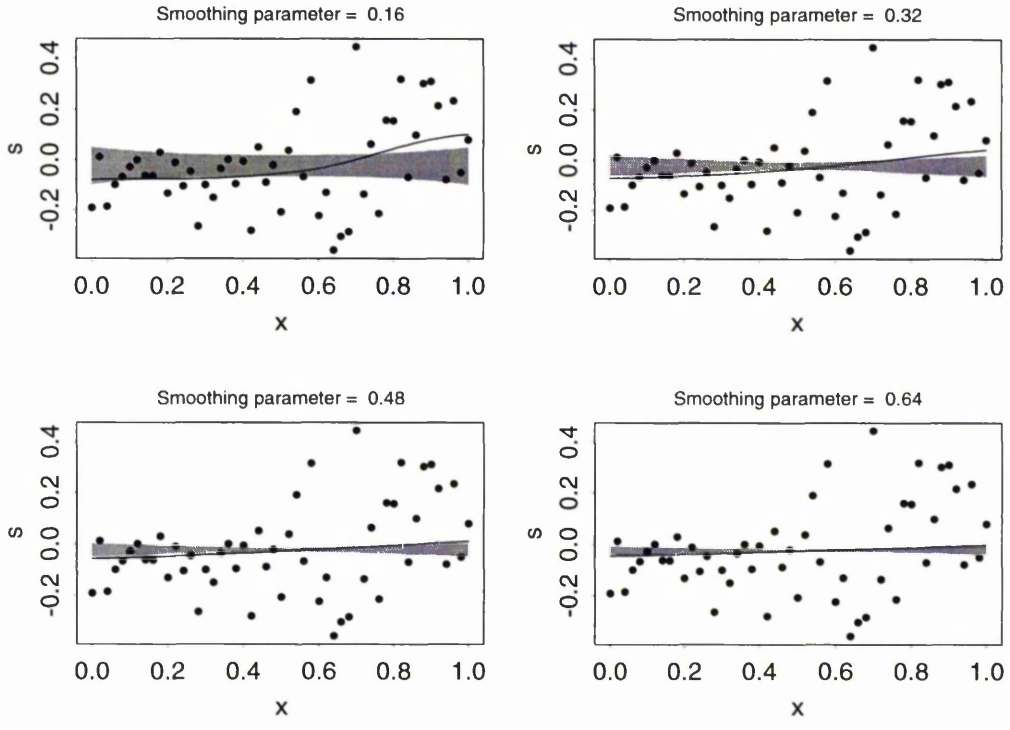

Figure 3.6. Reference bands for a range of smoothing parameters with the simulated Data

reference bands is the same, demonstrating that the particular choice of this parameter is not crucial. There will, of course, be cases, where the conclusion does change with h. As an automatic technique, a "plug-in" bandwidth selection technique, such as the one described by Gasser et al (1991), could be used. These techniques assume independent observations. However, it has already been demonstrated in Section 3.4 that the adoption of this approximation does not greatly affect the results.

In Figure 3.7 plots of reference bands and smooth curves are shown for

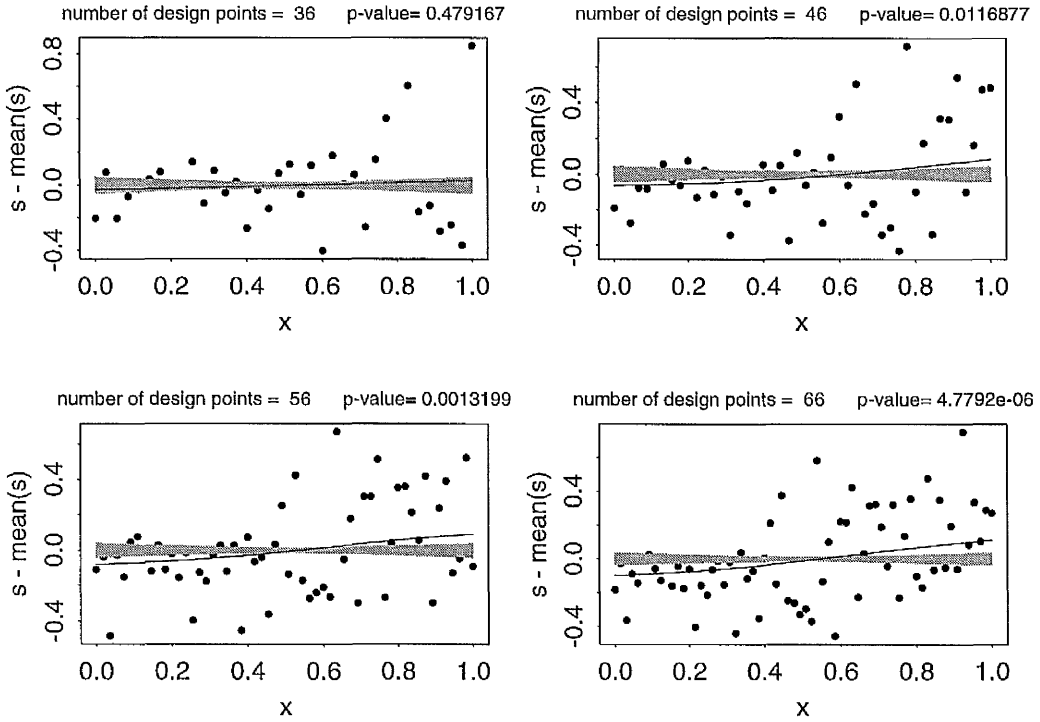

Figure 3.7. Reference bands for different number of design points. The data were obtained by simulation. The variances of errors follow linear function of Table 3.1.

the case where the functional dependence of the variance of errors as a linear function of the design points for different sample sizes in the interval [0,1]. The corresponding p-values for the test are printed. This graph is an illustration of the dependence of the power of the test on the number of design points and the relationship with the relative position of the points on the smooth curve and the reference bands. The variance of errors was the same linear function for the simulation and bootstrap studies on table 3.1 and 3.2. The value for the smoothing parameter here was 0.32. As expected, the power of this test

has a strong dependence on the number of design points. However from this figure and the simulation studies it is possible to conclude that for a reasonable number of design points the test has good power to detect heteroscedasticity. Even in cases such as those of Figure 3.7, where the dependence of variance on the design points is not obvious, the test can warn about changes in the variance.

Even though the assumption of indepndent errors was used for the linear model here, this assumption might be relaxed in a further study. This extension would include a large number of different linear models as those involving time series and spatial linear processes. In these cases the concept of heteroscedasticity can be extended to the concept of dependence. This will be explored in the following chapter.

		$\sigma_3(x) = 0.25 + 4(x-0.5)^2$				$\sigma_2(x) = 0.25exp(ln(5)x)$			
		non-ind. resid.	indep. resid.	boots-trap	C. & W.	non-ind. resid.	indep. resid.	boots-trap	C. & W.
n	h	%	%	%	%	%	%	%	%
30	0.08	55.8	56.2	69.2	30.0	49.4	50.6	50.2	63.2
	0.16	63.2	64.6	75.0		63.0	64.8	55.8	
	0.32	32.0	35.4	45.2		73.6	75.4	82.0	
35	0.08	66.4	65.2	78.4	32.4	56.4	55.2	67.4	92.8
	0.16	74.0	73.4	85.6		72.8	73.0	81.8	
	0.32	45.0	44.4	70.0		82.6	82.6	89.4	
40	0.08	76.8	77.6	80.0	29.2	64.0	66.4	72.4	95.8
	0.16	82.4	84.2	87.8		78.4	80.6	85.8	
	0.32	54.0	56.4	75.0		87.2	87.6	92.8	
45	0.08	84.2	85.0	86.2	31.4	70.4	70.4	80.8	95.6
	0.16	86.0	86.4	92.0		84.4	84.8	90.8	
	0.32	63.8	64.4	82.2		90.4	90.4	94.2	
50	0.08	88.2	88.4	93.2	30.4	81.4	82.2	84.4	98.6
	0.16	92.0	92.2	96.4		89.0	89.6	92.6	
	0.32	69.2	70.4	89.2		94.0	94.2	96.6	
55	0.08	90.4	90.6	93.6	30.6	85.0	84.8	88.8	98.8
	0.16	93.6	93.8	97.8		93.2	93.2	96.0	
	0.32	75.0	75.4	94.0		96.4	96.0	98.0	
60	0.08	91.4	92.2	94.8	32.0	86.4	87.4	89.6	99.6
	0.16	95.2	96.4	97.2		95.4	95.6	96.4	
	0.32	77.4	78.8	93.4		98.0	98.0	98.6	
65	0.08	93.8	94.4	97.0	29.0	90.0	91.0	93.4	99.8
	0.16	96.8	97.2	99.8		96.4	97.0	97.8	
	0.32	83.2	84.6	99.8		97.2	98.2	99.6	
70	0.08	97.2	97.2	97.8	35.6	93.4	93.8	93.0	99.6
	0.16	98.4	98.2	98.4		96.8	96.6	98.8	
	0.32	90.4	90.8	97.6		98.2	98.2	94.2	
75	0.08	97.8	97.8	99.2	33.4	94.4	94.6	97.4	99.8
	0.16	99.2	99.2	99.8		97.6	97.4	98.8	
	0.32	91.2	91.6	96.8		99.0	99.2	99.8	
80	0.08	99.0	99.2	99.8	31.4	97.8	97.4	96.6	99.8
	0.16	99.2	99.2	99.8		99.4	99.4	99.0	
	0.32	93.6	93.8	99.2		99.4	99.4	99.8	
85	0.08	98.8	98.8	99.8	34.4	96.8	97.0	98.0	99.8
	0.16	99.8	99.8	99.6		99.0	99.0	99.0	
	0.32	94.6	95.8	99.8		99.6	99.6	99.8	

Table 3.2. Power for the three versions of the nonparametric test of Cook & Weisberg (1983), using simulated data from a linear model $y_i = 1 + 2x_i + \varepsilon_i$, where ε_i is normal with zero-mean and standard deviation $\sigma_3(x_i)$ and $\sigma_4(x_i)$, $\imath = 1, \ldots, n$, with a variety of sample size n and smoothing parameters h.

Chapter 4

Testing for Constant Variogram

4.1 Introduction

Spatial phenomena are usually represented in the statistical context by a process $\{Y(s) : s \epsilon D\}$ where D is a suitable set in $\mathbb{R}^q$. Assumptions on the process are necessary in order to build an adequate model for the specific phenomenon to be analyzed. If a model can be proposed , then it will be possible to carry out inference on its parameters when a set of data is available. From the statistical point of view one of the most basic and important feature of a process is its correlation or covariance structure. Independence of the variables involved in a spatial model a is very common and convenient assumption because it makes the modeling distribution theory easier and more manegeable. However, the assumption of a suitable dependence or correlation structure is often more realistic. At least this feature should be checked at the initial steps of a statistical analysis. The variogram is the traditional quantity whose

values show how strongly correlation is linked to spatial locations.

The way followed traditionally in order to study the variogram of a particular process when a set of data is available is to estimate the variogram, to choose a particular family for the variogram with the information given for this estimate, and finally to fit a particular member of the family with the data. The tools to resort to for checking the estimated variogram in order to choose an adequate family to fit it, are traditionally graphical tools.

When examining correlation structures, an initial and very important question to be answered is whether the variables in the process are correlated or not. The methodology proposed here is an attempt to answer this question.

For a second-order stationary process, the covariance function and the variogram depend on the relative positions between each pair of points in the domain. Also, if its variance is constant (i.e. σ^2) through the domain, the covariance function and the variogram are related one to each other by:

$$2\gamma(\boldsymbol{h}) = 2\sigma^2 - 2C(\boldsymbol{h})$$

where $2\gamma(\boldsymbol{h})$ is the variogram at $\boldsymbol{h} = \boldsymbol{s_1} - \boldsymbol{s_2}$ and $C(\boldsymbol{h})$ is the covariogram. The **variogram** of the process as defined by Matheron (1962) is:

$$2\gamma(\boldsymbol{s_1} - \boldsymbol{s_2}) = var(Y(\boldsymbol{s_1}) - Y(\boldsymbol{s_2})), \quad for \quad all \quad \boldsymbol{s_1}, \boldsymbol{s_2} \epsilon D \qquad (4.1)$$

and the function γ is called **semi-variogram** of the said process. Then, a constant variogram implies a constant covariogram and, particularly interesting is the case of a null covariogram or uncorrelated process. If, aditionally,

the process is isotropic, the variogram is a curve in the plane and its properties can be assessed as those for a real function. Also, the local variation of the variogram as a curve is expected to be smooth. Hence, one natural approach to checking if the variogram is constant is to compare it as a function to a constant.

The general assumptions of a linear model, where the trend is a linear function of covariates which depend on the spatial locations or on other variables and the errors are normal zero-mean, may also be adopted. Transformed differences of square residuals are used as a measure of how much deviation from a constant the variogram displays. For this goal, it is shown here that nonparametric smoothing techniques are adequate. Among these techniques, those dealing with kernel estimation of smooth functions are considered, as in the case of homoscedasticity for the linear model. The test statistic proposed for the hypothesis of constant variogram, and, its approximated distribution, are similar (at some stages) to those described for the test for homoscedasticity in the linear model.

Reference bands for the kernel smoothed version of a transformation of differences of residuals under the null hypothesis of constant variogram are also built. These reference bands are constructed with the square root of absolute differences of residuals for each distance in the set of observed locations. They are again very useful graphical tools to check whether there is evidence of spatial correlation, and to assess its size.

The test and reference bands are applied to some data concerned with some chemical substances used as indicators of pollution in rivers and billabongs.

4.2 Previous literature

Traditionally, the subject concerned with the study of models for spatial data has been included in the discipline known as Geostatistics. Even though its early origin can be attributed to Matheron in the sixties, it was in the eighties when several contributions dealing with the development of specific statistical tools for problems coming from Geology and Mining Engineering gave these tools a distintict identity. With the passing of time, several of these tools have been used in other scientific contexts but still nowdays some of them keep their original names. One example is an approach used for spatial prediction which is recognized as kriging as it was called by Matheron in honour of Krige, a mining engineer who developed empirical methods for determining empirical ore-grade distributions.

From the statistical point of view one of the most important distinguishing features of the models traditionally considered as a part of Geostatistics deals with the characterization of the set of indexes for the spatial process to be modelled. In fact, the processes $\{Y(\boldsymbol{s}) : \boldsymbol{s}\epsilon D\}$ usually modelled in Geostatistics assume the set D to be a non-zero volume set in $\mathbb{R}^q$, or the spatial index $\boldsymbol{s}$ varies continuously over a subset of $\mathbb{R}^q$. However methods usually associated with processes defined on sets of indexes with other characteristics, such as point-pattern or lattice processes, can be borrowed from a different type of process. In the approach presented here the set D is assumed to be a non-zero volume set in $\mathbb{R}^q$. The whole work is fixed in a set in two-dimensional lattice even though some of the tools used to develop are in $\mathbb{R}^2$ all the results can be extended straightforwardly to higher dimensions.

Another distinguishing feature for the models dealing with Geostatistics and which is the subject of the approach considered here is the capacity of the model to recognize spatial variability at both large and small scale. In other words, geostatistical models are usually able to include both trend and spatial correlation.

One of the first people to stress the importance of modelling spatial correlation was Watson (1972). He compared two different approaches, with and without spatial correlation structure, and pointed out that most geological problems exhibite strong positive correlation between data at nearby spatial locations.

The hypothesis of independence in spatial models when the "real" data are dependent or correlated can distort considerably results on estimation, prediction and designs as pointed out by Cressie (1991). He considered an example of $Z(1), \ldots, Z(n)$ independent and identically distributed Gaussian variables with unknown mean and known variance σ_0^2. The minimum-variance and unbiased estimator of μ is the widely known sample mean $\bar{Z}$ which is also Gaussian with mean μ and variance σ_0^2/n. If the data are not independent but they are positively correlated and $cov(Z(i), Z(j)) = \sigma_0^2.\rho^{|i-j|}, \quad i, j = 1, \ldots, n, \quad 0 < \rho < 1$ then the variance of the sample mean is: $var(\bar{Z}) = \{\sigma_0^2/n\}[1 + 2\{\rho/(1-\rho)\}\{1 - (1/n)\} - 2\rho/(1-\rho)^2(1-\rho^{n-1})/n$. If n=10 and $\rho = 0.26$ the variance of $\bar{Z}$ for the correlated data is $var(\bar{Z}) = (\sigma_0^2/n)1.608$, i.e. 1.608 times the variance of the same estimator under the assumption of independence. Other interesting examples of the effect of dependence (under the assumption of indendence) on prediction and experimental designs can be found in this excellent book by Cressie.

Also in the context of estimation, Haining (1988) compares the variance of

the sample mean assuming independence with its variance assuming positive dependence for random variables in $\mathbb{R}^2$ which are conditionally specified autoregressions and simultanously specified autoregressions and moving averages.

In the context of estimation, a natural estimator for the variogram of a process with constant mean, based on the method of moments is:

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} (Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))^2$$

where $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h};\ \ i, j = 1, \ldots, n\}$ (Matheron (1962)). This estimator is unbiased if the process $\{Y(\boldsymbol{s}) :\ \ \boldsymbol{s} \in D\}$ is intrinsically stationary (constant mean and second-order stationary). If the process is also gaussian the squared differences $(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))^2$ are distributed as $2\gamma(\boldsymbol{h}).\chi^2(1)$, where $\chi^2(1)$ is the chi-square distribution with one degree of freedom. Then, $E((Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))^2) = 2\gamma(\boldsymbol{h})$. But in this case (gaussianity and intrinsic stationarity), the distribution of these squared differences are highly skewed.

On the other hand, the distribution of the square root of the absoluted value of these differences, $|Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)|^{1/2}$ is approximately normal in the sense that it has a skewness and kurtosis closer to those of a normal distribution (see, Cressie and Hawkins (1980)). These properties encouraged Cressie and Hawkins (1980) to propose an estimator for the variogram of a process $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ which is based on the variables $|Y(\boldsymbol{s}_i - Y(\boldsymbol{s}_j)|^{1/2}$ for $\boldsymbol{s}_i, \boldsymbol{s}_j \in D$. This

estimator is:

$$2\tilde{\gamma}(\boldsymbol{h}) = \left\{ \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} |Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)|^{1/2} \right\}^4 \Big/ (0.457 + \frac{0.494}{|N(\boldsymbol{h})|})$$

where the constant $0.457 + \frac{0.494}{|N(\boldsymbol{h})|}$ is the correction for bias. Another important reason to resort to this transformation rather than the classical squares of the previous differences is their behaviour under dependence of the variables in the process. In fact, if the normal variables X_1, X_2 have a correlation coefficient $\rho = corr(X_1, X_2)$, then the square variables have correlation coefficient $corr(X_1^2, X_2^2) = \rho^2$ whereas the $corr(|X_1|^{1/2}, |X_2|^{1/2})$ is less than ρ^2. This property increases the efficiency of the averaging of these variables in estimating the variogram.

Chauvet (1989) suggested that the plot of the variogram-cloud in a given direction $\boldsymbol{e}$ is a very usuful graphical tool. The variogram-cloud is a two-dimensional x-y plot where x is the distance h between two points and y is the value of $(Y(\boldsymbol{s}_i + h\boldsymbol{e}) - Y(\boldsymbol{s}_j))^2$, where $\boldsymbol{s}_i + h\boldsymbol{e} = \boldsymbol{s}_j$ and $(i, j) \in N(\boldsymbol{h})$

Following this direction, Cressie (1991) suggested to resort to a square-root-differences cloud as a more efficient graphical tool.

Robinson (1990) gave a stringent condition for the existence of the variogram rather than the covariogram. In fact, there is a quite widely known example, which is the model to describe Brownian motion which has variogram but not covariogram. The variogram for this case is: $2\gamma(h) = (1/2)\sigma^2 h$ where σ^2 is the constant variance of the process. However the covariance for two variables $Z(t+h)$ and $Z(t)$ is: $cov(Z(t+h), Z(t)) = (1/2)\sigma^2 t$. This function depends

on t and, consequently the covariagram is not defined.

The variogram is a crucial parameter also in the estimation procedure called kriging. This method is a minimum-mean-square method which usually depends on the second-order properties of the process involved, i.e. on the variogram and/or covariogram. More precisely, kriging is a method to search for an optimal predictor for the value of a function g of the process Z at the point s_0. In a general approach, kriging is the procedure followed to calculate a predicted value $\hat{p}(\boldsymbol{Z}; s_0) = \boldsymbol{\lambda}'\boldsymbol{Z}$, where s_0 is a spatial location, $\boldsymbol{Z}$ a vector of a realization of a spatial process, and $\boldsymbol{\lambda}$ depends on the variogram or covariogram. Either the covariogram or the variogram used in kriging should be fitted to a valid family rather than be estimated directly because the estimators do not often verify properties of non positive definiteness (for the variogram) or non negative definiteness (for the covariogram).

These are, among others, more than enough reasons to justify and also encourage the search for adequate tools to check properties of the variance-covariance structure of a stochastic process.

4.3 A test statistic

Suppose the spatial data $\{Y(s_1), Y(s_2), \ldots, Y(s_n)\}$ observed at known spatial locations $\{s_1, s_2, \ldots, s_n\}$ in a set $D \subseteq \mathbb{R}^q$ with positive q-volume are modelled as a collection of random variables generated by the random process:

$$Y(\boldsymbol{s}) = \sum_{i=1}^{m} \beta_i x_i(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$$

where $s \epsilon D$, $x_i, i = 1, \ldots, m$ is a collection of m non-random explanatory variables which may or not depend on the spatial locations $\boldsymbol{s}$ and $\varepsilon(.)$ is an error random process with zero-mean and finite variance. An example of this model is the case in which the explanatory variables are a constant $x_1(\boldsymbol{s}) = 1$ and the coordinates of spatial locations, $x_2(\boldsymbol{s}) = s_1$ and $x_3(\boldsymbol{s}) = s_2$, where $(s_1, s_2) = \boldsymbol{s}$.

It is useful to write the model for the observed data in matrix notation:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where X is a nxm matrix whose (ij)th entry is $x_j(\boldsymbol{s}_i)$ the observed value of the variable x_j at the location $\boldsymbol{s}_i$, $\boldsymbol{\beta}$ is a m-vector of unknown parameters, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)^T$ and $\boldsymbol{\varepsilon}(\boldsymbol{s}) = (\varepsilon(\boldsymbol{s}_1), \varepsilon(\boldsymbol{s}_2), \ldots, \varepsilon(\boldsymbol{s}_n))^T$, the n-vector of errors at each location $\boldsymbol{s}_i$, i=1,...,n.

The process $\boldsymbol{\varepsilon(s)}$ is assumed here as a **second-order stationary** process,e.g.,

$$E(\varepsilon(\boldsymbol{s})) = 0, \quad for \ \ all \ \ s \epsilon D \tag{4.2}$$

$$cov(\varepsilon(\boldsymbol{s_1}), \varepsilon(\boldsymbol{s_2})) = C(\boldsymbol{s_1} - \boldsymbol{s_2}), \quad for \ \ all \ \ \boldsymbol{s_1}, \boldsymbol{s_2} \epsilon D \tag{4.3}$$

Furthermore, it is assumed in this study that the covariance of two random variables $Y(\boldsymbol{s_1})$ and $Y(\boldsymbol{s_2})$ in the process is a function only of the distance

$h_{12} = \|s_1 - s_2\|$. In other words, the process Y is assumed **isotropic**

$$Cov(Y(s_1), Y(s_2)) = C(\|s_1 - s_2\|), \quad for \quad all \quad s_1, s_2 \epsilon D \tag{4.4}$$

As a consequence of the latter property, the variance of each variable Y(s) in the process is constant,

$$var(Y(s)) = \sigma^2, \quad for \quad each \quad s \epsilon D \tag{4.5}$$

The **variogram** of the process 4.1:

$$2\gamma(s_1 - s_2) = var(Y(s_1) - Y(s_2)), \quad for \quad all \quad s_1, s_2 \epsilon D$$

is also a function of the distance between points in D if the process is isotropic. In this case:

$$2\gamma(s_1 - s_2) = var(Y(s_1) - Y(s_2)) = 2\sigma^2 - 2cov(Y(s_1), Y(s_2))$$

$$= 2\sigma^2 - 2C(s_1 - s_2) = 2\sigma^2 - 2C(\|s_1 - s_2\|) = 2\gamma(\|s_1 - s_2\|), \quad for \ all \ s_1, s_2 \epsilon D$$

If the variogram is constant for each value of $h = \|s_1 - s_2\|$ then the covariogram $cov(Y(s_1), Y(s_2)) = cov(h)$ is also constant.

For the **distribution** of the vector of errors,

$$(\varepsilon(s_1), \varepsilon(s_2), \ldots, \varepsilon(s_n))^T,$$

a multivariate normal distribution $N(\boldsymbol{O}, \Sigma_{\boldsymbol{\varepsilon}})$ is considered.
The hypothesis to be tested can be written as:

$$\begin{aligned} H_0 : 2\gamma(h) &= 2\sigma_0^2 \\ H_1 : 2\gamma(h) &= \textit{smooth} \text{ function of } h, \quad \textit{for all } h. \end{aligned}$$

where σ_0^2 is a constant.

If $\sigma_0^2 = \sigma^2 = var(\varepsilon(\boldsymbol{s}))$ then $cov((Y(\boldsymbol{s_1}), Y(\boldsymbol{s_2}))$=0 and, under the normality assumption, the process $\boldsymbol{\varepsilon}$ is a white-noise process.
If the null hypothesis is true but $\sigma_0^2 \neq \sigma^2$ the covariance is a non-zero constant (c=$\sigma^2 - \sigma_0^2$) and the process shows a particular correlation structure.

Hence, under the null hypothesis, the variance-covariance matrix of errors can be written as:

$$\Sigma_{\boldsymbol{\varepsilon}} = \begin{bmatrix} c+\sigma_0^2 & c & \dots & c \\ c & c+\sigma_0^2 & \dots & c \\ . & . & \dots & . \\ . & . & \dots & . \\ c & . & \dots & c+\sigma_0^2 \end{bmatrix}$$

where c is the constant value for the covariance.

The case in which $c = 0$ is particularly interesting because it implies diagonal variance-covariance matrix of errors, $\Sigma_{\boldsymbol{\varepsilon}}$, and consequently under the hypothesis of normality, it means independence of errors. On the other hand, the

case $c \neq 0$ suggests an extremely particular and unusual covariance structure. Hence, from now and up to be explicitly expressed, the case considered here will be $c = 0$. In other words, a null hypothesis which is equivalent to independency of errors will be considered. Then, the least squares estimated values for the linear model (1) under the null hypothesis are:

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T Y$$

The least square residuals are $\boldsymbol{r} = \hat{\boldsymbol{Y}} - \boldsymbol{Y}$ where $\hat{\boldsymbol{Y}} = X\hat{\boldsymbol{\beta}}$.

Because the variogram is defined as the variance of the difference of two variables in the process, a linear model based on these differences will be built now. Through this model a natural way will be shown in which the tools applied to testing homoscedasticity in a linear regression can be suitably modified to be applied in this context.

For each value h of the distance between two observed locations $\boldsymbol{s_1}, \boldsymbol{s_2} \epsilon D$, $h = \|\boldsymbol{h}\| = \|\boldsymbol{s_1} - \boldsymbol{s_2}\|$, the differences

$$U(\boldsymbol{h}) = Y(\boldsymbol{s} + \boldsymbol{h}) - Y(\boldsymbol{s})$$

can be modelled as

$$U(\boldsymbol{h}) = \sum_{l=1}^{m} \beta_l (x_l(\boldsymbol{s} + \boldsymbol{h}) - x_l(\boldsymbol{s})) + (\varepsilon(\boldsymbol{s} + \boldsymbol{h}) - \varepsilon(\boldsymbol{s}))$$

$$or \quad U(\boldsymbol{h}) = \sum_{l=1}^{m} \beta_l v_l(\boldsymbol{h}) + \varepsilon^*(\boldsymbol{h})$$

where $v_l(\boldsymbol{h}) = x_l(\boldsymbol{s} + \boldsymbol{h}) - x_l(\boldsymbol{s})$ and $\varepsilon^*(\boldsymbol{h}) = \varepsilon(\boldsymbol{s} + \boldsymbol{h}) - \varepsilon(\boldsymbol{s})$.

For the observed spatial points $\{\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_n}\}$ in a set $D \subseteq \mathbb{R}^q$, the corresponding differences between them are $\boldsymbol{h_{ij}} = \boldsymbol{s_i} - \boldsymbol{s_j}$, for $i, j = 1, \ldots, n$ and the observed values $U(\boldsymbol{h_{ij}})$ can be expressed as:

$$\boldsymbol{U} = V\boldsymbol{\beta} + \boldsymbol{\varepsilon^*}$$

where $\boldsymbol{U} = (U(\boldsymbol{h_{12}}), \ldots, U(\boldsymbol{h_{1n}}), \ldots, U(\boldsymbol{h_{23}}), \ldots, U(\boldsymbol{h_{2n}}), \ldots, \ldots, U(\boldsymbol{h_{(n-1)n}}))$; V is a Nxm $(N = n(n-1)/2)$ matrix whose $(ij)^{th}$ entry is $v_j(\boldsymbol{h_{kl}})$ for some k,l such that $(\boldsymbol{h_{kl}})$ is the vector between the two observed points $\boldsymbol{s_k}$ and $\boldsymbol{s_l}$ in D such that $1 \leq k < l \leq n$ and $\boldsymbol{\varepsilon^*}$ is the vector:

$$\boldsymbol{\varepsilon^*} = (\varepsilon^*(\boldsymbol{h_{11}}), \ldots, \varepsilon^*(\boldsymbol{h_{1n}}), \ldots, \varepsilon^*(\boldsymbol{h_{23}}), \ldots, \varepsilon^*(\boldsymbol{h_{2n}}), \ldots, \ldots, \varepsilon^*(\boldsymbol{h_{(n-1)n}}))$$

In order to simplify the notation, one subcript for the differences between observed points and their norms will be used afterwards. Then, each vector $\boldsymbol{h_{ij}} = \boldsymbol{h_i} - \boldsymbol{h_j}$ will be expresed as $\boldsymbol{h_k}$ for some k as the result of a bijection between the set of pairs $\{(1,2), \ldots, (1,n), (2,3), \ldots, (2,n), \ldots, \ldots, (n-1,n)\}$ and the set of positive integers $\{1, 2, \ldots, N\}$ with $N = n(n-1)/2$ as mentioned before. With this notation, the new vector of errors is

$$\boldsymbol{\varepsilon^*} = (\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_N^*)$$

has a normal distribution $N(\boldsymbol{O}, \Sigma_{\boldsymbol{\varepsilon^*}})$. The variance-covariance matrix $\Sigma_{\boldsymbol{\varepsilon^*}}$ has entries σ_{ij}^* where:

$$\sigma_{ii}^* = var(\varepsilon_i^*) = var(\epsilon(\boldsymbol{s} + \boldsymbol{h_i}) - \epsilon(\boldsymbol{s})) = 2\sigma^2 - 2cov(\|h_i\|)$$

$$\sigma_{ij}^* = cov(\varepsilon_i^*, \varepsilon_j^*) = E((\varepsilon(\boldsymbol{s_k} + \boldsymbol{h_i}) - \varepsilon(\boldsymbol{s_k}))(\varepsilon(\boldsymbol{s_l} + \boldsymbol{h_j}) - \varepsilon(\boldsymbol{s_l})) =$$

$$= E(\varepsilon(\boldsymbol{s}_k + \boldsymbol{h}_i)\varepsilon(\boldsymbol{s}_l + \boldsymbol{h}_j)) - E(\varepsilon(\boldsymbol{s}_k + \boldsymbol{h}_i)\varepsilon(\boldsymbol{s}_l)) - E(\varepsilon(\boldsymbol{s}_l + \boldsymbol{h}_j)\varepsilon(\boldsymbol{s}_k)) + E(\varepsilon(\boldsymbol{s}_k)\varepsilon(\boldsymbol{s}_l)) =$$

$$= cov(||\boldsymbol{h}_i - \boldsymbol{h}_j + \boldsymbol{s}_k - \boldsymbol{s}_l||) - cov(||\boldsymbol{h}_i + (\boldsymbol{s}_k - \boldsymbol{s}_l)||) - cov(||\boldsymbol{s}_k - \boldsymbol{s}_l - \boldsymbol{h}_j||) - cov(||\boldsymbol{s}_k - \boldsymbol{s}_l||$$

$$= cov(||\boldsymbol{h}_i - \boldsymbol{h}_j + \boldsymbol{h}||) - cov(||\boldsymbol{h}_i + \boldsymbol{h}||) - cov(||\boldsymbol{h} - \boldsymbol{h}_j||) + cov(||\boldsymbol{h}||)$$

for each value of $\boldsymbol{h}$ such that $h = ||\boldsymbol{h}||$ is the distance between two points in D. Under the null hypothesis, the covariance function of $\boldsymbol{\varepsilon}$ is constant and $cov(||\boldsymbol{h}||) = \sigma^2 - \sigma_0^2$, for $\boldsymbol{h} \neq \boldsymbol{0}$ and $cov(||\boldsymbol{h}||) = var(\varepsilon) = \sigma^2$, for $\boldsymbol{h} = \boldsymbol{0}$. Hence, the variance-covariance matrix of $\boldsymbol{\varepsilon}^*$, $\Sigma_{\boldsymbol{\varepsilon}^*}$, under the null hypothesis has entries,

$$\sigma_{ij}^* = \begin{cases} 2\sigma_0^2 & if \quad i = j \\ \sigma_0^2 & if \quad \boldsymbol{s}_k + \boldsymbol{h}_i = \boldsymbol{s}_l + \boldsymbol{h}_j \quad or \quad \boldsymbol{s}_k = \boldsymbol{s}_l \\ -\sigma_0^2 & if \quad \boldsymbol{s}_k + \boldsymbol{h}_i = \boldsymbol{s}_l \quad or \quad \boldsymbol{s}_l + \boldsymbol{h}_j = \boldsymbol{s}_k \\ 0 & otherwise \end{cases}$$

In terms of U, the hypotheses considered before can be rewritten as:

$$\begin{aligned} H_0 : var(U(h_i)) &= 2\sigma_0^2, \quad for \quad 1 \leq i \leq N \\ H_1 : var(U(h_i)) &= smooth \ \text{function of} \ h_i, \quad for \quad 1 \leq i \leq N \end{aligned}$$

Then, in terms of the linear model $\boldsymbol{U} = V\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$ the statistical hypotheses are the same as those for the linear regression model considered in another chapter. However the assumptions in both cases are rather different. In fact, the variance-covariance matrix of errors under the null hypothesis for the simple linear model is diagonal. It is not diagonal for the new model. This problem could be sorted out at least in a theoretical way by looking for a different estimator for the linear regression parameters (such as weighted least squares). However, in practice, the sizes of the matrices involved are very big even for a moderate number of sampled points.

As in the approach for testing homoscedasticity in a linear model, the behaviour of a function of the least squares residuals is considered in checking the hypotheses above. The presence of the linear regression parameters $\boldsymbol{\beta}$ then be handled easily by constructing the least squares residuals from the model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. In fact, if $\hat{U} = V\hat{\beta}$ where $\hat{\boldsymbol{\beta}} = (X^TX)^{-1}X^TY$ the vector of residuals residuals here is $\boldsymbol{R} = \boldsymbol{U}(h) - \hat{\boldsymbol{U}}(h)$ and each component $R_i = R(h_i)$ can be expressed as a function of the residuals $\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$. In fact,

$$R_i = R(h_i) = U(h_i) - \hat{U}(h_i) = U(h_i) - \sum_{l=1}^{m} \hat{\beta}_l v_l(h_i)$$

$$= Y(\boldsymbol{s} + \boldsymbol{h_i}) - Y(\boldsymbol{s}) - \sum_{l=1}^{m} \hat{\beta}_l x_l(\boldsymbol{s} + \boldsymbol{h_i}) + \sum_{l=1}^{m} \hat{\beta}_l x_l(\boldsymbol{s})$$

$$= Y(\boldsymbol{s} + \boldsymbol{h_i}) - \hat{Y}(\boldsymbol{s} + \boldsymbol{h_i}) - Y(\boldsymbol{s}) + \hat{Y}(\boldsymbol{s}) = r(\boldsymbol{s} + \boldsymbol{h_i}) - r(\boldsymbol{s})$$

for $i = 1, \ldots, N$ and $\boldsymbol{s}$ a point in D such that $\boldsymbol{s} + \boldsymbol{h}_i \in D$. The vector $\boldsymbol{R}$ has a normal distribution $N(\boldsymbol{O}, \Sigma_R)$ where the entries σ_{Rij} of the matrix Σ_R can be calculated, under the null hypothesis, from the variance-covariance of residuals, $\sigma^2 H = \sigma^2(I - X(X^TX)^{-1}X^T)$ by the expressions:

$$cov(R_i, R_j) = cov(r(\boldsymbol{s} + \boldsymbol{h_i}) - r(\boldsymbol{s}), r(\boldsymbol{t} + \boldsymbol{h_j}) - r(\boldsymbol{t})) =$$

$$cov(r(\boldsymbol{s}+\boldsymbol{h_i}), r(\boldsymbol{t}+\boldsymbol{h_j})) - cov(r(\boldsymbol{s}), r(\boldsymbol{t}+\boldsymbol{h_j})) - cov(r(\boldsymbol{s}+\boldsymbol{h_i}), r(\boldsymbol{t})) + cov(r(\boldsymbol{s}), r(\boldsymbol{t}))$$

where each term of the last expression is an entry of the matrix H multiplied by σ^2, and $\boldsymbol{s}, \boldsymbol{t} \in D$.

Each component R_i of $\boldsymbol{R}$ has a normal distribution $N(0, \sigma^2_{R_i})$ where

$$\sigma^2_{R_i} = var(r(\boldsymbol{s} + \boldsymbol{h_i}) - r(\boldsymbol{s})) = 2\sigma^2 - cov((r(\boldsymbol{s} + \boldsymbol{h_i}), r(\boldsymbol{s}))$$

Hence, again the distributions of the variables $S_i = \sqrt{|R_i|} - E_0(\sqrt{|R_i|})$, where E_0 indicates the expected value calculated under the null hypothesis, are approximately normal and they can be used to build the test statistic:

$$T = \frac{\sum_{i=1}^{N}(S_i - \bar{S})^2 - \sum_{i=1}^{N}(S_i - \tilde{S}_i)^2}{\sum_{i=1}^{N}(S_i - \tilde{S}_i)^2} \tag{4.6}$$

or, as a ratio of two quadratic forms:

$$T = \frac{S^T A S}{S^T B S}$$

As in the case of the test for homocedasticity in a linear model, for each observed value t_0 of T the p-value of this test can be calculated as:

$$p \;=\; P(\; S^T A S - t_0 S^T B S > 0 \;\mid\; H_0 \;\; true) \;=\; P(\; Q(t_0) > 0 \;\mid\; H_0 \;\; true)$$

where the distribution of the quadratic form $Q(t_0) = S^T A S - t_0 S^T B S == S^T(A - t_0 B)S$ is approximated by a shifted χ^2 distribution as in the test statistic for homoscedasticity in a linear model. As pointed out in that occasion, the moments of the shifted χ^2 distribution can be matched to the moments of the quadratic form $Q(t_0)$. The moments of $Q(t_0)$ are available from the traces of powers of the product $(A - t_0 B)\Sigma_S$ where Σ_S is the variance-covariance matrix of the vector S.

It is to be noticed that in this approach the variance-covariance matrix Σ_S can be singular. Even though, the approximation used in the case of the test for the linear model, by matching the moments of this quadratic form in approximately normal variables and a random variable V=a+bU(c) where a and b are constants and U(c) is a χ^2-random variable with c degrees of

freedom, is known to be still valid (see, for instance, Mathai & Provost (1992)).

4.3.1 Adapting the test statistic for a process defined on a regular grid

Let the sampled locations be the set of points $\{\boldsymbol{s}_i : i = 1, \ldots, l\}$ and suppose they are the vertices of a regular grid. In this case, the entries $\sigma_{ij} = 2\sigma^2 - 2\gamma(h_{ij})$, $i, j = 1, \ldots, l$ for the same value of h_{ij} are all equal. In this case the size of the variance-covariance matrix of S (necessary to calculate the aproximate distribution of the test statistic) could be reduced considerably.

In fact, the number of distinct distances between two points on a grid of this type is considerably less than the total number of distances between them. This reduction is also substantial in the calculation of the variance-covariance matrix of the variables S in the test statistic. For example, for a 10x10 grid, this matrix has size 10000x10000. If only different distances between points in the grid are considered, the corresponding matrix would have a size of 50x50.

Suppose h_i is a distance between two points on the grid and n_i the number of times that this distance is repeated $1 \leq i \leq k$. Then there exists n_i pairs of points $(\boldsymbol{s}, \boldsymbol{t})$ on the grid such that the variable $R_i = R(h_i)$ as defined above has n_i observed values, say, $R_i = r(\boldsymbol{s}) - r(\boldsymbol{t})$ for n_i pairs of points $(\boldsymbol{s}, \boldsymbol{s})$ on the grid such that $\boldsymbol{s} \in D, \boldsymbol{t} \in D$ and $\| \boldsymbol{s} - \boldsymbol{t} \| = h_i$

Suppose there are $\bar{N}$ different distances between the points on the sampled grid. Then, if a permutation is applied to the components of the vector:

$(h_1, h_2, \ldots, h_N)$ such that the order of the components is not decreasing, the resultant vector has the first n_1 equal components, the following n_2 equal components and so on. Perhaps with abuse of notation the different values of these distances will be called: $h_1, h_2, \ldots, h_{\bar{N}}$. Suppose now that the same permutation is applied to the vector with components R_i for $i = 1, \ldots, N$, so that the first n_1 values of $R_i = r(\boldsymbol{s}) - r(\boldsymbol{t})$ correspond to points $\boldsymbol{s},\boldsymbol{t} \in D$ such that $\| \boldsymbol{s} - \boldsymbol{t} \| = h_1$, the following n_2 values of $R_i = r(\boldsymbol{s}) - r(\boldsymbol{t})$ correspond to points $\boldsymbol{s},\boldsymbol{t} \in D$ such that $\| \boldsymbol{s} - \boldsymbol{t} \| = h_2$, and so on. In order to distinguish the values R_i as coming from a difference of residuals associated with points at equal or different distances, a new subscript will be introduced. Then:

$$R_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, \bar{N}$$

The new variables $\bar{S}_i$ and $\tilde{\bar{S}}_i$ corresponding to S_i and $\tilde{S}_i$, respectively, are defined as:

$$\bar{S}_i = \frac{1}{n_i} \sum_{j=n_{i-1}+1}^{n_i} \left[|R_{ij}|^{1/2} - E_0(|R_{ij}|^{1/2}) \right], \quad i = 1, \ldots, \bar{N} \tag{4.7}$$

and

$$\tilde{\bar{S}}_i = \sum_{j=1}^{\bar{N}} w_{ij} \bar{S}_j \text{ where } w_{ij} = \frac{\exp - \left(\frac{h_i - h_j}{b} \right)^2}{\sum_{j=1}^{\bar{N}} w_{ij}}, \; i = 1, \ldots, \bar{N}$$

where b is the bandwidth. With this notation the new test statistic is:

$$T = \frac{\sum_{i=1}^{\bar{N}} n_i (\bar{S}_i - \bar{\bar{S}})^2 - \sum_{i=1}^{\bar{N}} n_i (\tilde{\bar{S}}_i - \bar{\bar{S}})^2}{\sum_{i=1}^{\bar{N}} n_i (\tilde{\bar{S}}_i - \bar{\bar{S}})^2}$$

where $\bar{\bar{S}} = \frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} \bar{S}_i$. The variables $\bar{S}_i$, $i = 1, \ldots, \bar{N}$ are also approximately

normal with zero-mean and variance-covariance matrix with entries:

$$cov(\bar{S}_i, \bar{S}_j) = \frac{1}{n_i n_j} \sum_{l=1}^{n_i} \sum_{u=1}^{n_j} \left[E_0(|R_{il}|^{1/2}|R_{ju}|^{1/2}) - E_0(|R_{il}|^{1/2}) E_0(|R_{ju}|^{1/2}) \right]$$

for $i, j = 1, \ldots, \bar{N}$.

The expected values $E_0(|R_{il}|^{1/2}|R_{ju}|^{1/2})$ can be calculated by numerical integration by using the bivariate joint density distribution of (R_{il}, R_{ju}), $l = 1, \ldots, n_i$; $u = 1, \ldots, n_j$; $i, j = 1, \ldots, \bar{N}$ as explained above. More generally, when the correlations induced by the correlation of residuals are ignored, the covariance of $\bar{S}_i$ and $\bar{S}_j$ can be calculated exactly using the hypergeometric function, as reported by Cressie (1993), page 76.

Even though the number of calculations of $E_0(|R_{il}|^{1/2}|R_{ju}|^{1/2})$ is still quite large, the advantage of this procedure for points on a regular grid is that these values do not need to be stored. In fact, only the values of $cov(\bar{S}_i, \bar{S}_j)$, $i, j = 1, \ldots, \bar{N}$ are necessary to obtain the p-value for the test.

Under the null hypothesis, as in the case for the test for homoscedasticity for the linear model, the residuals can be considered approximately independent without loss of accuracy. However, in this case the variance-covariance Σ_S necessary to calculate the p-values for the test is still not diagonal in the general approach of the random process with a linear trend depending on the spatial locations. Despite this, the number of numerical calculations can be reduced considerably. Details of this approach are given now.

4.3.2 Independent residuals

The approximate by independent residuals were found to be a very good approximation in the context of the test statistic for homoscedasticity in the linear model. In this case the advantage of this approximation was the big increasing in the efficiency of calculations to obtain a p-value for the test. In fact, under the assumption of independent residuals, the variables used to calculate the final quadratic form were also independent given that each of those variables were a function of a single residual. This independence produces a diagonal variance-covariance matrix of these variables whose diagonal entries can be calcualted straightforwardly. In the approach presented here for testing constant variaogram, improving the efficiency of the calculations is even more important because the nature of the problem requires a larger number of calculations and space of storage in its implementation in the computer.

In this case, the variables for the final quadratic form are: $S_i = |r(\boldsymbol{s}+\boldsymbol{h}_i) - r(\boldsymbol{s})|^{1/2} = |R_i|^{1/2}$. Then, each variable S_i, used here is not a function of each residual but a function of the difference of two residuals and, consequently, independent residuals do not imply independence of these variables. Nevertheless, these covariances achieve simpler expressions than in the general case and again the calculations of the entries of the variance-covariance matrix can be done straightforwardly when some assumptions are added to those for the original model.

As explained before, if $i \neq j$ and $1 \leq i \leq \bar{N}$ and $1 \leq j \leq \bar{N}$ the covariance between the differences of residuals R_i, R_j can be written as: $cov(R_i, R_j) = cov(r(\boldsymbol{s}+\boldsymbol{h}_l), r(\boldsymbol{t}+\boldsymbol{h}_k)) - cov(r(\boldsymbol{s}), r(\boldsymbol{t}+\boldsymbol{h}_k)) - cov(r(\boldsymbol{s}+\boldsymbol{h}_l), r(\boldsymbol{t})) + cov(r(\boldsymbol{s}), r(\boldsymbol{t}))$

for some $\boldsymbol{s}$, $\boldsymbol{t} \in D$ and $\boldsymbol{h}_l$, $\boldsymbol{h}_k$ vectors between two points in D.

Then,

$$cov(R_i, R_j) = \begin{cases} \sigma^2_{r_u} \quad if \quad \begin{cases} \boldsymbol{s} + \boldsymbol{h}_l = \boldsymbol{t} + \boldsymbol{h}_k \quad (\sigma^2_{r_u} = var(r(\boldsymbol{s} + \boldsymbol{h}_l))) \\ or \\ \boldsymbol{s} = \boldsymbol{t} \qquad\qquad\quad (\sigma^2_{r_u} = var(r(\boldsymbol{s})) \end{cases} \\ -\sigma^2_{r_u} \quad if \quad \begin{cases} \boldsymbol{s} + \boldsymbol{h}_l = \boldsymbol{t} \quad (\sigma^2_{r_u} = var(r(\boldsymbol{t})) \\ or \\ \boldsymbol{t} + \boldsymbol{h}_k = \boldsymbol{s} \quad (\sigma^2_{r_u} = var(r(\boldsymbol{s})) \end{cases} \\ 0 \qquad otherwise \end{cases}$$

for $1 \leq i, j \leq \bar{N}$.

$$var(R_i) = var(r(\boldsymbol{s} + \boldsymbol{h}_l)) + var(r(\boldsymbol{s})) = \sigma^2_{r_l} + \sigma^2_{r_k} \quad 1 \leq i \leq \bar{N}$$

Under the null hypothesis, the variances of residuals $\sigma^2_{r_l}$ are diagonal elements of the projection matrix $I - X(X^TX)^{-1}$ multiplied by the constant variance of the process σ^2_0 which can be assumed equal one because its value is cancelled in the expression for the test statistic. It is straightforward to proof that the $cov(S_i, S_j) = cov(|R_i|^{1/2}, |R_j|^{1/2})$ has the same value when the $cov(R_i, R_j)$ has a different sign. The variances $var(S_i) = var(|R_i|^{1/2}) = \frac{\sqrt{2\sigma^2_{R_i}}}{\pi}(\sqrt{\pi} - \Gamma^2(3/4)) = \frac{\sqrt{2(\sigma^2_{R_l} + \sigma^2_{R_k})}}{\pi}(\sqrt{\pi} - \Gamma^2(3/4))$ as explained in the chapter for the linear model. Then, if the sample size is n, the number of numerical calculations needed for the variance-covariance matrix Σ_S is only $n^2 - n$ and the test is very efficient from the point of view of its implementation in a

computer.

An interesting case in which the assumption of independent residuals is very useful is the case in which the process is assumed to be stationary in mean. Suppose that the process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, observed at the points $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\} \subseteq D$ is stationary, isotropic and has a normal distribution with zero-mean. In this case the differences of residuals R_i are the differences of the observed values, say, $R_i = Y_l - Y_k$, for some l and k, $1 \leq l, k \leq N$ and $1 \leq i \leq N$. Hence, under the null hypothesis of constant variogram or independence, the covariances $cov(R_i, R_j)$, $1 \leq i, j \leq N$, have values as explained before but where now the values $\sigma^2_{r_u}$ are all equal to the constant variance σ^2_0 and the variance $var(R_i)$ are also all constant and equal to 2 σ^2_0. Then, the joint distribution of each pair (R_i, R_j) is normal bivariate zero mean and variance covariance matrix $\Sigma_{R_i R_j}$ with entries $\sigma_{11} = \sigma_{22} = 2\sigma^2_0$ and $\sigma_{12} = \sigma^2_0$ or $\sigma_{12} = -\sigma^2_0$ or $\sigma_{12} = 0$. Then, the variance-covariance matrix of the vector $(S_1, S_2, \ldots, S_N)$ has all the elements on the diagonal equal to $var(S_i) = \frac{\sqrt{2.2}}{\pi}(\sqrt{\pi} - \Gamma(3/4)^2) = 0.1724015$ and all the elements out of the diagonal equal to zero or to $E_0(S_i S_j) - E_0(S_i)E_0(S_j) = E_0(S_i S_j) - E^2_0(S_i) = \sigma_0(c - \left(\frac{\sqrt{\sqrt{2}.\sqrt{2}}}{\pi}\Gamma(3/4)\right)^2) = \sigma_0(c - 0.9559776)$ where c is the common value of $E_0(S_i S_j)$, $1 \leq i, j \leq \bar{N}$ and can be calculated by numerical integration. This value , as obtained by using Nag subroutines is found to be approximatly c=0.9868. Then, $cov(S_i, S_j) = \sigma_0 0.031$.

4.4 Simulation results

The size and power of the test were analyzed through a simulation study. In order to do this procedure more efficient from the computational point of view and without any loss of generality, the linear model considered was:

$$Y(\boldsymbol{s}_i) = \beta_0 + \varepsilon(\boldsymbol{s}_i)$$

where the points $\boldsymbol{s}_i$, $i = 1, \ldots, n$ were points on a regular grid in $[0, 1)^2$ and where $\varepsilon(\boldsymbol{s}_i)$, $i = 1, \ldots, n$ are gaussian random variables with zero-mean and covariance function of the form $c(h) = \exp(-ch)$ and $c(h) = \exp(-ch^2)$, for each distance h between two points on the grid. The "degree of dependence" was controlled considering different values for the constant c. In fact, the case $c = 0$ corresponds to the null hypothesis of constant variogram or independence as explaned before. The shapes of the semivariogram with these covariance functions for the different values of c are shown in Figure 4.1.

For $c = 2$ in the exponential model, the process presents quite a long-range dependence compared to the other extreme considered case of $c = 50$ where, as displayed in Figure 4.1, two variables in the process at a distance equal or greater than 0.2 are "almost" independent.

In each case 1000 samples were generated by using the simulation method proposed by Wood and Chan (1995) and the subroutine by Chan and Wood (to appear) under the alternative hypothesis. Different approaches for the distribution of the test statistic were also considered. As shown in tables 4.1, 4.2 and 4.3. One of the options was the χ^2 approximation for the distribution

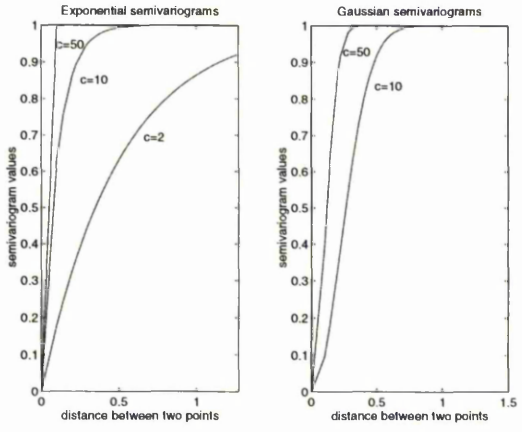

Figure 4.1. Display of the different models for the gaussian semivariogram used in the simulations. On the left, the exponential model: $\gamma(h) = \sigma^2 - \exp(-ch)$. On teh right, the gaussian model: $\gamma(h) = \sigma^2 - \exp(-ch^2)$ The corresponding curves for different values of c and a variance $\sigma^2 = 1$ are shown in this figure.

of the test statistic, another the ordinary bootstrap and the third one was the resampling ideas by using permutations of one random sample to generate others.

In the permutation test, the observed responses are permuted around the spatial locations. Under the null hypothesis of no spatial correlation this permutation will not affect the distribution of the test statistic. An empirical p-value for the test can, therefore, be calculated by counting the number of statistics from permuted data which are larger than the observed statistic.

In the bootstrap test, the mechanism is very similar. The difference is that at each location a response from the original set is sampled with replacement from the entire collection of responses.

Both of these techniques are described in grater detail in chapter 5.

The simulation was also performed for different grid sizes and different bandwidths. From the simulation studies, some of whose results are shown in Table 4.1 and Table 4.2 and 4.3 is possible to conclude:

1. In all cases, the size of the test, indicated by the results for $c = 0$, is close to the target value of 5%. In addition, there is very little difference between the performance of the bootstrap, permutation test and the approximation of the distribution of the test statistic by χ^2 distribution.

2. Under the general assumption of a general linear model, the calculation of the variance-covariance matrix of the random variables in the quadratic form of the test statistic is highly time-computer comsuming because, even for a small number of spatial points, the number of all the possible distances is usually big. Nevertheless, under the null hypothesis, the approximation of this matrix to that corresponding to independent residuals is very good in the case considered in chapter 3. This possibility makes the test for constant variogram a useful tool even in cases where the number of different distances is big.

3. Despite the large computing effort when the procedure of resampling is based on the idea of permutation of one realization of the process, this method has the advantage of generality. The size and power of the test calculated with this approach are reasonably similar to those calculated with the χ^2-approximation.

4. The most relevant parameters related to the power of the test are the grid size n and the "strenght of dependence" of the underlying process. This degree of dependence is, indeed, measured by the value of the

constant c in the expression for the gaussian variogram used in this simulation study.

5. The value of the smoothing parameter does have some effect on the power and size of the test. However, in contrast to the case of the test for constant variance in a linear model, the more common and applicable variogram models are naturally not-decreasing monotonic functions of the distances between points.

		Exponential variogram: $\gamma(h) = 1 - \exp(-ch)$					
		$c = 50$		$c = 10$		$c = 2$	
		χ^2-approx.	bootstrap	χ^2-approx.	bootstrap	χ^2-approx.	bootstrap
n	h	%	%	%	%	%	%
5	0.15	4.00	4.50	7.70	8.60	48.30	45.00
	0.25	5.20	5.80	8.40	9.40	49.10	43.20
	0.35	5.60	6.50	8.20	10.20	50.80	47.50
7	0.15	6.50	8.40	19.80	23.50	84.60	83.00
	0.25	6.80	7.00	16.40	20.10	77.00	70.20
	0.35	5.00	6.50	14.70	18.50	71.40	72.00
10	0.15	19.80	25.30	44.50	46.00	97.80	95.00
	0.25	16.40	15.00	35.50	38.70	91.00	90.80
	0.35	14.70	13.50	30.00	35.00	85.30	84.20

Table 4.1. Size and power of the test for constant variogram as calculated from the χ^2 and the bootstrap approximations for constant variogram for a gaussian isotropic stationary process. The data were generated on a $n \times n$ regular grid in $[0, 1]^2$. Different grid sizes n and bandwidths h were also considered. The data were generated under an exponential model for the variogram.

		Gaussian variogram: $\gamma(h) = 1 - \exp(-ch^2)$					
		$c = 0$		$c = 50$		$c = 10$	
		χ^2-approx.	boots-trap	χ^2-approx.	boots-trap	χ^2-approx.	boots-trap
n	h	%	%	%	%	%	%
5	0.15	3.80	4.50	5.80	6.90	55.70	52.00
	0.25	4.30	5.00	8.30	8.50	46.70	53.50
	0.35	4.50	4.80	8.00	10.50	44.70	41.00
7	0.15	5.20	5.10	27.50	32.00	94.00	91.00
	0.25	5.10	4.80	17.20	25.00	77.60	87.40
	0.35	5.80	5.40	14.80	15.60	69.40	65.00
14	0.15	7.90	4.50	78.00	79.80	100.00	98.00
	0.25	7.10	5.60	44.80	51.20	95.00	90.00
	0.35	7.20	6.50	36.40	35.00	86.20	87.50

Table 4.2. Size and power of the test for constant variogram as calculated from the χ^2 and the bootstrap approximations for constant variogram for a gaussian isotropic stationary process. The data were generated on a $n \times n$ regular grid in $[0, 1]^2$. Different grid sizes and bandwidths were also considered. The data were generated under a gaussian model for the variogram.

4.5 Examples

In order to illustrate the performance and application of the test statistic for the constant variogram, three examples will be now given. In the first example two simulated sets of data from a stationary and isotropic process on the points on a regular grid in $\mathbb{R}^2$ are considered. One of them assumed a gaussian model for the covariance function of the form $c(h) = exp(-h^2)$ and in the other the covariance function is null, implying independence and consequently constant variogram. The second and third examples are concerned with applying the method proposed here to data coming from the "real world". In the second example the data are the measures of two chemical substances, PO4 and NO2, which are indicators of pollution in an area nearby the Clyde river

		Exponential variogram: $\gamma(h) = 1 - \exp(-ch)$					
		$c = 10$			$c = 2$		
		χ^2-approx.	boots-trap	permut. test	χ^2-approx.	boots-trap	permut. test
n	h	%	%	%	%	%	%
6	0.15	15.50	16.50	17.50 *	81.40	85.00	81.00 *
	0.25	17.80	15.20	16.00 *	72.40	74.50	72.00 *
	0.35	14.50	19.80	15.60 *	65.00	68.30	67.00 *
7	0.15	19.80	23.50	25.00 *	84.60	83.00	85.00 *
	0.25	16.40	20.10	22.00 *	77.00	70.20	78.00 *
	0.35	14.70	18.50	19.00 *	71.40	72.00	72.00 *
10	0.15	44.50	46.00	52.00 *	97.80	95.00	92.00 *
	0.25	35.50	38.70	45.00 *	91.00	90.80	99.00 *
	0.35	30.00	35.00	32.00 *	85.30	84.20	85.00 *

Table 4.3. Power of the test for constant variogram as calculated from the χ^2 and the bootstrap approximations for constant variogram for a gaussian isotropic stationary process. The data were generated on a $n \times n$ regular grid in $[0,1]^2$. Different grid sizes n and bandwidths hw ere also considered. The results in the table for the permutation test version are over 100 simulations. The rest over 1000.

in the U.K. The third example is concerned with data which are residuals of linear models where the response variables are measures of three variables of the water column and surface sediments. In this case the spatial locations are points in some billabongs in Australia.

		Exponential variogram: $\gamma(h) = 1 - \exp(-ch^2)$					
		$c = 10$			$c = 50$		
		χ^2-approx.	boots-trap	permut. test	χ^2-approx.	boots-trap	permut. test
n	h	%	%	%	%	%	%
5	0.15	55.70	50.50	58.00 *	5.80	6.40	5.00 *
	0.25	56.70	55.30	67.00 *	8.30	9.00	15.00 *
	0.35	42.30	41.10	56.00 *	5.60	5.00	3.00 *
7	0.15	94.00	91.00	90.00 *	27.50	20.10	10.00 *
	0.25	77.60	76.50	85.00 *	17.20	15.40	21.00 *
	0.35	69.40	65.80	62.00 *	15.40	12.30	12.00 *

Table 4.4. Power for the three versions of the test for constant variogram for an isotropic stationary process. Using simulated data from the model $y_i = \mu + \varepsilon_i \quad i = 1, \ldots, n$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is normal with semivariogram $\gamma(h) = 1 - \exp(-ch^2)$ for $c = 10$ and $c = 50$. The data were generated on a $n \times n$ regular grid in $[0, 1]^2$ for different values of n. The results in the table for the permutation test version are over 100 simulations. The rest over 1000.

		Constant variogram: $\gamma(h) = 1$		
		χ^2-approx.	boots-trap	permut. test
n	h	%	%	%
6	0.15	5.70	4.80	4.00 *
	0.25	4.50	5.10	5.00 *
	0.35	5.20	4.90	5.50 *
7	0.15	5.20	5.00	4.50 *
	0.25	5.10	4.80	3.70 *
	0.35	5.80	5.40	5.20 *

Table 4.5. Size for the three versions of the test for constant variogram for an isotropic stationary process. Using simulated data from the model $y_i = \mu + \varepsilon_i \quad i = 1, \ldots, n$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is normal with semivariogram $\gamma(h) = 1$. The data were generated on a $n \times n$ regular grid in $[0, 1]^2$ for different values of n. The results in the table for the permutation test version are over 100 simulations. The rest over 1000.

4.5.1 Example 1: Simulated data

Exponential semivariogram

A process considered here was:

$$Y(\boldsymbol{s}_i) = \beta_0 + \varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n$$

where $\boldsymbol{s}_i, \quad i = 1, \ldots, n$ are points on a regular grid in $\mathbb{R}^2$. In this case, the grid was a 10×10. The variables $\varepsilon_i, \quad i = 1, \ldots, n$ are jointly normal, zero-mean with variance-covariance matrix Σ whith entries:

$$\sigma_{ij} = \begin{cases} \sigma_0^2 = 1 & if \quad i = j \\ 1 - \exp(-2 \parallel \boldsymbol{s}_i - \boldsymbol{s}_j \parallel) & if \quad i \neq j \end{cases}$$

for $i, j = 1, \ldots, 10$. The generated data in their corresponding l ocations are displayed on the left of Figure 4.2.

In the top left panel of Figure 4.3, the classical estimator of the semivariogram (in dashdotted line) and the thoretical semivariogram are displayed. The p-value for the test of constant variogram was p=0.0001. In fact, the strenght of dependence given by the covariance function of this generated process was expected to imply a significantly small p-value. In the top right panel of Figure 4.3, the observed values of "raw s", defined as the average of all the squared root of absolute differences $Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)$, for each pair $(\boldsymbol{s}_i, \boldsymbol{s}_j)$ such that $\parallel \boldsymbol{s}_i - \boldsymbol{s}_j \parallel = h$, for each h which is a distance between two points on the grid, and the smooth version, or kernel-weighted average of these differences are shown. The bandwidth of the these kernels functions was 0.25. This

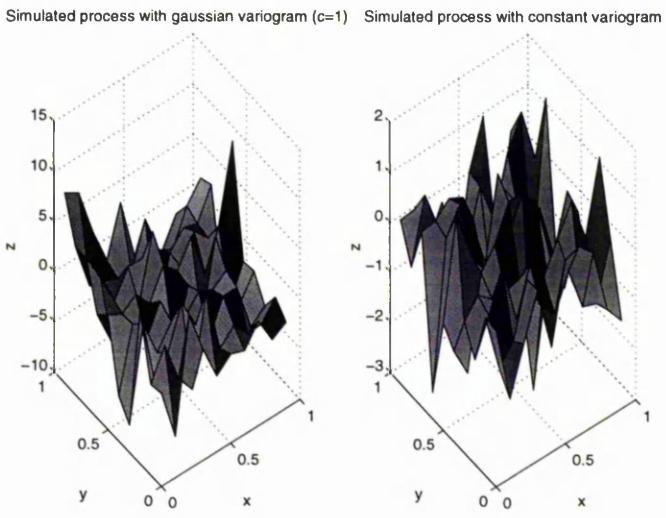

Figure 4.2. Simulated normal processes on a 10×10 regular grid in $[0,1]^2$. On the left panel, the theoretical variogram is $\gamma(h) = 1 - \exp(-2h)$, for each distance h. On the right, it is a constant.

scatter plot also shows the expected increasing tendency of these variables.

4.5.2 Constant semivariogram

As an example where there is no spatial dependence or in terms of the variogram, it is constant, a stationary isotropic gaussian process zero mean and diagonal varaiance-covariance was generated. The grid and bandwidth of the kernel functions considered were as in the previous example. The generated data are displayed in th right panel of Figure 4.3. The p-value for the test of constant variogram was $p = 0.80$. In the left botton panel of Figure 4.3, the estimated semivariogram and the theoretical semivariogram are displayed. On the right the observed values of the variables "raw s" as explained above and the its smooth version are displayed. Both curves illustrate the "constant

behaviour" of the underlying variogram.

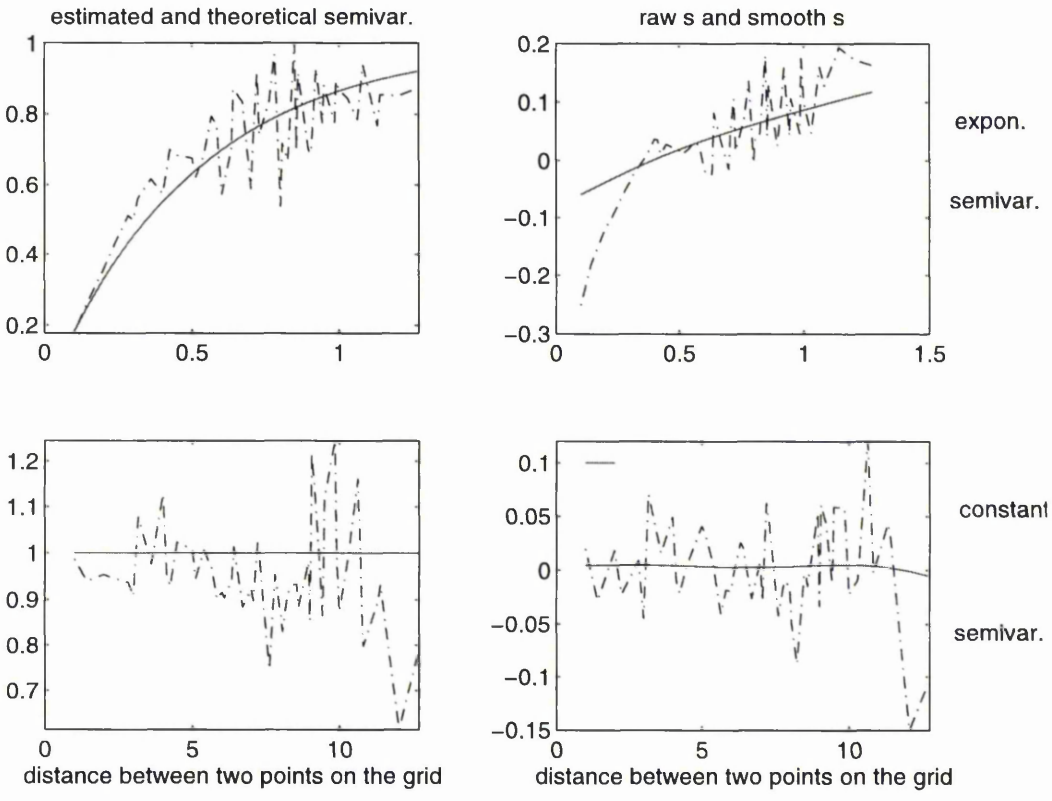

Figure 4.3. On the left, scatter plots of the estimated and theoretical semivariogram of the simulated processes in Figure 4.2 are shown. On the right values of "raw s" and "smooth s" as defined in the text are displayed. The two scatter plots on the top corresponding to the gaussian semivariogram, and those on the bottom to the constant semivariogram.

4.5.3 Example 2: In areas nearby the Clyde River

The Clyde River Purification Board has the responsibility of monitoring pollution levels in the river and sea areas within its sphere of operation. These data shown in this example were collected by the Board vessel based at Greenock

which routinely collects water samples from the positions in the area of the River Clyde and its stuary.

A variety of pollutants and other indicators are assessed on each sample. In this example, two indicators of pollutions are considered, PO4 and SlO2 form a single survey in November 1994. Their observed values are shown in Figure 4.4.

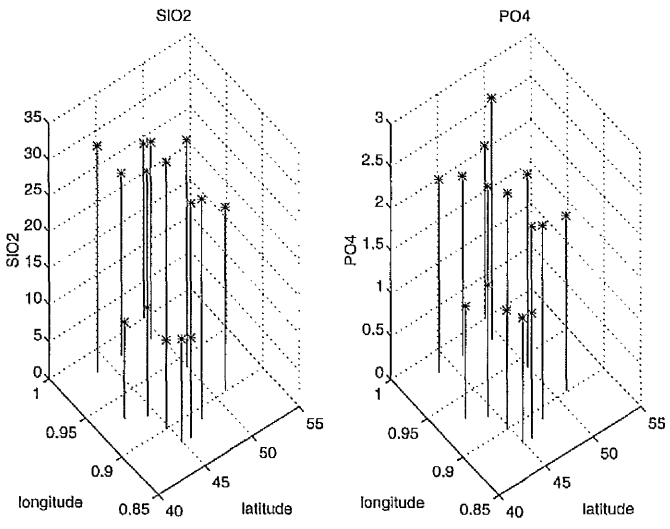

Figure 4.4. This figure shows the values of the two indicators of pollution in the area nearby the Clyde river. On the left, the values of SlO2 and, on the right, the values PO4 are displayed.

One question of interest is whether there is scope for reducing the number of sampling points if the spatial correlation between neighbouring observations is sufficiently large. To answer this question and others concerned with the statistical behaviour of the variables of interest, a preliminary exploratory study of the most elemental parameter involved in the joint distribution of these variables is compulsory to perform. With this aim, the structure of the

variance-covariance patterns of two indicators of pollution has been carried out. These indicators have been assumed to be random gaussian processes with constant mean. The model for each indicator Y at the location $\boldsymbol{s}$, say, $Y(\boldsymbol{s})$ is considered as:

$$Y(\boldsymbol{s}_i) = \mu + \varepsilon(\boldsymbol{s}_i) \quad i = 1, \ldots, n$$

where $\boldsymbol{\varepsilon} = (\varepsilon(\boldsymbol{s}_1, \varepsilon(\boldsymbol{s}_2, \ldots, \varepsilon(\boldsymbol{s}_n))$ is normal zero-mean and variance-covariance matrix Σ whose entries depend on the distances between the pairs of the locations, and n, is the number of locations. In this case $n = 15$. The simplest question to be answered here is whether or not the matrix Σ is diagonal, for each indicator. In other words, whether or not the variogram of the underlying processes are constant.

Because all the locations are on a non-regular grid, the number of different distances is high enough to consider an estimation of the variogram by averaging the square differences not only for each distance but for a set of distances in a "tolerance" (as called by Cressie, (1991)) region. Then,

$$2\hat{\gamma}(h_i) = \{Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j) : (\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(h); h \in T(h_i)\}$$

where $T(h_i) = \{h : \| h - h_i \| \leq h_b\}$ and where the value of h_b considered here was the bandwidth for the kernel weighted variables for the quadratic form of test statistic. In this context the estimated variogram is calculated for graphical purposes and consequently more details are irrelevant in this context. On the left of Figure 4.5 these estimated semivariograms for both indicators, SIO2 and PO4, are displayed.

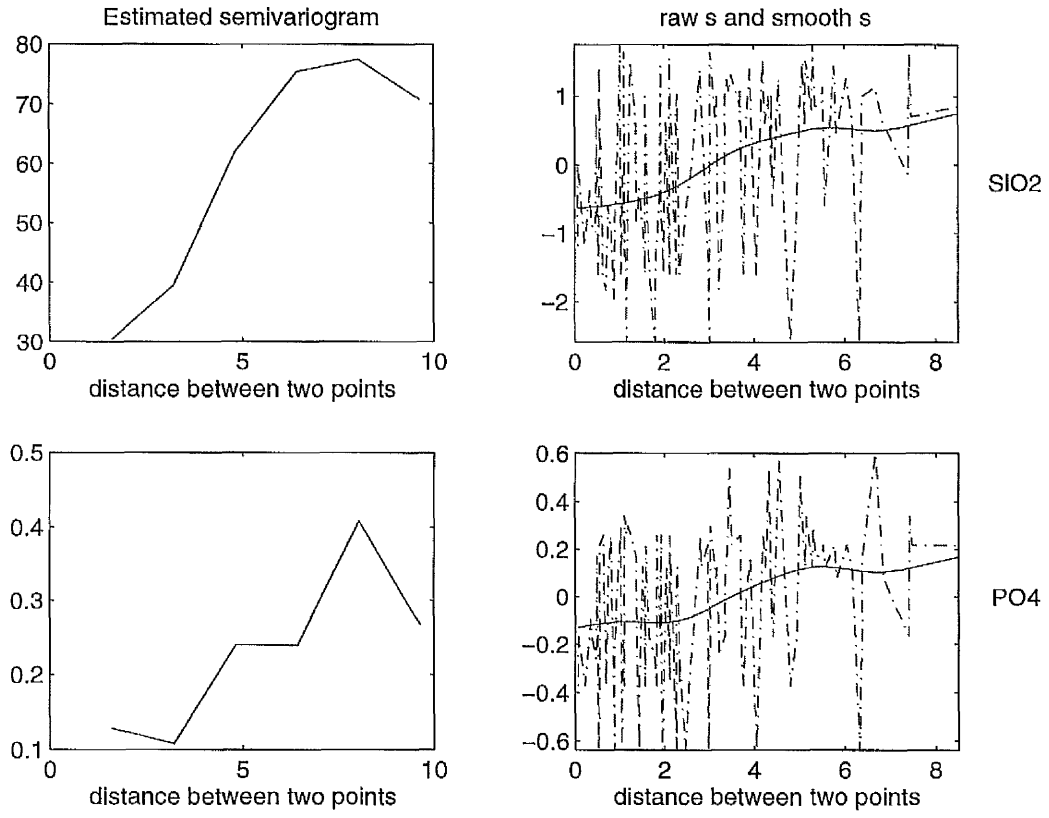

Figure 4.5. On the left, scatter plots of the estimated semivariogram of the data corresponding to the indicators of pollution SlO2 and PO4 are displayed. On the right values of "raw s" and "smooth s" for both indicators are shown. The two scatter plots on the top corresponding to SlO2 and those on the bottom to PO4.

On the right of Figure 4.5, the "raw s" and "smooth s" variables for both indicators, with the same meaning through the approach for testing constant variogram, are shown. In this case, the p-values were, $p = 0.0001$ for SlO2, and $p = 0.04$ for PO4. These p-values were calcualted with a bandwidth, h_b equal to the range of distances over the number of different distances. For other values of bandwidths, $2h_b$ and $3h_b$, the respective p-values are 0.0008 and 0.0015 for SlO2 and 0.048 and 0.05 for PO4.

Hence, there is not enough evidence of independence of the variables at different locations in both cases and this premise should be present in any further study involving these variables.

4.5.4 Example 2: In some billabongs

Billabongs are standing water bodies on the surface of floodplains, and in Australia, as with the rest of the world, they have received little scientific attention in spite of broad-scale modifications resulting from human land-use practices. The natural environment of the study region has been extensively modified by farming practices and the construction of dams, which has altered the flow and flood regimes of the rivers which replenish billabongs.

The purpose of this study is to characterise the regional ecology of eastern Murray Valley billabongs, and see if there are any differences between the ecology of billabongs on farms versus more "pristine" sites in forest, and between billabongs on regulated and unregulated sections of the region's rivers. The study is composed of two parts: an ecological survey of 43 sites, measuring variables of the water column and surface sediments, and an historical survey examining changes through time in the sedimentary history of ten billabongs. The data were kindly supplied by Ralph Ogden and Ross Cunningham of the Australian National University.

The data analyzed here were the residuals of the linear model:

$$Y_i = \beta_j + \varepsilon_i, \quad j = 1, 2; \quad i = 1, \ldots, 43$$

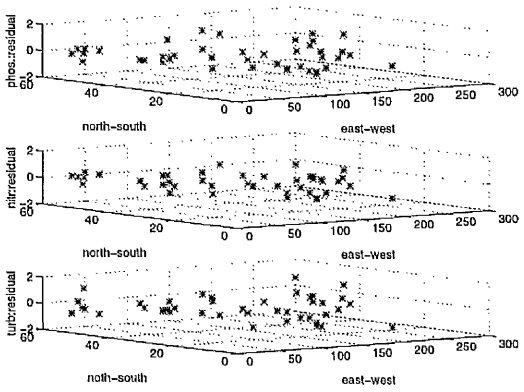

Figure 4.6. Observed values of the processes of residuals in linear models for the of Values of phosphorus, nitrogen and turbidity in some billabongs in Australia, as explained in the text.

where the factors β_j correspond to low ($j = 1$) and high ($j = 2$) levels of farming in sections of the region's rivers. The response variables were nitrogen and phosphorus. A third variable, turbidity, was also examined, using a model which also contained factors for regulated/unregulated water flow and for different river system.

The residuals $r_i = \hat{Y}_i - Y_i, i = 1, \ldots, n$ of the models corresponding to these three variables are displyed in 4.6.

The motivation to consider the study of the structure of dependence of the process of residuals on their spatial locations is supported by the interest in the geographical characterization of these billabongs.

As in the previous example, the spatial locations here are not on a regular grid. This results in a large number of different distances. Hence, an appropriate estimator of the variogram is again of the form considered in the example

2. These estimated semivariograms are displayed on the left of Figure 4.7. On the right, the scatter plots of "raw s" and "smooth s" are shown for each substance. These graphs illustrate the results obtained for the p-values of the test of constant variogram. In fact they were, $p = 0.90$, 0.87, and 0.91 for phosphorus, nitrogen and turbidity, respectively, showing that there is no evidence of spatial correlation in the data.

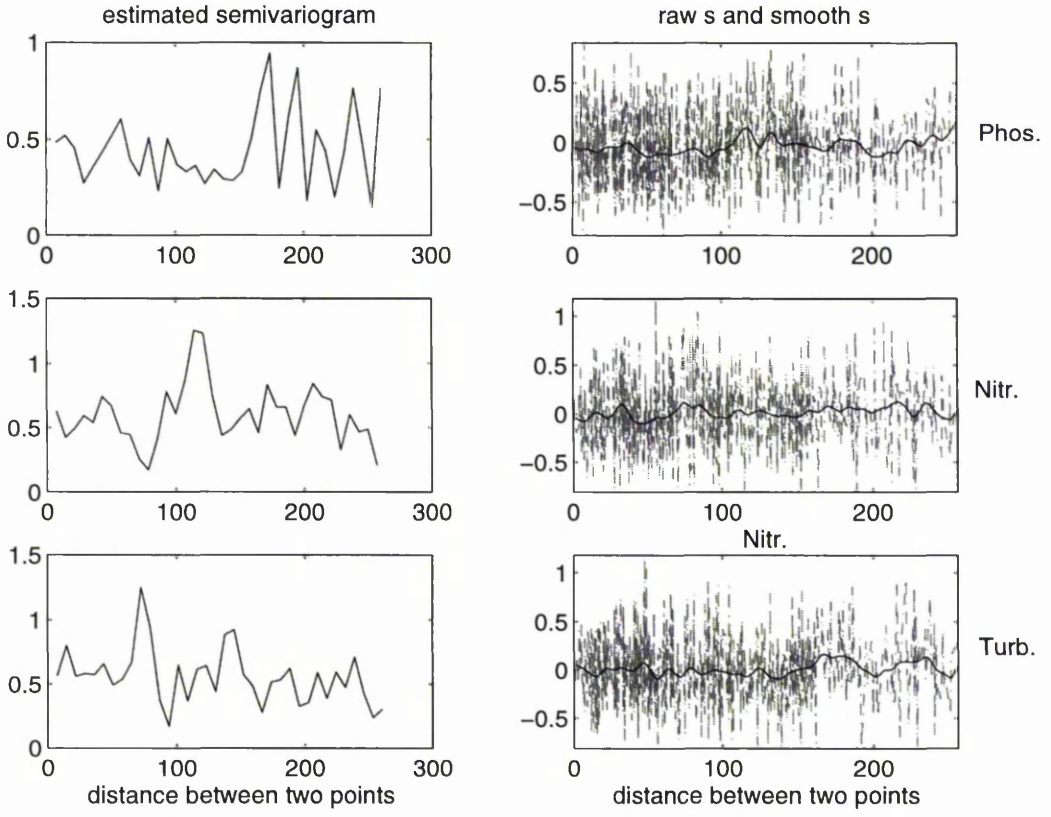

Figure 4.7. On the left, plots of the estimated semivariograms of the residuals in the fitted models for phosphorus, nitrogen and turbidity are displayed. On the right values of "raw s" and "smooth s" for the residuals in the three models are shown.

4.6 Reference bands

As in the approach for checking homoscedasticity in a linear model, reference bands for checking indendence or constant variogram in a graphical approach can be built. The reference bands here are based on the variables $S(h) = |R(h)|^{1/2}$ where R(h) is the difference of residuals for a given value h of the distance between two points on the domain. The idea is based on the estimated smooth function $\tilde{S}$. This function is calculated for each value of h in the interval [a,b] of the distances between all the points in the sample as before. For each value of h the random variable S(h) is, under the null hypothesis, approximately normal. Then, the random variable Q(h)

$$Q(h) = \frac{\tilde{S}(h) - \bar{S}}{\sqrt{var(\tilde{S}(h) - \bar{S})}} \tag{4.8}$$

is approximately standard normal. Then, the $\gamma 100\%$-reference band for the smooth function $\tilde{S}$ can be defined as:

$$Rb_\gamma = \{ \ (h, y) \ | \ a \leq h \leq b; \quad q_1(h) \leq y \leq q_2(h)\} \tag{4.9}$$

where $q_1(h) = -q_0\sqrt{var(\tilde{S}(h) - \bar{S})}$, $q_2(h) = q_0\sqrt{var(\tilde{S}(h) - \bar{S})}$, where q_0 is the $(1+\gamma)/2$ - quantile of the standard normal distribution. The expression $var(\tilde{S}(h) - \bar{S})$ can be calculated as:

$$var(\tilde{S}(h) - \bar{S}) = var(\sum_{i=1}^{N}(w_i(h) - (1/N))S_i =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}(w_i(h) - (1/N))(w_j(h) - (1/N))cov(S_i, S_j)$$

where $w_l(h)$ is the kernel function corresponding to the weight for S_l and calculated at the point h, $1 \leq l \leq N$.

Under the null hypothesis it is expected that each point h, where $a \leq h \leq b$, the variable y= $\tilde{S}(h) - \bar{S}$, should belong to the Rb_γ with a probability of $\gamma 100\%$. This methodology can be used as an approximate graphical tool for checking constant variogram (independence).

The reference bands and their ability to show whether, in a graphical and consequently approximate way, there is or is not evidence of spatial dependence or correlation (under the hypothesis of normality) are now calculated for the examples considered above. In Figure 4.7 the 95% reference bands for the two simulated examples are shown. As in the conclusions drawn from the p-values for the test statitstic in both cases, the 95% reference band of the simulated process with gaussian variogram (top panel) does not contain the smooth curve, as expected, whereas the 95% reference band of the simulated process with constant variogram (bottom panel) includes this curve.

Reference bands for the sample of the indicators of pollution PO4 and SlO2 are displayed in Figure 4.9 where there is some evidence of spatial correlation. In fact, in both cases the p-value to test independence of the underlying processes are less than 0.05.

The third example of reference bands corresponds to the example concerned with data from some billabongs in Australia. In all the three models studied above, the p-values for the test of spatial dependence were reasonably big. This characteristic is once agin coherent with the graphical approximation of the reference bands. Figure 10 illustrate the 95%-reference band for each

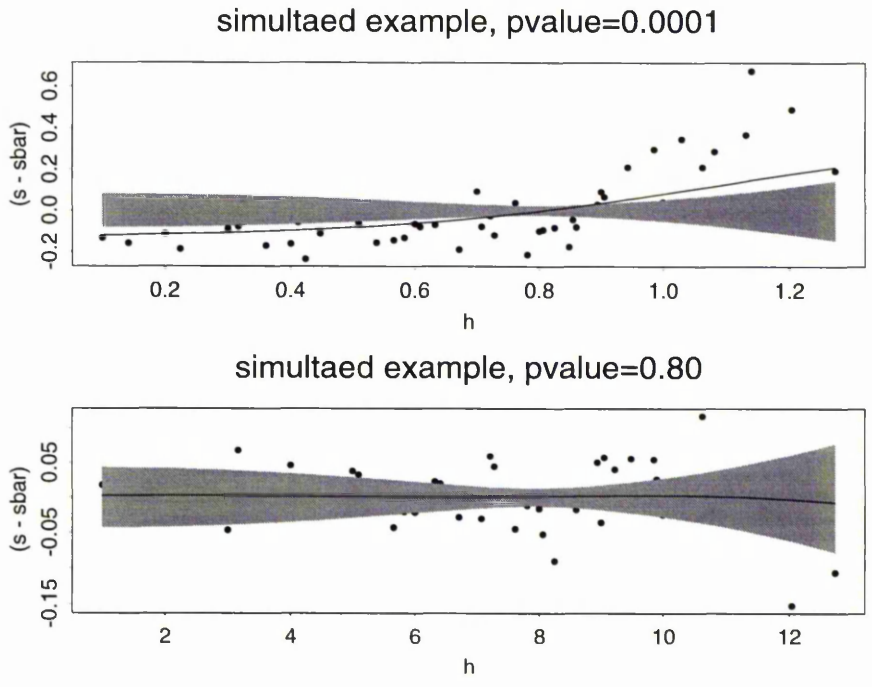

Figure 4.8. 95%-reference bands for the simulated processes in the example 1 are shown. In the top, the semivariogram is exponential with $c = 2$, in the botton is constant.

case.

4.7 Comments

In parametric modelling the estimation of the parameters of the underlying distribution is essential. The translation of this sentence to the language of spatial statistics in its simplest approach is the estimation of the mean and variogram or covariance-function of the underlying process. In this chapter an attempt to develop a tool to check the most basic property of the variogram

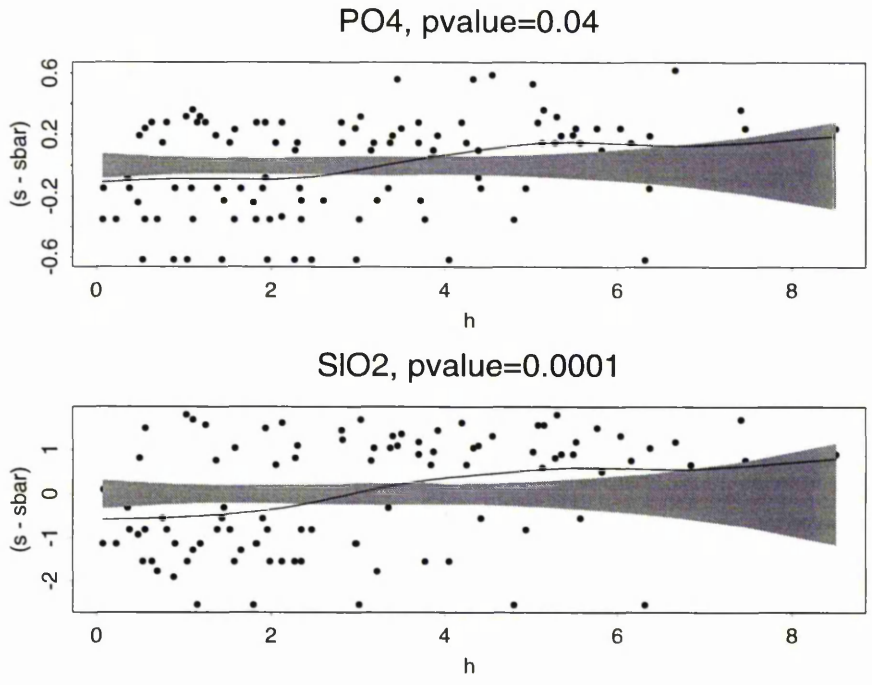

Figure 4.9. The 95%-reference bands for two indicators of pollutions as taken from the area nearby the Clyde river, U.K. are shown in this figure. On the left, the reference band corresponds to PO4, and on the right the reference band corresponds to SIO2 are displayed.

has been made. Some general comments can be summarized as follows:

1. The reference bands can be considered as an useful and adequate graphical tool to search for the behaviour of the spatial dependence.

2. The test proposed here for the null hypothesis of constant variogram can be an appropriate tool to resort to, under the assumptions of second order stationarity and isotropy.

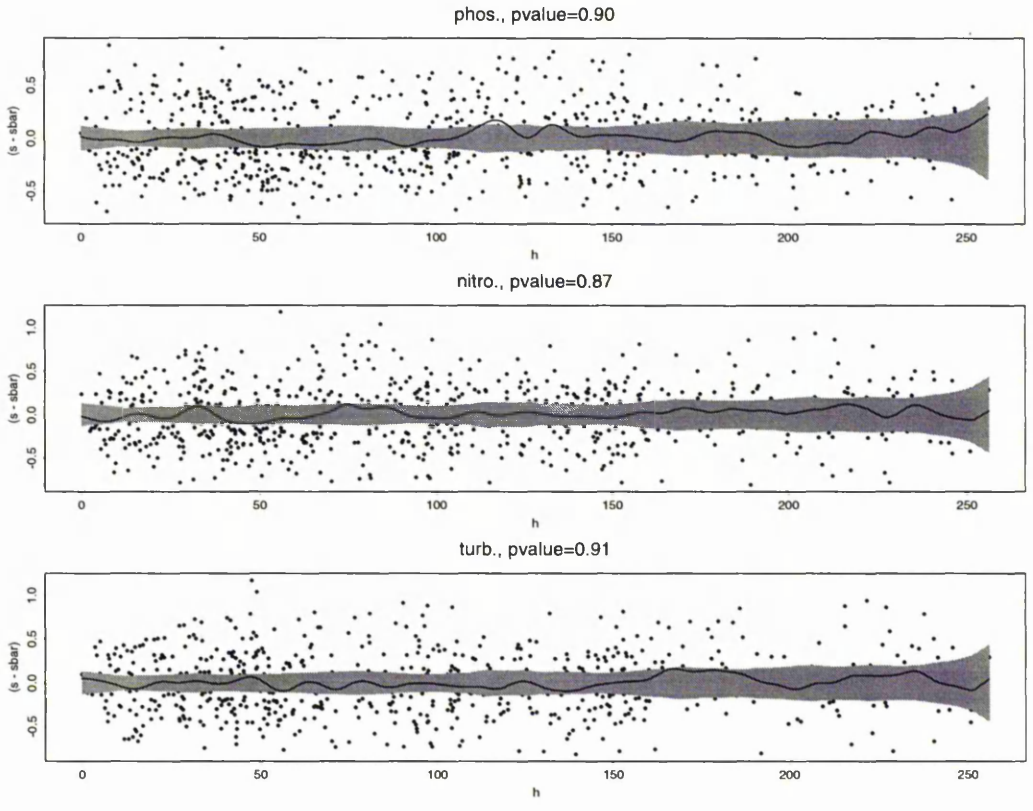

Figure 4.10. The 95%-reference bands for the process of residuals taken form the linear model as explained in the text, for phosp., nitr. and turb. for the example on the billabongs, Australia. In all of these cases the corresponding underlying processes show a structure of independence from their locations. These three graphs remark again the usfulness of this graphical tool.

3. Under the assumption that the mean of the underying spatial process is a linear function of the locations, the χ^2 approximation for the distribution of the test statistic may be computationally not very efficient. In this case, the empirical distribution of the test statistic under the null hypothesis is more efficient.

4. Under the assumption of constant mean, the approximation of the distribution of the test statistic to a χ^2 distribution is straightforward.

5. If the set of spatial locations is a regular grid, or there are many equal distances between these points, the version of averaging the variables considered here is an appropriate approach either in the version for the test statistic or for the graphical approach of the reference bands.

6. The ideas of the reference bands under the null hypothesis of constant variogram might be extended to build reference bands under the null hypothesis of a specific model or smooth curve for the variogram. These ideas have been considered in the context of noparametric checking of models by Bowman & Young (1995).

Chapter 5

Simulating and re-sampling methods for spatial data

5.1 Introduction

The generation of a spatial process with a specific model for the variogram is a tool of indisputable necessity in the context of spatial statistics. In fact, a good algorithm to simulate realizations coming from a model with specific characteristics encourages the development of tools for inference such as test statistics, estimators for parameters of a spatial process, etc. An example of this requirement is the test statistic proposed in the previous chapter for testing for a constant variogram.

The simulation of a process with a specific model for its variance-covariance structure sometimes leads to numerical difficulties at the computational stage.

It usually happens, for instance, that the variance-covariance matrix is positive definite but some eigenvalues are negative when they are calculated in practice, due to numerical problems.

Even when a realization of a process with a given variance-covariance matrix can be obtained by using one of the algorithms proposed in the literature, the generated result may not be a good approximation for the problem which originated its simulation. In fact, many of these algorithms are based on approximations and the accuracy of the approximation should be analysed in each context.

Some of the algorithms proposed for simulation of a process require such a large amount of space to store matrices that they are not very useful in practice. Even when all of the disvantages explained above can be sorted out, the computer time required may be extremely high. In the context of spatial data, as in time series, the necessity of an algorithm to generate a realization of a process efficiently is indisputable.

These are among the reasons which have made some researchers think and write about different ways to generate this kind of process. Most of these methods combine different tools, but they are usually constructed by focusing on one particular characteristic. With this criterion in view, the work dealing with algorithms to generate a process can be classified into three categories: methods which are based on a factorisation of the variance-covariance matrix (similarity or LU factorisation), methods concerned with a projection onto a space of lesser dimensions (the turning band method), and methods based on properties of circulant matrices (embedding the covariance matrix in a circulant matrix) dealing with spectral approaches. Several modifications in

order to improve or/and to extend their respective scopes have been made in all cases.

The first approach has been widely used because of its generality and simplicity. In fact, it is based on the property of positive definiteness of the variance-covariance matrix and does not require assumptions about stationarity. Davis (1986) is among the people who have worked on it.

The idea of simulating a process in several dimensions through its projection on one dimension seems to be due to Matheron (1973). Journel (1974) has worked with turning bands and their applications on mining in his doctoral thesis. He also worked with Huijbregts (1981) on a posterior analysis of the properties and other details in this approach. Afterwards, Brooker (1985), Luster (1985), Mantoglue (1987) and Christakos (1987) deserve to be mentioned for their contributions in this direction. The turning bands algorithm has enjoyed great popularity in the mining environment.

Under the assumption of second order stationarity, another characteristic of a process generated on a regular grid which can be exploited for simulation is the Teoplitz (in one dimension) or block Teoplitz (more than one) structure of the variance-covariance matrix of the process. This leads to tools available in spectral theory and its humble requirement of computer memory has attracted several investigators to adopt it. Some of them are Borgman, Taheri, and Hagan (1984), Davis, Hagan and Borgman (1981), Mejia and Rodriguez Iturbe (1974), Shinozuka and Jan (1972) and Wood and Chan (1994).

Some of these approaches are explained in section 5.2.

A spatially correlated Gaussian process, simulated from an independent Gaussian process by using two-dimensional kernel averaging is proposed. In this, the covariance function is not specified in advance but the strength of dependence can be controlled through the kernel functions. It is the two dimensional analogue of the concept of a moving average process for time series. This method is outlined in section 5.3.

Re-sampling methods for the analysis of a single set of observed data are also discussed. They are the non-parametric bootstrap approach and the permutation technique. These approaches are very attractive because it is not necessary to make assumptions about the distribution or specific structure for the variance-covariance matrix. Both techniques are considered here as useful tools under the assumption of independence. An example in which this assumption is valid is that for the distribution of the statistic proposed in the previous chapter under the null hypothesis. Some general ideas are given in section 5.4.

It is known that the ordinary bootstrap is not appropriate under an assumption of dependence of observations from the underlying distribution. In an effort to amend the bootstrap to apply to dependent data, some modifications have been proposed. They have led to the method which is called the block (or moving) bootstrap. To study the behaviour of this statistic and build local confidence intervals the technique of block bootstrap is used. The coverage of these intervals is then analysed. This approach, as well as a small simulation study, is presented in section 5.5.

A general discussion on techniques to simulate a process is summarised in section 5.6.

Some more general comments are considered in section 5.7.

5.2 Simulation methods

When a process is simulated to check properties of its variogram, as is the case in the study in the previous chapter, the nonnegative definiteness of its variance-covariance matrix can be guaranteed under some circumstances. To analyse this property, the two general situations of constant and non-constant variogram are now considered.

Under the assumption of constant variance,

$$2\gamma(\boldsymbol{h}) = 2var(Y(\boldsymbol{s})) - 2cov(Y(\boldsymbol{s}+\boldsymbol{h}), Y(\boldsymbol{s})) = 2\sigma^2 - 2C(\|\boldsymbol{h}\|) = 2\sigma_0^2$$

. Hence, the covariance function is constant and $C(\boldsymbol{h}) = 2\sigma^2 - 2\sigma_0^2 = c = constant$ and the variance-covariance matrix $\Sigma_{\boldsymbol{\varepsilon}}$ has constant diagonal entries σ^2 and constant off the diagonal entries c. Then the determinant of this matrix is:

$$|\Sigma_{\boldsymbol{\varepsilon}}| = (\sigma^2 - c)^{(n-1)}(\sigma^2 + (n-1)c) = (\sigma^2 - c)^n + nc(\sigma^2 - c)$$

(see Graybill (1983), page 231). Therefore, if $\sigma^2 > c$ then $|\Sigma_{\boldsymbol{\varepsilon}}| > 0$ and $\Sigma_{\boldsymbol{\varepsilon}}$ is positive definite. Clearly, if c=0, $\Sigma_{\boldsymbol{\varepsilon}}$ is diagonal with positive elements on it and, consequently, it is positive definite.

If the variogram is not constant, the non-negative definiteness of the variance-covariance matrix can be guaranteed if the variogram $\gamma(., \boldsymbol{\theta})$ is a member of

a "valid" family:

$$S(\boldsymbol{\theta}) = \{\gamma(.,\boldsymbol{\theta}) : \boldsymbol{\theta} = (c_0, c_s, a_s)', \quad c_0 \geq 0, c_s \geq 0, \quad and \quad a_s > 0\}$$

as, for example, the family known as **spherical**:

$$\gamma(h,\boldsymbol{\theta}) = \left(c_0 + c_s\left((3/2)(h/a_s) - (1/2)(h/a_s)^3\right)\right) I_{(0,a_s]}(h) + (c_0 + c_s) I_{[a_s,\infty]}(h)$$

or the **exponential**:

$$\gamma(h,\boldsymbol{\theta}) = c_0 + c_s\left(1 - exp(-(h/a_s))\right) I_{(0,\infty)}(h)$$

for each $h \geq 0$ which is a distance between two points in the set where the process is defined.

These families, as well as others (see for example, Cressie, 1991, page 60), guarantee the conditional negative definiteness of the variogram. This property can be written as:

$$\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j 2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \leq 0$$

for any finite set of spatial locations $\{\boldsymbol{s}_i : i = 1, \ldots, l\}$ and any finite set of real numbers $\{a_i : i = 1, \ldots, l\}$. Thus, if the variance-covariance matrix of errors is $\Sigma_\epsilon = (\sigma_{ij})$ where $\sigma_{ij} = 2\sigma^2 - 2\gamma(h_{ij})$ and where 2γ is a member of a family as above and h_{ij} is the distance between the points i and j, then the variance-covariance matrix Σ_ϵ is positive definite (Matheron 1971).

Hence, the nonnegative definiteness of the variance-covariance matrix for a process as studied here can be guaranteed at least from a theoretical point of

view. Despite these theoretical considerations, numerical problems can occur in the implementation of a simulation in a computer.

If $\{Y_1, Y_2, \ldots, Y_n\}$ are the observed values of a realization of a stationary process on a regular grid in $\mathbb{R}$, its variance-covariance matrix has the characteristic that all the elements on each super-diagonal are equal and all the elements on each sub-diagonal are equal. In other words, the variance-covariance matrix of a realization of a stationary process in $\mathbb{R}$ is Teoplitz. If the process is defined in $\mathbb{R}^2$, there exists an order of the points on the grid for which its variance-covariance matrix can be split in Teoplitz sub-matrices. In other words, the variance-covariance matrix of a stationary process defined in $\mathbb{R}^2$ is block-Teoplitz. It is widely known that a Teoplitz matrix can be embedded in a symmetric circulant matrix (a $k \times k$-matrix with elements a_{ij} such that if the rest of i+j-2 (module k) is equal to the rest of p+q-2 (module k), then $a_{ij} = a_{pq}$) and a block-Teoplitz matrix can be embedded in a block circulant matrix. The symmetric (block) circulant matrices can be factorized in the complex field by using the fast Fourier transform in a very convenient way in order to simulate the required process. This simulation can be reduced finally to the simulation of independent random variables as in the algorithm which uses the similarity and Cholesky factorisation of the variance-covariance matrix.

The description of some of the algorithms for simulating a process will now be described. For the methods presented in this section the common characteristic will be that the process has a specific variance-covariance matrix. Also the properties of isotropy, normality and zero-mean will be assumptions even though for many of these procedures some or all of these assumptions can be relaxed.

The algorithms will be described for processes in $\mathbb{R}^2$. Nevertheless all the methods in this section can be extended, with more or less computational difficulty, to more than two dimensions.

5.2.1 Using a similarity transformation of the covariance matrix

Let $\{Y(s) : s \in D\}$, $D \subseteq \mathbb{R}^2$ be such that a realization $\{Y_1, Y_2, \ldots, Y_n\}$ at the locations $s_i \in D$, $i = 1, \ldots, n$ is to be simulated. Suppose that a specific $N \times N$ variance-covariance matrix Σ_Y, where $N = n(n-1)/2$ is constructed from a valid family for the variogram. Then, Σ_Y is positive definite and an orthogonal matrix P exists such that: $P^T \Sigma_{\boldsymbol{\varepsilon}} P = \Lambda$ where Λ is the diagonal matrix with the eigenvalues of $\Sigma_{\boldsymbol{\varepsilon}}$ in the diagonal. Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the characteristic roots of $\Sigma_{\boldsymbol{\varepsilon}}$. These values are all positive because $\Sigma_{\boldsymbol{\varepsilon}}$ is a positive definite matrix. Multiplying by P^T in the equation $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, this model is transformed in the model,

$$P^T \boldsymbol{Y} = P^T X \boldsymbol{\beta} + P^T \boldsymbol{\varepsilon} = P^T X \boldsymbol{\beta} + \boldsymbol{\eta}$$

where $\boldsymbol{\eta} = P^T \boldsymbol{\varepsilon}$ has a normal distribution with zero-mean and variance-covariance matrix

$$\Sigma_{\boldsymbol{\eta}} = P^T \Sigma_{\boldsymbol{\varepsilon}} P = \Lambda$$

Therefore $\boldsymbol{\eta}$ is a n-normal vector with components η_i, $i = 1, \ldots, n$ zero-mean independent normal variables with variance λ_i, $i = 1, \ldots, n$. Hence, this algebraic theory can be used in order to generate a spatial process Y with a matrix $\Sigma_{\boldsymbol{\varepsilon}}$ given above. The steps followed for this generation are:

1. Consider a variogram 2γ from a specific valid family and calculate its values $2\gamma(\boldsymbol{h})$, for every vector $\boldsymbol{h}$ which is the difference between two points in D.

2. Calculate the charactistic roots $\lambda_1, \ldots, \lambda_N$ and the matrix P with columns the corresponding characteristic vectors for the matrix $\Sigma_{\boldsymbol{\varepsilon}}$ with entries σ^2 in the diagonal.

3. Generate normal independent random variables $\eta_i \sim N(0, \lambda_i)$, $i = 1, \ldots, n$.

4. Calculate $\boldsymbol{\varepsilon} = P\boldsymbol{\eta}$ which has normal distribution with variance-covariance matrix: $\Sigma_{\boldsymbol{\varepsilon}} = P\Lambda P^T$, and the values of $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

5.2.2 LU triangular decomposition of the covariance matrix

The method by Davis (1986) is based on the idea that if a set of data on a grid are kriged, they are smoother than reality. Then, one way to reproduce the spatial variability is to predict (or krige) one value on each point on the grid, to subtract the predicted value from the original (non-conditional) value in order to obtain a correlated error. These errors are then added to the unconditioned kriged data on the grid to obtain a conditional simulation. These ideas can be summarised as follow:

1. Given a data vector $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)$, a predicted (kriged) value at the point s_0 is given by: $Z^*(s_0) = \boldsymbol{c}' K^{-1} \boldsymbol{Z}$ where $\boldsymbol{c} = (cov(s_0, s_1), \ldots, cov(s_0, s_n))'$ and K is the covariance matrix of the vector Z if simple kriging with zero-mean is assumed (see for instance, Cressie (1991),page 123).

2. Suppose that $\boldsymbol{Z}$ is a vector of data values and $\boldsymbol{Y}$ is a simulated vector at the data locations, then the equation $\boldsymbol{U}^* = C' K^{-1}(\boldsymbol{Z} - \boldsymbol{Y})$, where C is the matrix whose columns are the vectors $\boldsymbol{c}$ as explained previously, can be expressed as $\boldsymbol{U}^* = \boldsymbol{Z}_{cs} - \boldsymbol{Z}_{us}$ with $\boldsymbol{Z}_{cs}$ vector of conditioned simulated values and $\boldsymbol{Z}_{us}$ a vector of unconditioned simulated values.

3. The previous equation can be solved directly and a vector $\boldsymbol{Z}_{cs}$ can be found. To this aim a matrix C given by:

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{11} is the covariance matrix of the data points (K), C_{12} is the covariance matrix between the data points and the grid points, $C_{21} = C_{12}$ and C_{22} is the covariance matrix between the grid points.

4. By using a Cholesky decomposition, the matrix C can be written as

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & O \\ L_{21} & L_{22} \end{bmatrix} + \begin{bmatrix} U_{11} & U_{12} \\ O & U_{22} \end{bmatrix}$$

5. With the notation below a required vector can be written as: $\boldsymbol{Z}_{cs} = L_{21} L_{11}^{-1} \boldsymbol{Z} + L_{22} \boldsymbol{W}$ where $\boldsymbol{Z}$ is the vector of data (previously transformed so that its components are independent standard normal distributed) and so is $\boldsymbol{W}$.

This technique can be applied, among others, to simulate realizations of a process in a part of the grid. It has the disadvantage that its implementation requires matrices of quite big proportions. In order to sort out this problem, the same author has introduced a modification of this methodology (Davis, 1986(b)). He proposed to approximate the square root of the covariance matrix by a minimax matrix polynomial and to use the block Teoplitz structure of the covariance matrix to minimise storage.

5.2.3 Embedding the covariance matrix in a circulant matrix

Wood and Chan (1994) have created a methodology to generate a process $\{Y(\boldsymbol{s}_j) : 1 \leq j \leq n, \boldsymbol{s}_j \in \mathbb{R}^d\}$ with a specific covariance function $C : \mathbb{R}^d \longrightarrow \mathbb{R}$. The assumptions for the covariance function here is stationarity with respect to translation in $\boldsymbol{R}^d$. The assumption of isotropy is not necessary in this approach. The general idea behind this procedure is based on the fact that the variance-covariance matrix of a stationary process defined on a regular grid in $\mathbb{R}^d$ is Teoplitz when $d = 1$, block Teoplitz when $d = 2$ and nested block Teoplitz when $d \geq 3$. This property guarantees that this matrix can be embedded in a symmetric circulant matrix when $d = 1$, block symmetric circulant matrix when $d = 2$, and nested block symmetric circulant matrix when $d \geq 3$. The main idea is to simulate from a longer vector whose covariance matrix is the circulant, and then select the sub-vector whose covariance matrix has the appropriate Teoplitz form. The properties of circulant matrices are used to seek for an adequate factorisation of the covariance matrix. This factorisation

leads to a linear transformation of independent normal variables which produce a process with the required variance-covariance matrix. The description of this algorithm is presented below for the case of a process defined on a two-dimensional grid. Its extension to higher dimensions is straightforward.

Let the set of points on the regular grid be

$$D = \left\{ \left(\frac{j_1}{n_1}, \frac{j_2}{n_2} \right) : 0 \leq j_l \leq n_{i-1}, \quad 1 \leq i \leq 2, 1 \leq l \leq 2 \right\} \in \mathbb{R}^2$$

where $(j_1, j_2) = J$ and $(n_1, n_2) = N$ are elements of $\mathbb{Z}^2$, $\mathbb{Z}$ being the set of integer numbers. In order to simplify the notation, the pairs $\left(\frac{j_1}{n_1}, \frac{j_2}{n_2}\right)$ in D are defined as the result of the component-wise fashion division in Z^2: $\frac{J}{N} = \left(\frac{j_1}{n_1}, \frac{j_2}{n_2}\right)$. With this notation, it is required to simulate a process $\left\{ Y\left(\frac{J}{N}\right) : \frac{J}{N} \in D \right\}$ such that its covariance function is a known function $C : \mathbb{R}^d \longrightarrow \mathbb{R}$, $C\left(Y\left(\frac{I}{N}\right), Y\left(\frac{J}{N}\right)\right) = C\left(\frac{J-I}{N}\right)$ or, more generally, $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = c(\boldsymbol{s}_i - \boldsymbol{s}_j)$ for an appropriate function c and $\boldsymbol{s}_i, \boldsymbol{s}_j \in D$.

It is widely known that the points on a regular grid can be ordered in such a way that the variance-covariance matrix Σ_Y of the process is block-Teoplitz (a partition of Σ_Y exists such that each block on it is a Teoplitz matrix). It is also well known that each block-Teoplitz matrix can be embedded in a block-circulant matrix.

If the process Y is isotropic and its covariance function c is defined as:

$$c(\boldsymbol{h}) = \sigma^2 - \gamma(\boldsymbol{h}), \quad \boldsymbol{h} \in \mathbb{R}^2$$

where the semi-variogram γ belongs to a valid family, then the variance-covariance matrix Σ_Y is non-negative definite (see, for instance, Cressie (1990)). For these families it is also valid that:

$$\sum_{j \in \boldsymbol{Z}^2} \left| C\left(\frac{J}{N}\right) \right| < \infty$$

and the spectral density:

$$g(t) = (2\pi)^{-2} \sum_{j \in \boldsymbol{Z}^2} C\left(\frac{J}{N}\right) exp(-2\pi i J^T \boldsymbol{t})$$

is strictly positive for all $\boldsymbol{t} \in [0,1]^2$. Then, there exists an integer $u_0 = u_0(N, C)$ such that for all $M = (m_1, m_2)$ that satisfy $min\{m_1, m_2\} > u_0$ the matrix Σ_Y can be embedded in a positive definite matrix of order $\bar{m} = m_1 m_2$.

For each pair of positive integer numbers $K = (k_1, k_2)$ the sets I(K) and $I^*(K)$ are defined as:

$$I(K) = I(k_1, k_2) = \{(j_1, j_2) : 0 \leq j_i \leq k_i - 1, 1 \leq i \leq 2\}$$

$$I^*(K) = I(k_1, k_2) = \{(j_1, j_2) : 0 \leq |j_i| \leq k_i - 1, 1 \leq i \leq 2\}$$

The set $I(K) = I(k_1, k_2)$ has $\bar{k} = k_1 k_2$ elements. Consequently, there exists a bijection b between this set and the set $I_{\bar{k}} = 0, 1, 2, \ldots, \bar{k} - 1$. One of the possible bijections is $b_k : I(K) \rightarrow I_{\bar{k}}$, $b(j_1, j_2) = j_1 + k_1 j_2$ for each $(j_1, j_2) \in K$. This notation is used now to define a circulant variance-covariance matrix (and its corresponding stochastic process) such that it contains as a sub-matrix the variance-covariance matrix Σ_Y.

With the notation above, let the pair $M = (m_1, m_2)$ be one of the pairs that verifies that $min\{m_1, m_2\} \geq u_0$ (as defined before). For this M it is possible to define a matrix of order $\bar{m} \times \bar{m}$ with entries $\sigma_{ij} = c(i-j)$ where c is function defined as: $c(\boldsymbol{h}) = C\left(\frac{\tilde{\boldsymbol{h}}}{N}\right)$ with $N = (n_1, n_2)$ where n_1, n_2 are the numbers of points on each side of the grid and $\tilde{\boldsymbol{h}}$ is defined as follow:

$$\tilde{\boldsymbol{h}} = \begin{cases} \boldsymbol{h} & if \quad 0 \leq h_i \leq \frac{m_i}{2}, 1 \leq i \leq 2 \\ \boldsymbol{h} - M & if \quad \frac{m_i}{2} \leq h_i \leq m_i - 1, 1 \leq i \leq 2 \\ \boldsymbol{h} + M & if \quad \frac{m_i}{2} \leq -h_i \leq m_i - 1, 1 \leq i \leq 2 \end{cases}$$

for each $\tilde{\boldsymbol{h}}, \boldsymbol{h} \in I^*(M)$. The matrix Σ with entries σ_{ij}, $i, j = 1, \ldots, \bar{m}$ defined as explained before has the matrix Σ_Y as a sub-matrix and is a circulant matrix. At this stage, the idea is to build a process Z such that the matrix Σ is its variance-covariance matrix, in other words, $\Sigma_Z = \Sigma$.

Because Σ_Z is a circulant matrix, there exist a $\bar{m} \times \bar{m}$ matrix Q such that $\Sigma_Z = Q\Lambda Q^*$ (see, for instance, Brillinger (1981), pag.73) where Λ is the diagonal matrix with the latent values $\lambda_i, i = 1, \ldots, \bar{m}$ of Σ_Z on the diagonal, Q is the matrix with the latent vectors of Σ_Z, and Q^* is the conjugate transpose of Q. The latent values and the latent vectors can be calculated very efficiently by using the finite Fourier transform:

$$\lambda_{k_1+m_1k_2} = \sum_{J \in I(M)} c(J) exp\left(-2\pi i \left(\left(\frac{J}{M}\right)^T .K\right)\right)$$

where $K = (k_1, k_2), J = (j_1, j_2) \in I(M)$ and the j^{th} column q_j of Q has components

$$q_{lj} = \bar{m}^{-1/2} exp(-2\pi i \frac{lj}{\bar{m}}), l = 0, \ldots, \bar{m} - 1$$

for $j = 0, \ldots, \bar{m} - 1$. Another important property of these matrices is that its

square root can be straightforwardly calculated as ${\Sigma_Z}^{1/2} = Q\Lambda^{1/2}Q^*$. Hence, if a random vector $\boldsymbol{U} = (U_0, U_1, \ldots, U_{\bar{m}-1})$ with components U_i independent and identically distributed as standard univariate normal is generated, then the vector $Z = {\Sigma_Z}^{1/2}\boldsymbol{U} = Q\Lambda^{1/2}Q^*\boldsymbol{U}$ has zero-mean, $\bar{m}$-normal distribution with variance-covariance matrix Σ_Z.

With this procedure, the vector $Y\left(\frac{J}{N}\right) = Z\left(\frac{J}{N}\right)$ for each $J \in I(N)$ has variance-covariance matrix Σ_Y given originally.

The Wood & Chan algorithm exploits the efficiency of the fast Fourier transform. Also, it is not necessary to store big matrices even in cases where the size of the grid is big. In fact, the vector $\boldsymbol{U}$ is not calculated in the fortran program by Chan & Wood. Other complex-random vectors are used in their places and the resource of the fast Fourier transform is then used.

Figure 5.1 shows simulations of two different processes simulated with a program from Chan & Wood (1996).

The steps to be followed to apply this algorithm to generate a set of data can be summarised as follows:

1. Given a specific covariance function C defined for each point on the grid in $[0,1]^2$ with $n_1 \times n_2$ points, calculate a pair of positive integer numbers $(m_1, m_2) = M$, the sets $I(M)$ and $I^*(M)$ and the new covariance function $c(h) = C(\frac{\tilde{h}}{N})$, where $N = (n_1, n_2)$

2. Calculate the latent values $\lambda_{k_1+m_1k_2}$ for $(k_1, k_2) \in I(M)$ and the latent vectors q_j, for $j = 1, \ldots, \bar{m}-1$ with components q_{lj}; $l = 0, \ldots, \bar{m}-1$

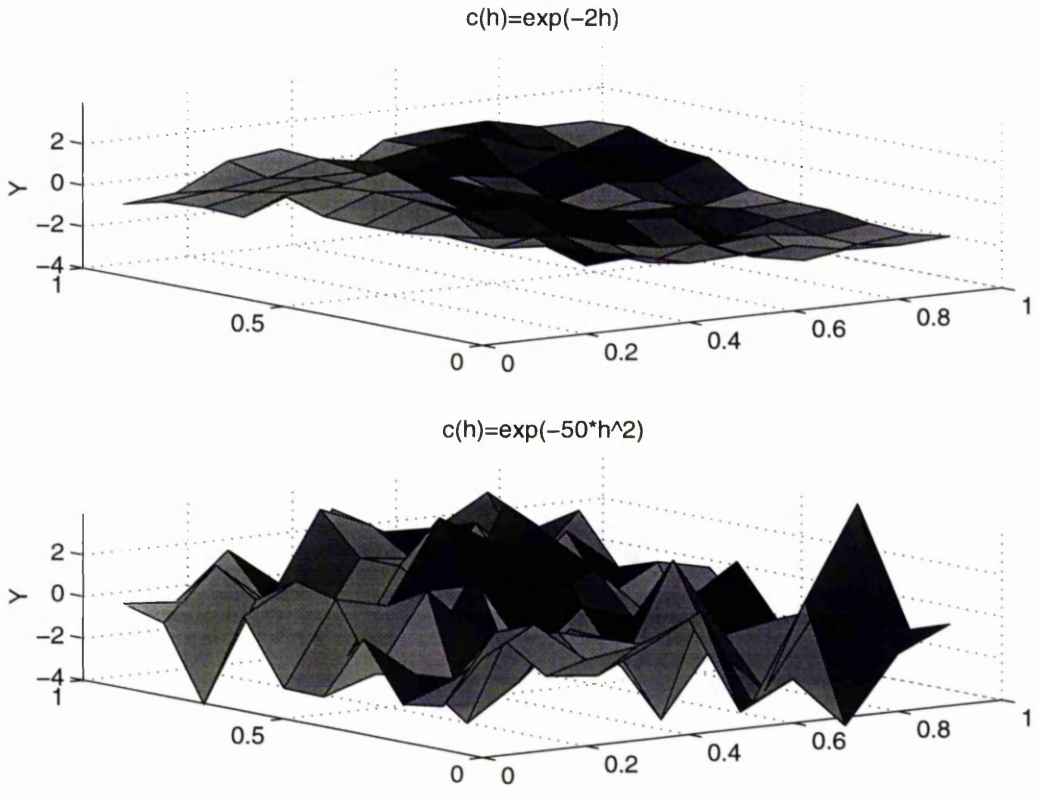

Figure 5.1. Two examples of simulated spatial normal data on a regular grid with Wood & Chan's algorithm. On the top panel, the variogram is $2\gamma(h) = 2 - 2exp(-2h)$. On the bottom, the variogram is $2\gamma(h) = 2 - 2exp(-50h^2)$. In both, the variance is $\sigma^2 = 1$

of the covariance function c by using the fast Fourier transform.

3. Generate a vector $\boldsymbol{U} = (U_0, U_2, \ldots, U_{\tilde{m}-1})$ with random independent normal components.

4. Do the transformation $Z = Q\Lambda^{1/2}Q^*\boldsymbol{U}$, where Q is the matrix with entries q_{lj} and Λ the diagonal matrix of the latent values, to calculate a realization of the process Z with covariance function c.

5. Extract a subset of the data Z obtained in the latter step.

5.2.4 The turning bands method

This method is based on the idea of simulating a process originally defined in a higher dimension through the simulation of several adequate one-dimensional projected processes. The general ideas of this approach are due to Matheron (1973). Journel (1974) has given a detailed development and applications in the context of mining in his doctoral thesis. Brooker (1985) has worked with the simulation of a two-dimensional process and found explicit relationships between two and one-dimensional covariances functions. The concept of space transformations and some of its properties are explored in Christakos (1987). He has also considered an extension of this approach to anisotropic and integrated processes.

The general theoretical background of the turning bands method deals with the Radon projection (Christakos, 1987). The main basic ideas are now explained.

If f is a function $fA :\rightarrow \mathbb{R}$, where A is a subset of $\mathbb{R}^n$, the **Radon-projection** of f over the hyper-plane $H \subset \mathbb{R}^n$ is defined as:

$$g_{\boldsymbol{u}}(\boldsymbol{x}.\boldsymbol{u}) = R[f(\boldsymbol{u})] = \int_H f(\boldsymbol{x}) dm(\boldsymbol{x})$$

where $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ is a unit vector defining the orientation of the hyper-plane in $\mathbb{R}^n$, $\boldsymbol{x}.\boldsymbol{u}$ is the inner product of the vectors $\boldsymbol{x}$ and $\boldsymbol{u}$, and dm is the Euclidean measure on H, provided the integral exists.

This definition and some of its properties lead to two operators, Ψ_n^{n-k} and

T^n_{n-k}, which assign to a function defined in R^n a function defined in a hyper-plane of R^n and vice versa. They have been defined as:

$$g_{n-k,u}(\boldsymbol{x}.\boldsymbol{u}_k) = T_n^{n-k}[f_n(\boldsymbol{x})] = \int_{\mathbb{R}^n} f_n(\boldsymbol{x})\delta(\boldsymbol{x}.\boldsymbol{u} - t)d\boldsymbol{x}$$

where f_n is a function defined in $\mathbb{R}^n$, t is the inner product between $\boldsymbol{x}$ and $\boldsymbol{u}$ which defines the position of the hyper-plane H_k in $\mathbb{R}^n$ and the delta function δ allows to selected the hyper-plane H_k from $\mathbb{R}^n$.

Another operator is given by:

$$f_n(\boldsymbol{x}) = \Psi^n_{n-k}[f_{n-k}(\boldsymbol{x}')] = \int_{\theta^n} v(\boldsymbol{u}) f_{n-k,u}(\boldsymbol{x}.\boldsymbol{u})d\boldsymbol{u}$$

where the integration is carried out over any closed surface θ^n enclosing the origin $\boldsymbol{u} = \mathbf{0}$ in $\mathbb{R}^n$, and $f_{n-k,\boldsymbol{u}}$ is a function defined on the union of a collection of hyper-planes H_{n-k} of $\Re^n$. This operator assigns to a function $f_{n-k,\boldsymbol{u}}$ a function f_n in $\mathbb{R}^n$.

With a view towards the simulation of a two-dimensional process starting from the simulation of a one-dimensional process, when a covariance function of the two-dimensional process is given, it is useful to consider how to obtain a covariance function for the one-dimensional process. In terms of the operators given above, the inverse operator of Ψ^n_{n-k} for $k = n-1$ and $u(\boldsymbol{s})$ equal to the volume of the n-sphere of unit radio is:

$$f_{1,\boldsymbol{u}}(t) = \Psi^1_{2m+1}[f_{2m+1}(\boldsymbol{x})] = \frac{(-1)^m S_{2m+1}}{2(2\pi)^{2m}} \frac{df^{2m}}{dt^{2m}} T^1_{2m+1}[f_{2m+1}(\boldsymbol{x})]$$

if $n = 2m + 1$ is odd, and, for $n = 2m$ even, it is:

$$f_{1,\boldsymbol{u}}(t) = Psi^1_{2m}[f_{2m}(\boldsymbol{x})] = \frac{(-1)^{m-1}S_{2m}}{2(2\pi)^{2m-1}} H\left\{\frac{df^{2m}}{dt^{2m}} T^1_{2m}[f_{2m}(\boldsymbol{x})]\right\}$$

where $t = \boldsymbol{x}.\boldsymbol{u}$ and H is the Hilbert transform defined as: $H[f(z)] = \frac{1}{\pi} G \int_{\mathbb{R}} \frac{f(y)}{y-z} dy$, where G indicates that the Cauchy value should be taken (Sneddon, 1972)).

Suppose, a covariance function c_2 for a two-dimensional Gaussian isotropic process is given and a process $\{Y(\boldsymbol{s}) : \boldsymbol{x} \in D \subseteq \Re^2\}$ is required to generate with the condition that c_2 is its covariance function. The procedure of turning bands reduces the simulation of this process to the one-dimensional process, say,

$$\{Z(s_{H_i}) : s_{H_i} \in H_i, H_i \subseteq D, \quad i = 1, \ldots, N\}$$

where N is a number of hyper-planes (lines for this case) to be determined and s_{H_i} is the projection of a point $\boldsymbol{s}$ on the hyper-plane H_i, as indicated in the 5.2. If the covariance function of this one-dimensional process is called c_1, the idea here is to find c_1 such that the operator $\Psi^n_{n-k} = \Psi^2_1$ applied to the function c_1 results in the function c_2. If, θ^n in the definition of Ψ is a circle θ^2 which contains the origin in $\mathbb{R}^2$, $\boldsymbol{u}_i$ is a unitary vector on H_i and the weighted function v is the constant $v = 1/\text{area of the united circle} = \frac{1}{\pi}$ then,

$$c_2(\boldsymbol{h}) = \int_{\theta^2} c_1(\boldsymbol{h}.\boldsymbol{u}_i) dr d\phi = \frac{1}{\pi} \int_0^{\pi} c_1(hcos\phi) d\phi$$

where $h = \| \boldsymbol{h} \|$ If c_1 is even and, because the two-dimensional process is isotropic it gives:

$$c_2(h) = \frac{2}{h\pi} \int_0^h \frac{c_1(x)}{\sqrt{1 - \left(\frac{x}{h}\right)}} dx \tag{5.1}$$

This operator should be inverted to find the appropriate function c_1 to simulate realizations of the one-dimensional process. Applying the expression above, for n even (n=2.m), the function c_1 is (Brooker, 1985):

$$c_1(h) = c_2(0) + \int_0^h h(h^2 - t^2)^{-1/2} \frac{d}{dt}[c_2(t)]dt \tag{5.2}$$

With the covariance function c_1 is possible to simulate a one-dimensional process Z on a line (hyper-plane) through the origin in $\Re^2$.

Now consider N lines H_i for $i = 1, \ldots, N$ with corresponding directions $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_N$ uniformly distributed over the unit half circle. Suppose it is required to simulate a process $Y(\boldsymbol{s})$ for $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n$. These points are projected on the lines H_i for $i = 1, \ldots, N$ and one-dimensional processes $\{Z(s_{H_i}) : s_{H_i} \in H_i, H_i \ \ line \subseteq D, \ \ i = 1, \ldots, N\}$ with covariance function c_1 are simulated independently on each line.

Suppose now that $\boldsymbol{s}$ is one of the points $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ (the subscript is omitted for simplicity), and $s_{H_1}, s_{H_2}, \ldots, s_{H_N}$ are its respective projections on the lines $H_1, H_2, \ldots, H_N$, the value of $Y(\boldsymbol{s})$ at $\boldsymbol{s}$ is defined as:

$$Y(\boldsymbol{s}) = \frac{1}{N^{1/2}} \sum_{i=1}^{N} Z(s_{H_i})$$

The covariance function c_2 of the process Y at $h = \| \boldsymbol{h} \|$, the distance between the points $\boldsymbol{s}$ and $\boldsymbol{s} + \boldsymbol{h}$ is:

$$c_2(h) = cov(Y(\boldsymbol{s}+Y(\boldsymbol{h})), \boldsymbol{s}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} cov(Z((\boldsymbol{s}+\boldsymbol{h}).\boldsymbol{u}_j), Z(\boldsymbol{s}.\boldsymbol{u}_i)) = \frac{1}{N} \sum_{i=1}^{N} c_1(\boldsymbol{h}.\boldsymbol{u}_i)$$

because $Z((\boldsymbol{s}+\boldsymbol{h}).\boldsymbol{u}_i)$ and $Z(\boldsymbol{s}.\boldsymbol{u}_j))$ are independent by construction when $i \neq j$, $i, j = 1, \ldots, N$

In practice, instead of the projections $\boldsymbol{s}.\boldsymbol{u}_i$, for $i = 1, \ldots, N$, each line H_i is divided in segments of equal longitude, and, simultaneously, the whole area is divided in bands (the **turning bands**) with width equal to this longitude. Then, one value of the realization of Z is generated at the middle point of this segment. When passing to the two-dimensional realization $Y(\boldsymbol{s})$, the value for the unidimensional process generated in this middle point, Zs in 5.2, is given to all points in two dimensions which lie within this band.

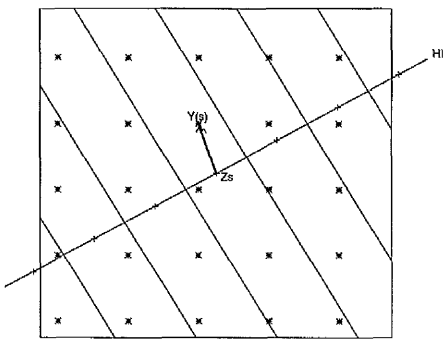

Figure 5.2. The points on the grid where the two dimensional process Y is required to be simulated (indicated with '*'), one of the hyper-planes Hi, and the turning bands corresponding to this hyper-plane. The value Zs of the simulated one dimensional process is assigned to all the points on the corresponding turning band as a value of $Y(\boldsymbol{s})$.

Then, the procedure is repeated for all the hyper-planes in a uniform distribution of them within the unit circle (see Figure 5.3).

The steps to simulate a two-dimensional zero-mean Gaussian isotropic process by using turning bands can be sumarized as follow:

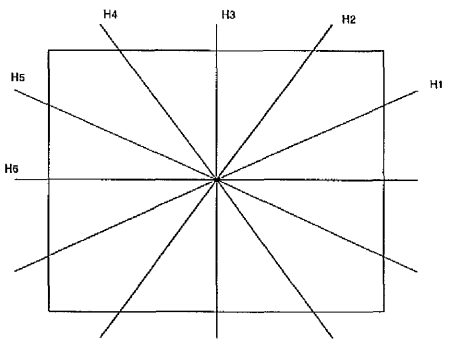

Figure 5.3. An example of hyper-planes H_1, H_2, H_3, H_4, H_5 and H_6 uniformly distributed in the unit circle in $\mathbb{R}^2$. For each hyper-plane H_i, the turning bands are built as shown in Figure 5.2.

1. Given a covariance function c_2, calculate the corresponding covariance function c_1 for a one-dimensional process by using the equations 5.1 or 5.2 above.

2. Given a set of spatial locations $\{\boldsymbol{s}_1, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \subseteq \mathbb{R}^2$, consider a rectangle, say R, which contains this set of data. The centre point of this rectangle will be considered the origin in $\mathbb{R}^2$.

3. Consider a positive integer N and calculate the angle $\phi = \pi/N$.

4. Trace a line H_i through the origin O and divide this line in segments of equal longitudes. Consider also an unitary vector $\boldsymbol{u}_i$ in the direction of H_i.

5. Trace perpendicular lines to the line H_i by the extreme points of the segments and define a **turning band** as the intersection between the band between two of these perpendicular lines and the rectangle R.

6. Simulate an univariate normal zero-mean process at each middle point

of the segments in the line H_i with covariance function c_1. Call this process Z_{H_i}.

7. Repeat the steps 3 to 5 below in such a way that the angle between two consecutive lines is ϕ.

8. For each point $\boldsymbol{s}$ in the given set $\{\boldsymbol{s}_1, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ define as the value at $\boldsymbol{s}$ for a realization of the two-dimensional process Y, $N^{1/2}$ times the average of the simulated values of all the unidimensional processes which are in the same turning bands as $\boldsymbol{s}$:

$$Y(\boldsymbol{s}) \;=\; N^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N} Z(\boldsymbol{s}.\boldsymbol{u}_i)\right) \;=\; \frac{1}{N^{1/2}}\sum_{i=1}^{N} Z(\boldsymbol{s}.\boldsymbol{u}_i)$$

where $\boldsymbol{s}.\boldsymbol{u}_i$ indicates the middle point of the turning band where $\boldsymbol{s}$ is intersection H_i.

The approximation of the distribution of the generated two-dimensional process to a normal distribution is justified by the central limit theorem. It is therefore important to consider an adequate minimum value of the number N of turning bands. Journel (1974) gives an experimental minimum number of N for a specific model (spherical) for the covariance function. In fact, some experience is required to select an appropriate value of N. The influence of small changes on the one-dimensional covariance function is translated into important changes on the two-dimensional covariance function. This property gives a warning that it is necessary to look for a way in which the one-dimensional processes can be simulated with good accuracy.

This method can be applied to simulate data in three (or more) dimensions as is often required in problems dealing with mining. Also, this method has

been extended to anisotropic and integrated processes (see Christakos, 1987).

5.3 Incorporating dependence with kernel smoothers

When a specific variogram or variance-covariance matrix for a process is not required, but it is required to generate processes with varying degrees of dependence, smoothing techniques can offer useful tools.

Suppose $\{Z_1, Z_2, \ldots, Z_n\}$ are observations at locations $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ coming from a white noise process. Then, the smoothed variables:

$$Y_i = \sum_{j=1}^{n} w_{ij} Z_j, \quad i = 1, \ldots, n$$

where w_{ij} are weights which depend only on the locations $\boldsymbol{s}_i, \boldsymbol{s}_j$, are correlated and the "degree" of dependence is controlled by the weight functions.

Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ be the vectors of the variables Y and Z, and W the matrix with entries $w_{ij}, i = 1, \ldots, n; j = 1, \ldots, n$ as defined above. Then, the n-equations above, can be written as: $\boldsymbol{Y} = W\boldsymbol{Z}$ and $\boldsymbol{Y}$ has a normal distribution with zero mean and variance-covariance matrix $\Sigma_{\boldsymbol{Y}}$ given by $\Sigma_{\boldsymbol{Y}} = WW'$. Then, if the weighting functions w_{ij} are chosen as functions of a distance between the points $\boldsymbol{s}_i, \boldsymbol{s}_j$, say, $w_{i,j} = w(\| \boldsymbol{s}_i - \boldsymbol{s}_j \|)$ the covariance function of $\boldsymbol{Y}$ will be a function of the distances between the locations. In fact,

$$cov(Y_i, Y_j) = \sum_{k=1}^{n} w_{ik} w_{jk} = \sum_{k=1}^{n} w(\| \boldsymbol{s}_i - \boldsymbol{s}_k \|) w(\| \boldsymbol{s}_k - \boldsymbol{s}_j \|)$$

This expression shows that the simulated process may be not stationary if the kernel does not have compact support and the grid is, in fact, a compact set in $\mathbb{R}^2$. It implies a non regular distribution of the weights assigned by the kernel functions to two pair of points even when they can be at the same distance. Then, the covariance depends not only on the distance between the points but also on the positions of the points on the grid.

To illustrate the statement above, suppose the that the kernel functions w_i are proportional to the normal density centred at the point $\boldsymbol{s}_i, i = 1, \ldots, n$. Then,

$$w_i(\boldsymbol{s}) = c^{-1} \exp\left(-\frac{1}{2}\left[\left(\frac{x - x_i}{b_1}\right)^2 + \left(\frac{y - y_i}{b_2}\right)^2\right]\right)$$

where (x, y) are the coordinates of $\boldsymbol{s}$ and (x_i, y_i) are the coordinates of $\boldsymbol{s}_i$ and c is a constant to make the sum of the weights equal one:

$$c = \sum_{j=1}^{n} exp\left(-\frac{1}{2}\left[\left(\frac{x - x_i}{b_1}\right)^2 + \left(\frac{y - y_i}{b_2}\right)^2\right]\right)$$

Then, for a grid as in figure 4, the points $\boldsymbol{s}_{12} = (0.1, 0.1)$ and $\boldsymbol{s}_{13} = (0.2, 0.1)$ and the points $\boldsymbol{s}_{33} = (0.2, 0.3)$ and $\boldsymbol{s}_{34} = (0.3, 0.3)$ are at the same distance $\| \boldsymbol{s}_{12} - \boldsymbol{s}_{13} \| = \| \boldsymbol{s}_{33} - \boldsymbol{s}_{34} \| = 0.1$ whereas the covariances are $cov(Y_{12}, Y_{13}) = 0.0496$ and $cov(Y_{33}, Y_{34}) = 0.0339$. This is a result of the so called "edge-effects" of smoothing techniques.

The idea of smoothing a white noise process to obtain a dependent one is a generalisation of the moving average procedure, or more generally speaking with the filtering of a series, in the context of time series.

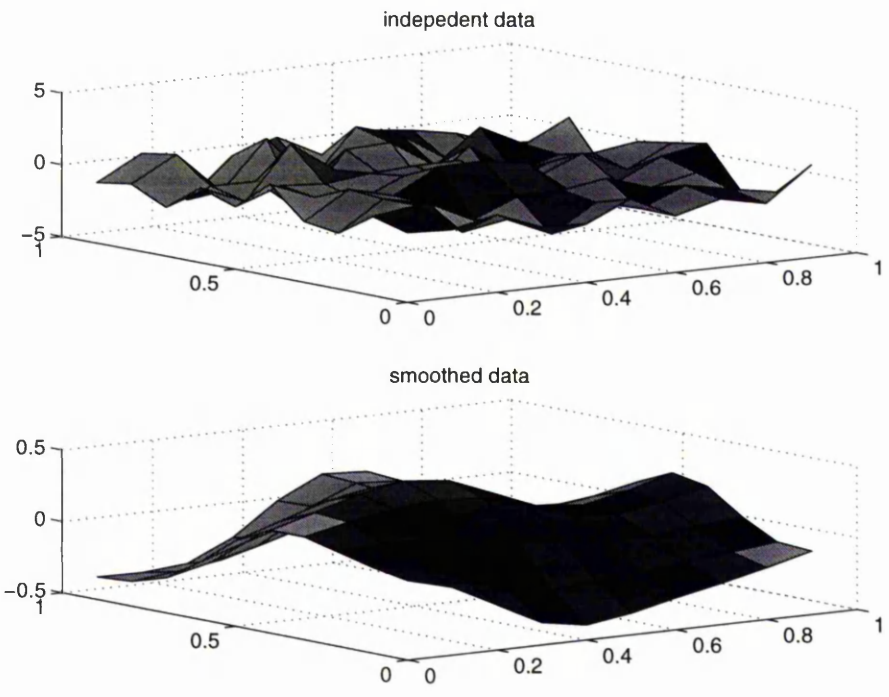

Figure 5.4. On the top panel a realization of a white noise process is shown. From these data a realization of a process with non diagonal variance-covariance matrix is generated through the application of a kernel smoother with weighted functions proportional to independent normal density functions with standard deviations b1=b2=0.15

One way to build a stationary process by using the ideas of smoothing techniques is to consider kernels with compact support. This approach can be thought of a tapering procedure in the time series scenario (see, for instance, Brillinger, 1981). To illustrate this method, suppose the kernel functions w_i are as considered before, and "tapers" or "windows" of the form

$$h(\boldsymbol{u}) = \begin{cases} 1 & if \parallel \boldsymbol{u} \parallel \leq b \\ 0 & \text{otherwise} \end{cases}$$

are used.

Then, a kernel with compact support can be constructed as the product of these two functions. For example; let b_1 and b_2 take the same value $b = b_1 = b_2$, $h_i(\boldsymbol{u}) = h(\frac{\boldsymbol{u}-\boldsymbol{s}_i}{b})$. The functions k_i,

$$k_i(\boldsymbol{u}) = h\left(\frac{\boldsymbol{u}-\boldsymbol{s}_i}{b}\right).w_i(\boldsymbol{u}) = \begin{cases} c.exp\left(-\frac{1}{2} \parallel \frac{\boldsymbol{u}-\boldsymbol{s}_i}{b} \parallel^2\right) & if \quad \parallel \frac{\boldsymbol{u}-\boldsymbol{s}_i}{b} \parallel \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where $c = \sum_{j=1}^{n} k_i(\boldsymbol{s}_j)$ for $j = 1, \ldots, n$ have support:

$$S(k_i) = \{\boldsymbol{u} : \parallel \boldsymbol{u} - \boldsymbol{s}_i \parallel \leq b\} = \text{circle in } \mathbb{R}^2 \text{ with centre at } \boldsymbol{s}_i \text{ and radius b.}$$

Figure 5.5 shows some contours of the kernels k_i and k_j where $\boldsymbol{s}_i = (0.5, 0.5)$ and $\boldsymbol{s}_j = (0.1, 0.1)$, a point close to the border of the rectangle $[0,1] \times [0,1]$ in which the grid is included. The function K_i has support included in the rectangle R but the support $S(k_j)$ of the function K_j is not included in this rectangle and, consequently, the data at the points in the neighbourhood of the point $\boldsymbol{s}_j$ are receiving different weights from those which are at the same distance but in the neighbourhood of the point $\boldsymbol{s}_i$. This characteristic leads to a loss of stationarity.

This border effect can be eliminated if a subset of the simulated data is considered. Suppose, a set of data is generated on a $m \times m$ grid in $R = [0,1] \times [0,1]$ with compact support kernels k_i with support set a circle with centre at $\boldsymbol{s}_i$, $i = 1, \ldots, m$ is simulated. Suppose also that k is the minimum positive integer such that $k/(m-1) \geq b$ and $n = m - 2k$, then the support of k_i is included in R for $(j-1)m+(k+1) \leq i \leq (j-1)m+m-k$ and $1 \leq j \leq n$. Consequently, if the subset of the simulated data $\{Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_{mm})\}$ at the points $\boldsymbol{s}_i$ with $(j-1)m + (k+1) \leq i \leq (j-1)m + (k+1) + n$ and

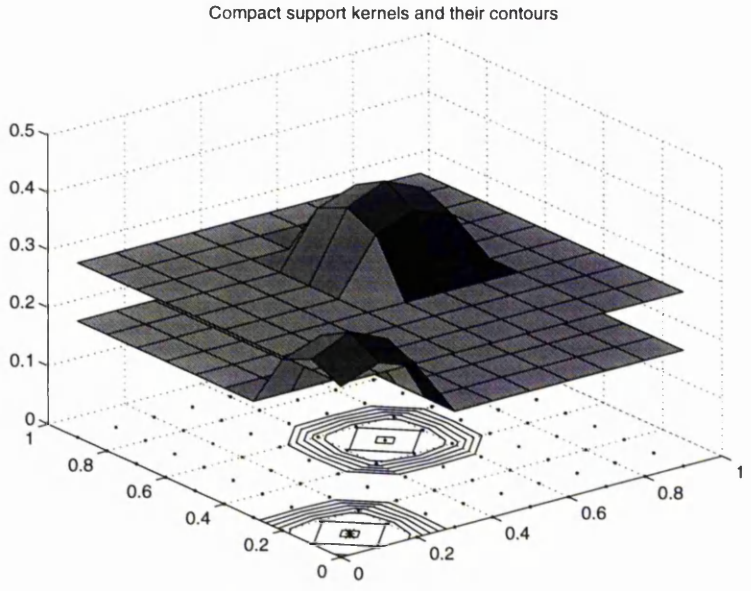

Figure 5.5. This figure shows two kernels with compact support centred at the points $(0.1, 0.1)$ and $(0.5, 0.5)$. They were built from a normal normal densities with means at these points and standard deviations $b_1 = b_2 = 0.15$ multiply by a tapering function h as described in the text with $b = 0.15$.

$k + 1 \leq j \leq (k + n)$ is extracted from that set, it will enjoy the property of stationarity and isotropy. The variance-covariance matrix of this subset is the matrix VV' where V is the sub-matrix of W (as explained above) which contains the rows with the same subscripts that the selected data. From this matrix is also possible to extract the values of the covariance function. One example of this procedure is shown in figure 5.6.

In conclusion, the use of kernels with compact supports, and the extraction of data from the interior of the region, enables stationary, dependent data to be simulated.

The symmetry of kernels used will guarantee isotropy.

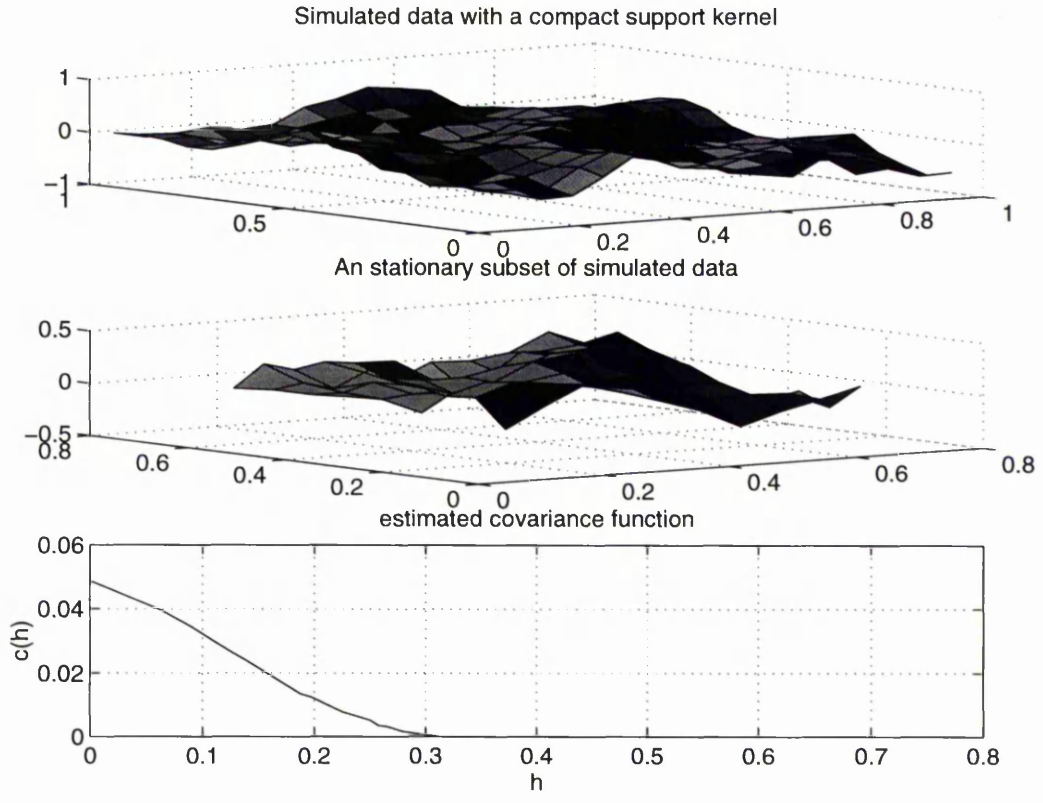

Figure 5.6. On the top panel a realization of a process simulated by using a normal kernel smoother on a 16×16 grid from a white noise process. The kernel smoother is proportional to the product of two independent normal densities with standard deviation $b_1 = b_2 = 0.15$. On the middle panel a stationary subset of data is shown. On the bottom panel the observed covariance function c from the latter set at each value of the distance between two points is graphed.

In terms of computation, the simulations are very straightforward to perform, both from the point of view of the calculations and storage capacity.

5.4 Re-sampling methods for hypothesis testing

Permutation or randomisation techniques have great attractions because of some of their properties. In fact, one of their more attractive properties is that they do not require assumptions about the distribution which the sample values were taken from. Another is the very simple idea of permutation or randomisation they are based upon. These ideas are also supported by a usually quite satisfactory efficiency. The popularity of this method is increasing along with the computer power which makes implementation straightforward. Most of the impetus to use distribution-free methods was originally in hypothesis testing as used here.

A method based on permutation of the values of a realization of a process was used here in order to calculate the empirical distribution of the test statistic for constant variogram. This approach is very similar to that of the non-parametric bootstrap in which a set of data is generated from the original set. The difference between these two approaches is that in the bootstrap approach every new set of data is obtained as a sample with replication form the first one.

As an example of an application of the permutation method the calculation of the p-value for the test statistic for constant variogram is considered now. In this case the steps followed were:

1. Calculate the value of the test statistic for the set of data.

2. Permute the spatial locations associated with the individual observations. This represents the null hypothesis of absence of spatial correlation.

3. Calculate the observed value of the test statistic with the permuted values.

4. Repeat steps two and three a large number of times to calculate the empirical distribution of the test statistic under the null hypothesis.

5. With the distribution calculated in the latter step, calculate the corresponding quantile for the observed value of the statistic obtained in the first step.

5.5 Building local confidence intervals for the variogram through block-bootstrap

5.5.1 Introduction

If the observations are independent the bootstrap methodology is quite useful to calculate the empirical distribution of a statistic based on these observations. However a similar statement cannot be made for dependent observations. The independent and identically distributed re-sampling scheme associated with the method fails to capture the underlying dependence in the joint distribution of the observations.

For this reason, some efforts have been made to extent this methodology

to the case of dependent observations. One technique which deals with the estimation of distributions under the assumption of dependence is the "block-bootstrap".

A statistic of interest in the context of spatial processes, and particularly in their applications to geostatistical data, is an estimator of the variogram. When the underlying distribution is Gaussian, some efforts have been made in the literature to approximate the distribution of the classical estimator of the variogram. However, these approximations usually assume normality and are based on large sample sizes.

Standard errors bars are sometimes displayed on top of estimated values of the variogram. There is a danger of misinterpreting these, because the underlying points are not independent. A motivation for the block-bootstrap is to attempt to represent the variability in the estimator of the variogram more accurately.

An algorithm based on a subsample or window due to Hall and Jing (1994) is explored here, to assess its effectiveness in calculating confidence intervals for the variogram. A Gaussian second order stationary and isotropic process with a specific variogram is generated. Local confidence intervals for the semi-variogram are calculated via this block-bootstrap approach and the coverage accuracy is calculated for some particular variogram models. These confidence intervals are also compared with those based on the approximate distribution for the classical estimator given by Baczkowski & Mardia (1987) under the assumption of normality.

5.5.2 Previous Literature

The bootstrap methodology as introduced by Efron (1979) to estimate the underlying distribution of a statistic when the observations are independent, fails when they are dependent. Singh (1981) (Remark 2.1) gave a simple example to show that, even in the case of weak dependendent processes, the bootstrap sample mean is not a consistent approximation to the sample mean.

The problems under the assumption of dependence have been recognised and several efforts have been made to extend the bootstrap methodology to the case of dependent observations. In fact, a re-sampling tool like this would be invaluable in the field of statistical inference.

Perhaps the first works concerning with bootstrap under assumptions of dependence are those due to Davis (1977), Feedman (1984) and Efron and Tibshirani (1986). In these three works the bootstrap method is applied to ARMA models by reducing consideration to innovations which are independent and identically distributed. Modifications of the classical bootstrap for specific autoregressive models have been tackled by Bose (1988) and Besawa et al (1989).

Carlstein (1985) has considered the use of sub-series of a stationary sequence to estimate the variance of a general statistic. The general idea is to calculate the values of a statistic from non-overlapping sub-series. These values are used to model the sampling variability of the statistic. The variance of the statistic is the sample variance as calculated from the sub-series.

The method known as "moving block bootstrap" has been independently formulated by Künsch (1988) and Liu and Singh (1991) in a significant breakthrough. The first one considered an extension of the bootstrap and jackknife methods of estimating standard errors to the case where the observations come from a general stationary sequence. The general idea about the bootstrap approach proposed by Künsch can be sumarized as follow.

Suppose that $Y_1, Y_2, \ldots, Y_N$ are observations from a stationary process and T_N is a statistic of the form:

$$T_N(Y_1, Y_2, \ldots, Y_N) = T(\rho_N^m)$$

where ρ_N^m is the empirical m-dimensional marginal given by:

$$\rho_N^m = (N - m + 1)^{-1} \sum_{t=0}^{N-m} \delta_{(Y_{t+1}, \ldots, Y_{t+m})}$$

where δ_x is the point mass at $x \in \mathbb{R}^m$. If a block of observations is denoted by: $B_t = (Y_t, Y_{t+1}, \ldots, Y_{t+m-1})$ and n=N-m+1, then the empirical marginal can be written, in terms of blocks of length l (n=kl) as: $\rho_N^m = n^{-1} \sum_{t=1}^{n} \delta_{B_t}$. Then, for a given $k \in N$, Künsch considers the random selection of k blocks through k-random numbers $S_1, S_2, \ldots, S_k$ uniform and identically distributed on $0, 1, \ldots, n - l$. With this notation, the bootstrap m-dimensional marginal can be written as:

$$\rho_N^{m*} = n^{-1} \sum_{j=1}^{k} \sum_{t=S_j+1}^{S_j+l} \delta_{B_t}$$

and the bootstrap statistic $T_N^* = T(\rho_N^{m*})$ with the bootstrap variance defined as:

$$\sigma_{boot}^2 = var^*(T_{N^*}) = E^* \left[(T_N^* - E^*[T^*])^2 \right]$$

Usually σ^2_{boot} and the distribution of $T^*_N - T_N$ has to be evaluated by simulation.

The distribution of the statistic T_N depends on the unknown distribution of $(Y_1, Y_2, \ldots, Y_N)$ and it is impossible to estimate this distribution from a finite marginal. For m=1, the case in which the statistic T_N is a function of univariate marginals, the proposed estimate for the distribution of $(Y_1, Y_2, \ldots, Y_N)$ is $(\rho^l_N)^{\otimes k}$ (the product of measures) which coincides with Efron's bootstrap estimate (independent observations) if each block contains only one point (l=1). When $m > 1$, the idea is to estimate the distribution of T_N as explained before.

An important remark here is that the use of observations from independent blocks are not convenient for the calculation of the marginal ρ^{m*}_N and a smooth transition between observations left out and observations with full weight is suggested. These ideas led to an empirical estimator of the form:

$$\rho^{m*}_N = \left(\sum_{t=1}^{n} W(t/l)\right)^{-1} \sum_{t=1}^{n} W(t/l)\delta_{B_t}$$

where $(W(t))_{t\in\Re}$ is a positive stationary process with continuous covariance function R(t) independent of (Y_t), for instance, $cov(W(t), W(s)) = max(1 - |t-s|/l, 0)$.

Künsch also performed a simulation study where the statistics T_N are the least squares estimators of the parameters for autoregressive models, AR(p), for p=1,2, and moving average models, MA(1). For these statistics, the moving block bootstrap is found to be more efficient than the method from Carlstein based on nonoverlaping sub-series (the highly nonlinearity of these statistics

requires much longer sub-series to achieve similar efficiency). The distributions of these statistics are also better approximated by a normal than those from the Efron and Tibshirani's method (Efron and Tibshirani (1986)).

Lahiri (1991) analysed second-order optimality (as in the case of independent bootstrap) for the approach proposed by Künsch for stationary dependent data. She found that, for statistics based on the sample mean and, under appropriate conditions, the overlapping or moving block bootstrap enjoys a second order optimality. This work is based on the assumption of weakly dependent observations. However, in another paper (1993), Lahiri considered the assumption of long-range dependent observations and showed that the moving block bootstrap provides a valid approximation to the distribution of normalised sample mean if and only if it is asymptotically normal.

Politis and Romano (1990, 1994) have proposed a bootstrap method by re-sampling blocks of random length. This method produces stationary bootstrap data but the selection of the probability function to choice randomly the bootstrap samples is its major barrier. Also Politis and Romano (1992,1993) introduced a a generalised moving block method to set a confidence interval for the spectral density of a stationary process.

5.5.3 The methodology

One of the methodologies used here is this from Hall and Jing (1993). It is called the "sampling window". The general idea is to build windows or sub-blocks from the whole sample. It can be summarised as follow.

Let the observed process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ where D is a lattice whose dimensions are $m_1 \times m_2 \times .... \times m_d$. A point $\boldsymbol{s}_i \in D$ has the form $\boldsymbol{s}_i = (s_{i_1}, s_{i_2}, \ldots, s_{i_d})$ where $1 \leq i_j \leq m_j, \quad 1 \leq j \leq d$. Therefore, if I is the set of indices:

$$I = \{(i_1, i_2, \ldots, i_d) : 1 \leq i_j \leq m_j; 1 \leq j \leq d\}$$

and $Y(\boldsymbol{s}_i) = Y_i, \quad i \in I$, the observed process can be written as: $\Upsilon = \{Y_i : i \in I\}$ and the sample size is $n = \prod m_i$.

Following this notation, a sampling window for $n' = \prod_{i=1}^{d} m_i'$ observed points can be defined as a set of indices:

$$W = \{0, \ldots, m_1' - 1\} \times \cdots \times \{0, \ldots, m_d' - 1\}$$

where $1 \leq m_j' \leq m_j; 1 \leq j \leq d$ and the ratios m_i'/m_i are similar to one another. This window is placed onto the set of indexes I to obtain the subset of indexes:

$$I_i = i + W = \{i + j : j \in W\}$$

where $i \in J = \{1, \ldots, m_1 - m_1' + 1\} x \cdots x \{1, \ldots, m_d - m_d' + 1\}$ With this notation a rectangular subset of data can be written as $\Upsilon_i = \{Y_j : j \in I_i\}$.

Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ be a process with covariance function $C(h) = \sigma^2 - \gamma(h)$, for each distance h between two points in D, where the variogram 2γ belongs, as a function, to an adequate specific family.

The classical and unbiased estimator of the variogram is that from Matheron:

$$G(h) = 2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{i,j \in N(h)} (Y_i - Y_j)^2$$

where $N(h) = \{(i,j) : \|s_i - s_j\| = h\}$ and $|N(h)|$=number of elements in N(h).

The idea here is to estimate via block-bootstrap the distribution functions $F_{G(h)}$ of G(h) for each h as a distance in D:

$$F_{G(h)}(x) = P(G(h) \leq x), \quad for \ \ each \ \ real \ \ x.$$

If $G_i(h)$ is the estimated variogram in h as calculated from the subset Υ_i, then a natural estimated distribution function for the estimated variogram at h is the average over all the subsets Υ_i:

$$\tilde{\mathrm{F}}_{G(h)}(x) = \frac{1}{M} \sum_{i \in J} I(G_i(h) \leq x)$$

An appropriate calibration of $\tilde{\mathrm{F}}$ is the function $\hat{F}$ as suggested by Richardson (1991):

$$\hat{F}_{G(h)}(x) = \{1 - (n\prime/n)\}\{2\Phi(x) - 1\} + (n\prime/n)\tilde{\mathrm{F}}(x)$$

where Φ is the standard normal distribution function. These empirical distribution functions were used to build confidence intervals for the variogram of a process under the assumptions mentioned previously.

A $\beta 100\%$-confidence interval for the value $2\gamma(h)$ of the variogram 2γ at the point h is defined as the interval

$$\{x : x_{(1-\beta)/2} \leq x \leq x_{(1+\beta)/2}\}$$

where $x_{(1-\beta)/2}$ and $x_{(1+\beta)/2}$ are the $(1-\beta)/2$ and $(1+\beta)/2$ respective quantiles of the distribution $\hat{F}_{G(h)}$.

Under the assumptions of second-order stationary and isotropy, the variables $U_i = \frac{G_i(h)-2\gamma(h)}{\hat{\sigma}(G_i(h))}$ and $V = \left(\frac{n}{nl}\right)^{1/2} \frac{G(h)-2\gamma(h)}{\hat{\sigma}(G_i(h))}$, where $G_i(h)$ is the version of
$U_i = \frac{G_i(h)-2\gamma(h)}{\hat{\sigma}(G_i(h))}$ and $V = \left(\frac{n}{nl}\right)^{1/2} \frac{G(h)-2\gamma(h)}{\hat{\sigma}(G_i(h))}$, where $G_i(h)$ is the version of
h, and $\hat{\sigma}(G_i(h))$ is its corresponding asymptotic estimated variance, $i \in J$, have asymptotical normal joint distribution (Hall & Jing, (1994)) and correlation coefficient of order $r = \left(\frac{n'}{n}\right)^{1/2} = nl^{-1}$. The order of magnitude of the remainder is minimised by taking nl of size $n^{1/7}$ in which case the remainder is of order $n^{-8/7}$.

5.6 Simulation results

5.6.1 Normal processes

Tables 5.1 and 5.2 show simulation results of coverage percentages from 500 simulations for each model for the exponential semi-variogram and each indicated window size. These percentage coverages were calculated as the number of times (divided by 500) that the parameter $\gamma(h)$ belonged to the 95%-confidence intervals for it. In all these cases, the realizations were from a zero-mean normal distribution. Two different methods of building confidence intervals were considered. One is the block-bootstrap approach in its "sampling window" variant, as suggested by Hall and Jing (1993).

Another approach shown in tables 1 and 2 is the calculation of confidence intervals for the semi-variogram under the assumption of a log-normal distribution for the classical estimator for the semi-variogram, when the process

is normal. Backowski & Mardia (1987) have considered this approach under the additional assumption that the differences $U_l = Y_i - Y_j$, for each pair of indexes (i, j) such that $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| = h_l$ are independent. Under these assumptions, the classical estimator for the semi-variogram at the distance h_l, $\hat{\gamma}(h_l) = \frac{1}{N(h_l)} \sum_{(i,j)} (Y_i - Y_j)^2$ for $N(h_l)$ the number of pairs (i, j) as explained before, has an asymptotic normal distribution $N(\gamma(h_l), 2\gamma^2(h_l)/N(h_l))$. These ideas led them to conclude that $log(\hat{\gamma}(h_l))$ has a normal distribution with mean equal to $log(\gamma(h_l))$ and variance $2/N(h_l)$. Hence, a $\beta 100\%$-confidence interval for the semi-variogram at the distance h, $\gamma(h)$ is given by:

$$\left[\; exp\left(-q_{(1+\beta)/2}\sqrt{\frac{2}{N(h)}}\right)\hat{\gamma}(h) \;\; , \;\; \exp\left(q_{(1+\beta)/2}\sqrt{\frac{2}{N(h)}}\right)\hat{\gamma}(h) \; \right]$$

where q_α is the α-quantile of the standard normal distribution.

From the simulation results it is possible to conclude:

1. The coverage percentage slightly depends on the different models for the variogram, both in the block-bootstrap approach and the log-normal approach.

2. The coverage percentage slightly increases with the value of h. This is not a surprising result given that the number of overlapping windows taking into account for the interval estimator of the semi-variogram increases as the value of h increases.

3. The coverage percentage for the block-bootstrap approach increases as the size of the window decreases from $nw = 8$ to $nw = 5$ and decreases as the size of the window decreases form $nw = 5$ to $nw = 3$, for a specific value of h. In fact, $nw = 8$ is too big compared with a grid side

size of $ns = 10$. It implies that the number of windows is too small to estimate the distribution of the semi-variogram estimator even when the sub-grids give highly correlated information in this case. When the window size decreases to $nw = 5$, the number of windows increases (respect to the previous size) and the dependence of the underlying distribution is still gathered in the three models of dependence structure considered here. When the window size decreases to $nw = 3$, the number of windows increases and the variance of the empirical distribution decreases, but its bias increases (see Hall & Jing (1994)) and its capacity to gather the underlying dependence is reduced.

4. Despite the asymptotic result (a window side size equal to the seventh root of the whole grid side size) from Hall and Jing (1992) about an optimum estimator for the block-bootstrap sampling, the simulation results for the models considered here show a reasonably good coverage for a window side size of a half of the whole grid side size in the case of a 10×10 grid.

5. In the approach of the log-normal distribution, the coverage percentages are similar over all ranges of h-values. These coverages are all under the 95%, the real level of confidence. In fact, the log-normal distribution is a result of the assumption of independence of the differences $Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)$, $i, j = 1, \ldots, n$ which does not hold for the models considered here. For the smallest window side size ($nw = 3$) the coverages are very similar to those in the bootstrap approach, where the dependence of the underlying distribution is broken because of the large number of non-overlapping windows.

Figure 5.7 shows the block-bootstrap distribution of the estimator of the semi-variogram for a window side size equal to 5 (side grid size of 10) for three different values of h. As remarked above its dispersion increases as h increases. The line on the horizontal plane joints the values of the theoretical (population) semi-variogram. In Figure 5.8 the block-bootstrap distribution

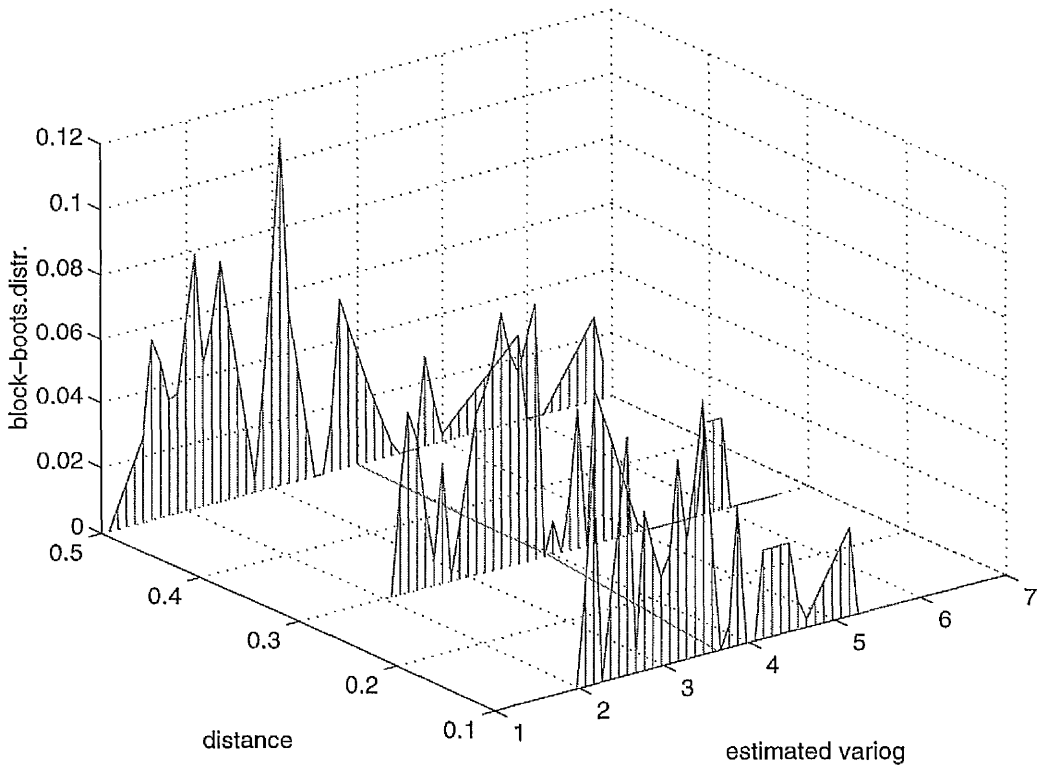

Figure 5.7. The window block-bootstrap distribution of the classical estimator of the variogram at the distances h=0.1, 0.3 0.5 between points on a regular grid in $[0,1]^2$ is shown in this figure. The line in the horizontal plane is the variogram of the simulated process.

for the same window and grid side sizes and the log-normal distribution for the estimator of the variogram at the distance $h = 0.50$ is shown.

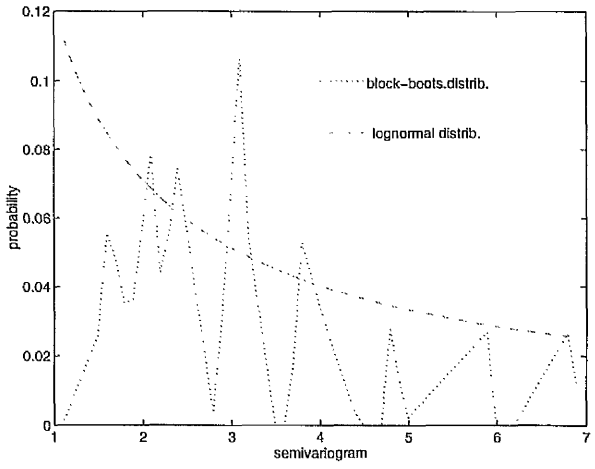

Figure 5.8. The window block-bootstrap distribution of the classical estimator of the semi-variogram at the distances $h = 0.5$ between points on a regular grid in $[0,1]^2$ is shown in this figure. The corresponding log-normal distribution is also shown in the figure. The process simulated is Gaussian with semi-variogram given by $\gamma(h) = 4 - \exp(-10h)$.

In figure 5.9, the extremes of the confidence intervals as calculated from the block-bootstrap approach and the log-normal distribution is graphed. The values of the theoretical semi-variogram, $\gamma(h) = 4 - \exp(-10h)$ is also represented.

5.6.2 Non-normal processes

When the distribution of the underlying process is not normal, the bootstrap methods are even more attractive because they are one of the few tools to resort to in many cases. Some examples about confidence intervals for the variogram of a non-normal stochastic process are analysed now. As explained

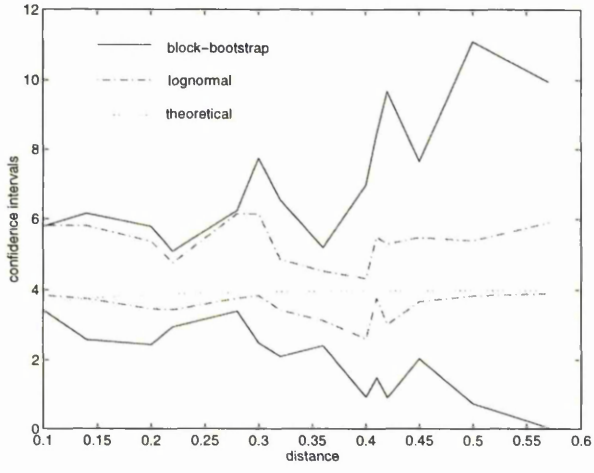

Figure 5.9. Local confidence intervals of the semi-variogram as calculated with the moving window block bootstrap method and the log-normal distribution of its classical estimator. The theoretical semi-variogram is given by $\gamma(h) = 4 - \exp(-10h)$, for each distance h between the points on the 10×10 grid in $[0,1]^2$

earlier in this chapter some approaches for the simulation of a Gaussian process have been developed in the literature. There are no works concerning with the simulation (in more than one dimension) of a non-Gaussian process. This reason leads to the search for a construction of non-Gaussian process from a Gaussian one. In fact, even when the distribution is not important for the block-method followed here, the study of the coverage percentage for the confidence intervals requires the knowledge of the variogram (semi-variogram) of the process.

Following the notation used above, it is required to generate a process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ where D is a subset of $\mathbb{R}^2$ (it might be $\mathbb{R}^m$, for $m \leq 2$ but it is considered $m = 2$ for simplicity). It is assumed that the rectangle $[0,1]^2$ is included in D, and a realization $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)^T = (Y_1, Y_2, \ldots, Y_n))^T$ of this process at the points $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\} \in D$ is required.

Let $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2), \ldots, Z(\boldsymbol{s}_n))^T$, be a realization of the Gaussian process $\{Z(\boldsymbol{s}) : \boldsymbol{s}$ at the same points $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\} \in D$. where $\boldsymbol{Z}$ has a normal distribution $N(\mathbf{0}, \Sigma_Z)$ whose ij^{th} entry is

$$\sigma_{ij} = \sigma^2 - \gamma_Z(\| \boldsymbol{s}_i - \boldsymbol{s}_j \|)$$

and where γ_Z is the variogram of the process defined in the same set D.

It will be assumed here that the realization of the process $\mathbf{Y}$ is obtained from the process $\boldsymbol{Z}$ through a transformation $H : \mathbb{R}^n \to \mathbb{R}^n$ where $H(\boldsymbol{Z}) = H(Z_1, Z_2, \ldots, Z_n) = (H_1(\boldsymbol{Z}), H_2(\boldsymbol{Z}), \ldots, H_n(\boldsymbol{Z})$.

Example

Let the transformation H above be defined as

$$H(\boldsymbol{Z}) = (Z_1^k, Z_2^k, \ldots, Z_n^k)^T = (Y_1, Y_2, \ldots, Y_n)^T$$

where k is a positive integer. To calculate the variogram γ_Y it is necessary to calculate the first and second-order moments: $E(Y_i)$ and $E(Y_iY_j)$, $i = 1, \ldots, n; j = 1, \ldots, n$. From this point of view, this distribution has the advantage that recursive formulas for $E(Y_iY_j) = E(Z_i^k Z_j^k)$ can be found in the literature (see, for instance, Cramer (1970)). From these expressions and for $k = 2$, is:

$$E(Y_iY_j) = E(Z_i^2 Z_j^2) = E(Z_i^2)E(Z_j^2) + 2E^2(Z_iZ_j) = var^2(Z_i) + 2cov^2(Z_i, Z_j) =$$

$$= \sigma^4 + 2(\sigma^2 - \gamma_Z(\| \boldsymbol{s}_i - \boldsymbol{s}_j \|)^2 = 3\sigma^4 + 2\gamma_Z^2(\| \boldsymbol{s}_i - \boldsymbol{s}_j \|) - 4\sigma^2\gamma_Z(\| \boldsymbol{s}_i - \boldsymbol{s}_j \|)$$

for $i = 1, \ldots, n; \quad j = 1, \ldots, n$ and $E(Y_i) = E(Z_i^2) = \sigma^2$, for $i = 1, \ldots, n$. Then,

$$cov(Y_i, Y_j) = E(Y_iY_j) - E(Y_i)E(Y_j) = 2\sigma^4 + 2\gamma_Z^2(\| s_i - s_j \|) - 4\sigma^2\gamma_Z(\| s_i - s_j \|)$$

$i = 1, \ldots, n; \quad j = 1, \ldots, n$. Consequently, for $i = j$,

$$var(Y_i) = 2\sigma^4, \quad i = 1, \ldots, n.$$

Then, the values of the variogram, γ_Y, of the process $\{Y(s) : s \in D\}$, at each distance h between the points on the grid can be calculated from the values of the variogram γ_Z and the variance σ^2 of the process $\{Z(s) : s \in D\}$ with the expression:

$$\gamma_Y(h) = 4\sigma^2\gamma_Z(h) - 2\gamma_Z^2(h)$$

where h is the distance between two points on the grid.

Tables 5.3 and 5.4 show the coverage percentage of confidence intervals for the variogram at each distance h between the points on a 10×10 grid in $[0, 1]^2$. The same window side sizes as for the normal case in tables 5.1 and 5.2 are considered. The total number of simulations is again 500. Some conclusions to be drawn from these simulations are:

1. From table 5.3, for a window 8×8, the coverage percentage is very small for almost all values of h as a result of the additional (to the normal case) non-normality of the underlying distribution (see Lahiri (1993))

2. Similar conclusions can be pointed out for the window side sizes of table

5.4. Nevertheless, the coverage percentage increases for the same values of h over those considered in table 5.4.

3. For each window side size, the coverage percentage increases as h increases as a result of the larger number of overlapping windows taken into account in the calculation of the corresponding confidence interval for $\hat{\gamma}(h)$.

5.6.3 Some remarks

The technique of block-bootstrap is an interesting tool for the estimation of confidence intervals for the variogram even in the case of normality of the underlying process. Nevertheless, some open problems of accuracy and adjustments are still remain:

1. Under normality of the underlying process, the selection of an adequate window side size could be obtained as a compromise between the number of windows and the size of it. As in the context of non-parametric kernel estimation the selection of the bandwidth is a compromise between bias and variance, the selection of the window size for the estimation of empirical distributions of a statistic, for a specific variogram model, is a compromise between these two parameters.

2. The dependence structure of the underlying process is another factor to take into account for the window side size selection. Even though, in the study here a window side size of a half of the whole grid side size was adequate, the relationship between the "degree of the dependence"

(as defined by Davis & Borgman (1979), for instance) and the window side size can simplify its choice. In fact, if the degree of the dependence is, say m, and $m < n_s$ (n_s is the side size of the whole grid), then a window side size of m (or greater) will be big enough to capture the underlying dependence. In the context of processes defined on points in one dimension, Léger, Politis and Romano (1992) have found that when the sample size is $n = 100$, the bootstrap estimators for independent data (block-size equal to one), seriously underestimate the variance and a large block-size (greater or equal to twenty) produces an estimates with a bias which is larger than the bias with size equal to one.

3. If the process is not Gaussian, a healthy advice is : "Do not use block bootstrap unless the distribution of the data is not "very" different from normal, and the whole grid side size is much bigger than the one considered here so that the relashionship between the window side and the grid size can be guided by the asymptotic results of Hall & Jing (1994)".

5.7 Some general comments

The simulation of a process with specific covariance matrix in a region of $\mathbb{R}^2$ or $\mathbb{R}^3$ is an invaluable tool in the area of spatial statistics. If the process is Gaussian some approaches have been proposed and analysed in the literature. The most widely known methods have also been described in this chapter. Some of these algorithms assumed a regular distribution of the points in the region (as the Wood and Chan). Others, even when this assumption is not

made, are not able to be used in practice because of their computational performance (LU-decomposition of the covariance matrix, turning bands).

The method of using kernel smoothers to generate a more general process from a white noise process is an alternative way to build an algorithm to generate a process in a non-regular grid. The tools coming from Fourier analysis can provide useful material to resort to in order to start with a specific variance-covariance matrix. This method can be easily extended to more than two dimensions and to weighted combinations of independent random variables not necessarily normal.

Building confidence intervals for the variogram of a Gaussian process was an aim in this chapter. In the environment of non-Gaussian processes this is still an open problem, given the behaviour of the block-bootstrap method under the relaxation of this assumption.

More generally, in the context of block-bootstrap methods and even for the case of a normal underlying process, its extension to non-rectangular grids is a challenge and a necessity. Most of the problems coming from the "real world" are not constrained to a rectangular region.

Despite the fact that the simulation methods shown in this chapter have been targeted to model characteristics of spatial processes, they can be easily adapted to the environment of time series modelling. For example, confidence intervals for the variogram and/or the CO-variogram of a process defined in points on a subset of the real line.

		$\gamma(h) = 4 - \exp(-ch)$					
		$c = 3$		$c = 10$		$c = 100$	
		block-boot.	logn.	block-boot.	logn.	block-boot.	logn.
n_w	h	%	%	%	%	%	%
	0.1000	63.00	80.20	63.60	80.00	63.20	79.80
	0.1414	65.20	81.60	65.60	81.20	64.80	81.00
	0.2000	65.00	82.80	65.20	82.40	65.60	81.60
	0.2236	63.60	68.20	63.80	66.80	63.20	66.60
	0.2828	65.00	83.40	65.40	81.40	64.40	83.00
	0.3000	68.00	86.00	67.80	85.40	67.60	85.00
	0.3162	64.80	70.80	64.20	70.80	64.60	70.60
	0.3606	69.60	71.00	68.40	70.60	67.20	71.00
	0.4000	68.60	86.40	68.40	88.20	68.20	88.20
	0.4123	65.00	72.60	67.60	73.60	67.00	77.20
	0.4243	68.60	83.40	70.20	83.20	71.00	82.40
	0.4472	67.80	73.60	70.00	74.80	68.00	76.80
	0.5000	66.00	66.20	68.40	69.80	67.80	71.60
	0.5099	69.80	72.80	70.60	76.40	71.00	77.80
8	0.5385	72.00	77.60	74.40	80.00	73.40	80.80
	0.5657	70.80	87.20	71.20	88.80	72.00	89.40
	0.5831	69.40	81.00	72.40	82.80	72.00	83.80
	0.6000	75.80	85.80	77.00	89.20	77.60	89.40
	0.6083	74.80	76.40	76.60	78.20	76.40	82.20
	0.6325	73.00	78.00	76.00	82.20	75.40	82.80
	0.6403	70.00	81.20	72.40	85.00	74.20	86.80
	0.6708	74.40	83.20	77.40	86.60	79.60	87.60
	0.7000	82.60	88.80	82.60	90.20	84.40	91.00
	0.7071	75.40	71.60	76.60	77.80	78.20	80.60
	0.7211	73.60	82.60	76.00	86.20	77.60	88.00
	0.7280	82.20	81.60	83.60	85.80	86.00	88.40
	0.7616	82.40	83.20	85.60	88.40	85.80	88.20
	0.7810	77.20	82.20	80.60	89.00	81.20	90.20
	0.8062	82.60	92.40	87.20	94.40	88.00	94.40
	0.8485	84.40	72.80	85.80	79.80	86.80	80.20
	0.8602	81.60	85.40	85.60	87.20	85.40	90.20
	0.9220	85.60	92.00	87.80	94.20	89.40	95.20
	0.9899	91.00	88.00	92.00	89.80	91.20	89.20

Table 5.1. Coverage percentages for the block-bootstrap 95% confidence intervals of a simulated process with a exponential semi-variogram. The process is simulated on a regular grid with $n = 10 \times 10$ locations. The window size is $n_w = 8$. The coverage percentage for the confidence intervals of the variogram as calculated from a log-normal distribution of its classical estimator is shown in this table.

		$\gamma(h) = 4 - \exp(-ch)$					
		$c = 3$		$c = 10$		$c = 100$	
		block-boot.	logn.	block-boot.	logn.	block-boot.	logn.
n_w	h	%	%	%	%	%	%
5	0.1000	86.80	81.00	86.00	79.80	84.20	79.60
	0.1414	89.40	80.80	89.80	80.00	87.00	79.60
	0.2000	88.60	83.00	89.00	82.00	86.80	81.40
	0.2236	89.00	67.80	88.20	66.00	87.20	65.60
	0.2828	91.20	83.60	90.40	81.40	87.80	82.20
	0.3000	93.80	84.40	93.40	84.60	89.00	84.00
	0.3162	93.40	70.40	92.60	71.00	90.40	70.20
	0.3606	92.00	71.20	92.20	71.40	93.00	70.60
	0.4000	97.20	86.40	97.60	88.00	95.60	87.40
	0.4123	96.40	71.60	97.20	74.00	95.20	76.00
	0.4243	96.40	83.80	96.20	83.20	97.40	82.20
	0.4472	97.00	74.20	97.00	74.20	97.40	76.40
	0.5000	98.60	67.00	98.20	70.00	98.20	71.40
	0.5657	99.00	74.20	99.10	76.60	99.00	77.80
3	0.1000	80.00	81.40	79.80	79.60	79.40	79.40
	0.1414	82.40	81.20	80.40	80.40	78.60	79.60
	0.2000	85.60	83.40	85.40	81.60	84.80	81.40
	0.2236	86.40	68.20	85.80	65.80	84.80	65.60
	0.2828	83.00	83.60	83.60	81.40	85.00	81.80

Table 5.2. Coverage percentage for the block-bootstrap 95% confidence intervals of a simulated process with a exponential semi-variogram. The process is simulated on a regular grid with $n = 10 \times 10$ locations. Different windows sizes n_w are considered. The coverage percentage for the confidence intervals of the variogram as calculated from a log-normal distribution of its classical estimator is shown in this table.

		$\gamma_Y(h) = 16\gamma_Z(h) - 2\gamma_Z^2(h)$		$\gamma_Z(h) = 4 - \exp(-ch)$
		$c = 3$	$c = 10$	$c = 100$
n_w	h	%	%	%
	0.1000	2.00	2.00	2.20
	0.1414	1.60	1.00	1.40
	0.2000	2.60	2.20	2.80
	0.2236	2.00	1.60	2.20
	0.2828	3.60	2.60	3.20
	0.3000	5.60	4.00	3.40
	0.3162	3.40	2.80	2.80
	0.3606	5.60	3.80	3.60
	0.4000	5.40	4.80	4.40
	0.4123	4.20	4.20	3.20
	0.4243	6.40	5.20	4.40
	0.4472	6.00	5.00	3.80
	0.5000	6.80	6.20	5.00
	0.5099	8.60	6.60	6.40
	0.5385	9.40	7.40	8.40
8	0.5657	7.00	7.00	6.00
	0.5831	8.20	7.00	6.80
	0.6000	17.80	18.60	20.20
	0.6083	14.60	15.00	15.40
	0.6325	13.60	13.60	13.40
	0.6403	12.20	10.80	12.00
	0.6708	18.20	18.00	19.20
	0.7000	40.80	41.80	44.00
	0.7071	27.40	30.40	30.00
	0.7211	22.60	23.60	24.20
	0.7280	36.40	38.60	38.80
	0.7616	41.00	44.60	44.80
	0.7810	32.20	33.60	34.00
	0.8062	46.00	47.60	50.40
	0.8485	46.80	50.80	52.60
	0.8602	54.60	60.60	62.20
	0.9220	68.80	74.60	76.00
	0.9899	88.80	91.40	90.80

Table 5.3. Coverage percentage for the block-bootstrap 95% confidence intervals of simulated processes $\boldsymbol{Y}$ obtained from quadratic and exponential transformations to normal processes $\boldsymbol{Z}$ with a exponential semi-variogram γ_Z as indicated on the top. The processes are simulated on a regular grid with $n = 10 \times 10$ locations. The window size is $n_w = 8$.

		$\gamma_Y(h) = 16\gamma_Z(h) - 2\gamma_Z^2(h)$		$\gamma_Z(h) = 4 - \exp(-ch)$
		$c = 3$	$c = 10$	$c = 100$
n_w	h	%	%	%
5	0.1000	16.20	17.80	18.80
	0.1414	19.20	19.60	19.60
	0.2000	24.60	25.20	25.80
	0.2236	25.40	27.00	26.00
	0.2828	35.00	33.00	35.20
	0.3000	33.20	34.20	34.40
	0.3162	35.00	35.80	36.40
	0.3606	35.80	37.00	36.20
	0.4000	57.60	59.20	60.20
	0.4123	53.00	53.80	54.80
	0.4243	52.60	52.60	53.20
	0.4472	61.20	64.60	66.00
	0.5000	72.20	73.80	77.00
	0.5657	89.60	93.00	92.80
3	0.1000	55.80	58.60	59.40
	0.1414	65.40	66.60	69.00
	0.2000	74.40	75.40	77.60
	0.2236	73.80	77.60	79.60
	0.2828	81.20	92.00	93.40

Table 5.4. Coverage percentage for the block-bootstrap 95% confidence intervals of simulated processes $\boldsymbol{Y}$ obtained from a quadratic transformation to normal processes $\boldsymbol{Z}$ with a exponential semi-variogram γ_Z as indicated on the top. The processes are simulated on a regular grid with $n = 10 \times 10$ locations. Different window sizes n_w are considered

Chapter 6

Reflections

Nothing is concluded.

Everything is the begining of a new way.

6.1 Introduction

At this stage, some general reflections can be made. Most of these thoughts will be about open problems in the context of the subjects which were tackled here.

The aim of this chapter is to motivate everyone who reads these pages to think about these problems.

6.2 Generalized linear models

1. The Generalized Additive Models approach provided a very attractive tool to modelling the density of mackerel egg as a spatial process whose mean at each location is a non-linear function of some covariates like latitude, longitude, bottom depth and distance to two hundred metres contour.

2. On the other hand, and as explained in chapter 2, the lack of a distributional theory for the estimators of these models led to the use of an "anology" with the generalized linear models as proposed by Hastie & Tibshirani in the book which is up the moment the only instrument available to learn their music. Indeed, as in the exercise presented here, under the assumption of independent errors, the bootstrap techniques can be applied to introduce some confidence in this analogy.

3. In the context of these models, a crucial question to be answered is: What is the distribution of the residuals, even under the simplest assumption of non-random covariates and independent normal distribution for the errors of the model?

4. After this question can be answered, a cluster of new open problems will be available to tackle, such, for instance, what is the variance-covariance structure of the underlying process? Can the tools developed here, in the context of the variogram, be extended to these models?

5. In fact, in the scenario of spatial statistics, the necessity of "good" (well developed) tools to fit data to models where a spatial random process is a non-linear function of covariates is indisputable.

6.3 Linear models

1. In fact, because the statistical theory of linear models is old and vastly developed, the starting point to tackle spatial problems where the data are fitted to a linear model is settled on better known territory. Consequently, the challenge of searching for new tools to check independence or constant variogram as considered in chapter 4, or to build confidence intervals for it as analyzed in chapter 5, may be less frustrating.

2. Nevertheless, also in the context of spatial processes (gaussian, second order stationary, isotropic) which are linear functions of other variables (covariates), the point reached in the way followed to look for tools to tackle problems dealing with the study of the covariance structure (variogram) of the underlying process is not very far from the begining.

3. Despite the work developed here for checking constant variogram and, consequently, independence of the variables in the underlying process, the distributions of estimators of the variogram or means to do inference about its shape and values at a specific distance between points is a very important problem to think about. In this work, the latter problem was considered in the approach proposed in chapter 5. Indeed, the block-bootstrap methodology was used to generate these intervals under the assumption of normal distribution of the underlying process.

4. The formal test and the graphical tool of the reference bands, could be also interesting ideas to extend to checking the null hypothesis that the variogram has a particular shape or belongs to a particular family γ_o as defined by Cressie (1991), say, exponential, gaussian, spherical, etc.

If the statistical hypotheses are:

$$\begin{aligned} H_o : \gamma(h_i) &= \gamma_o(h_i), \quad i = 1, \ldots, n \\ H_1 : \gamma(h_i) &= \text{smooth function of } \mathrm{h}_i, \quad i = 1 \ldots, n \end{aligned}$$

then, the residuals $r(h_i) = \hat{\gamma}(h_i) - \gamma(h_i), \quad \mathrm{i} = 1, \ldots, n$ from the fitted variogram under the null hypothesis can be used to check these hypothese. In fact, under the null hypothesis, the variables

$$s_i = |r(h_i)|^{1/2}, \quad \mathrm{i} = 1, \ldots, n$$

are expected to be approximately constant, and, consequently, the smooth versions

$$\tilde{s} = \sum_{j=1}^{n} w_{ij} s_j, \quad i = 1, \ldots, n$$

are expected to be approximatly equal to the average $\bar{s}$.

Otherwise, if H_o is not true, the variables $s_i = |r(h_i)|^{1/2}, \quad i = 1, \ldots, n$, where $r(h_i), \quad i = 1, \ldots n$ are the residuals from the data to the fitted variogram under H_o, and the corresponding smooth curve $\tilde{s}_i = \sum_{j=1}^{n} w_{ij} s_j, \quad j = 1, \ldots, n$ are expected to be more "different" from a constant than those corresponding to H_o true.

In other words, the statistic,

$$T = \frac{\sum_{i=1}^{n} (s_i - \bar{s})^2 - \sum_{i=1}^{n} (s_i - \tilde{s}_i)^2}{\sum_{i=1}^{n} (s_i - \tilde{s}_i)^2}$$

is expected to be "small" under H_o and "big" under H_1. Then, the test statistic presented in chapter 4, and the associated methods of inference, may be extended to these hypotheses.

5. If the assumption of normality is relaxed (non-gaussian geostatistics), the attempt to build confidence intervals for the variogram through block-bootstrap or any other appropriate technique or modification of block-bootstrap is still an open problem as showed in chapter 5.

6.4 Resampling

1. Not many problems in spatial statistics can reach the happy (or unhappy) end without a simulation of a spatial process. Hence, a good method to simulate a spatial process (gaussian, stationary, isotropic) is an invaluable tool for every statistician, consultant or researcher, who is involved in this area of statistics. Unfortunately, experience tells us that all the algorithms proposed until now fail in one or another model for the covariance function. For example, the algorithm proposed by Wood and Chan (1995) is a very interesting approach from the point of view of its efficiency in time and memory storage of a computer. Nevertheless this algorithm, or more precisely its implemenatation in the computer, cannot handle some combinations of covariance functions and grid sizes. Consequently, another algorithm to simulate a spatial process or a modification of the some methods proposed previously would be very welcome to our library.

2. The bootstrap environment also offers fresh herbs for those who like the taste of these approachs. In fact, when they can be applied, they offer a big help because of their generality. The modification of the block-bootstrap is a very useful tool when an appropriate block-size (under a fixed grid-size) can be found succesfully. Some other problems

also come across in its implementation, mostly for cases of long-range dependence and particularly, as shown in chapter 5, when the underlying distribution is not normal. But still, if it is normal there are some works in the literature concerned with the pour behaviour of the block-bootstrap (see, for instance, Lahiri (1983))

3. The behaviour of the block-bootstrap under the assumption of stationarity is another problem to improve when the block-bootstrap is used. There is also in this context some work, due to Politis and Romano (1990,1994), as an effort to give an answer to the best "design" for block-bootstrap in order to keep the underlying stationarity. There are some theoretical answers in this approach but also some unsolved problems which makes its current application poor.

4. The block-bootstrap method is designed for a regular grid. An extension for a non-regular set of points would be a very useful tool. In fact, most of the sets of data of the real world have this characteristic.

References

Abril, J.C. (1987) The approximate densities of some quadratic forms of stationary random variables *Journal of Time Series Analysis*, **8**, 249-259.

Aitchinson, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*,**50**, 901-908

Aitchison, J & Brown, J.A.C. (1957) The Lognormal Distribution. Cambridge University Press: Cambridge.

Ali, M.M. (1979) Analysis of stationary spatial-temporal processes: Estimation and prediction. *Biometrika*, **66**, 513-518.

Ali, M.M. (1984) An approximation to the null distribution and power of the Durbin-Watson distribution. *Biometrika*, **71**, 253-261.

Azzalini,A., Bowman,A.W. (1993) On the use of Nonparametric Regression for Checking Linear Relationships, J.R.Statist. Soc. B, 55, 549-557.

Baczkowski, A.J. & Mardia,K.V. (1987) Approximate lognormality of the sample semi-variogram under a gaussian process. *Commun. Statist. - Simula.*, **16** (2), 571-585.

Besawa, I.V., Mallik,A.K., McCormick,W.P & Taylor,R.L. (1989). Bootstrapping explosive autoregressive processes. *Ann. Statist.*, **17** (4), 1749-1486.

Bickel, P. & Freedman D. (1981). Some asymptotic theory for the bootstrap.

Ann. Statist., **9**, 1196-1217.

Borchers, D.L, Buckland, S.T., Priede, I.G. & Ahmadi, S. (1994) Improving the Precision of the Daily egg Production Method using Generalized Additive models. Accepted for publication by Can. J. Fish & Aquat. Sci.

Borgman, L, Taheri,M & Hagan,R (1984). Three dimensional frequency domain simulations of Geological variables. Geostatistics for Natural Resources Characterization Part I, eds G.Verly,M. David, A.G.Journel, and A.Marechal, Boston: D.Reidel, 517-541.

Bose, A. (1981). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.*, **16** (4), 1709-1722.

Bowman, A. & Young, S. (1996). Graphical Comparison of Nonparametric Curves. *Applied Statistics*, to appear.

Brillinger, D. (1981). Time Series, Data Analysis and Theory Holden-Day Series in Time Series Analysis, San Francisco.

Brooker, P. (1985). Two-dimensional Simulations by Tuerning Bands. *Mathematical Geology*, **17**, 81-91

Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. Unpublished.

Buja, A, Hastie, T.J. & Tibshirani, R.J. (1989) Linear Smoothers and Additive Models (with discussion). *Annals of Statistics*, **17**, 453-555

Chan, G. & Wood, A.T.A. (1979). An Algorithm for Simulating Stationary Gaussian Random Fields. *Algorithm Section of Applied Statistics*, to appear.

Christakos, G. (1987). Stochastic Simulations of Spatially Correlated Geo-Processes. *Mathematical Geology*, **19**, 807-831

Carlstein, E. (1986). The use of subseries values for estimationg the variance

of a general statistic form a stationary sequence. *The Annals of Statistics*, **14**, 1171-1179.

Carroll, R.J., (1982), Adapting for heteroscedasticity in linear models, The Annals of Statistics, vol 10, N 4, 1224-1233.

Carroll,R.J. and Ruppert,D. (1988), Transformation and Weighting in Regression, Chapman & Hall, New York-London.

Cleveland,W.S. (1979) Robust locally weighted regression and smoothing scatterplots. J.Am.Statist.Ass., 74, 829-836.

Cleveland,W.S. (1993); Visualizing Data; Hobart Press, Summit, New Jersey.

Cook,R.D. and Weisberg,S. (1983) Diagnostics for heteroscedasticity in regression, *Biometrika*, **70**, 1-10.

Cramer, H. (1967). Stationary and related stochastic processes: some properties and their applications. Wiley, New York.

Cramer, H. (1970). Random variables and probablity distributions. Cambridge University Press, Cambridge

Cressie, N. (1985) Fitting Variogram Models by Weighted Least Squares. *Mathematical Geology*, **17**, 563-586.

Cressie, N. (1991). Statistics for Spatial Data Wiley: New York.

Cressie, N. & Hawkins, D.M. (1980) Robust estimation of the variogram. *Mathematical Geology*, **12**, 115-125.

Davidian, M & Carroll, R.J. (1987) Variance function estimation. *JASA*, **82**, 1079-1091.

Davis, B.M. & Borgman, L.E. (1979). Some exact sampling distributions for variogram estimators. *Mathematical Geology*, **11**, 643-653.

Davis, B.M. & Borgman, L.E. (1982). A note on the asymptotic distribution

of the sample variogram. *Mathematical Geology*, **14**, 189-193.

Davis, M.W. (1986(a)). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, **19**, 91-98

Davis, M.W. (1986(b)). Generating Large Stochastic Simulations - The Matrix Polynomial Approximation Method. *Mathematical Geology*, **19**, 99-127

Davis, M.W., Hagan,R. and Borgman,L.E. (1981). A program for the finite Fourier transform simulations of realizations from one-dimensional random function with known covariance. *Computers and Geosciences*, **7**, 199-206

Davison, A. & Hinkley, R.J.(1996); Bootstrap methods and their applications. Chapman & Hall, Cambridge.

Diggle, P. (1979). On parameter estimation of goodness-of-fit testing for spatial point processes. *Biometrics*, **35**, 87-101.

Efron, B. (1979). Bootstrap methods: Another look at the jacknife. *The Annals of Statistics*, **7**, 1-26.

Efron, B. & Tibshirani, R.J. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy (with discussion). *Statist. Sci.*, **1**, 54-77.

Efron,B. Tibshirani,R.J.(1993); An Introduction to the Bootstrap, Chapman & Hall.

Eubank, R.L. (1988) Spline smoothing and nonparametric regression. Dekker,A.G., Marcel.

Eubank, R.L. & Spiegelman, C.H. (1990). Testing the goodness-of-fit of a linear model via nonparametric techniques. *J.Amer.Statist.Assoc.* **85**, 387–392.

Fisher, R.A. (1947). The analysis of covariance method for the relation between a part and the whole. *Biometrics* **3**, 65–68.

biblioitem Freedman, D. (1981). Bootstraping regression models. *Ann. Statist.* **9**, 1218-1228.

Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics*, **12** (3), 827-842.

Fuller, W.A. (1976). Introduction to Statistical Time Series. Wiley, New York.

Fuller,A.W., Rao,J.N.K. , Estimation for a linear regression model with unknown diagonal covariance matrix, *Ann. Statis.* **6**, 1149-1158.

Gasser, T., Kneip, A. and Kohler, W. (1991). A flexible and fast method for automatic smoothing. *J.Amer.Statistics.Assoc.* **86**, 643-652.

Gihman, I.I.& Skorohod, A.V. (1974). The Theory of Stochastic Processes. Springer-Verlag, Berling, Heidelberg, New York.

Gradshteyn, I.S. (1965). Table of integrals, series and products; Academic Press, New York.

Green, P.J. & Silverman, B.W. (1994) Nonparametric regression and generalized linear models. Chapman and Hall: London, New York, Tokyo, Melbourne, Madras.

Graybill, F.A. (1983). Matrices with applications in Statistics. Wadsworth International Group: California.

Haining, R. (1979). Trend-surface models with regional and local scales of variation with an application to aerial survey data. *Technometrics*, **29**, 461-469.

Haining, R.& Griffith, D.A. (1983). Simulating Two-dimensional Autocorrelated Surfaces. *Geographical Analysis*, **15**, 247-255.

Hall, P. (1985) Resampling a coverage pattern. *Stochastic Processes and*

their Applications, **20**, 231-246.

Hall,P. (1993); The Bootstrap and Edgworth Expansion; Srpinger-Verlag. New York.

Hall, P. & Jing, B. (1994) On sample re-use methods for dependent data. *Mathematical Geology*, **17**, 563-586.

Hastie, T.J. & Tibshirani, R.J. (1990) Generalized Additive Models. Chapman and Hall: London, New York, Tokyo, Melbourne, Madras.

Hawkins, R.& Cressie, N. (1984). Robust Kriging - A proposal. *Mathematical Geology*, **16**, 3-18.

Johnson,N.I. & Kotz,S.; Continuous Univariate Distributions, vol,2; Houghton Mifflin Company, Boston.

Journel, A.G. (1974) Simulation conditionelle de gisements miniers-theorie et pratique: These de Docteur Ingenieur, Université de Nancy.

Jowett, G.H. (1955). Sampling properties of local statistics in stationary stochastic series. *Biometrika*, **42**, 160-169.

Journel, A.G. & Juijbregts,C.T. (1978) Mining geostatistics. Academic Press, New York.

King, E., Hart, J.D. & Wehrly, T.E. (1992). Testing the equality of two regression curves using linear smoothers. *Statistics & Probability Letters* **12**, 2

Koenker, R.& Basset, G. (1981). Robust test for heteroscedasticity based on regression quantiles. *Econometrika* **50**, 43-61.

Kunsch, S. (1989). The jacknife and the bootstrap for general stat ionary observations. *Annals of Statistics* **17**, 1217-1241.

Lahiri, S. (1991). Second order optimality of stationary bootstrap. *Statisitics & Probability Letters* **11**, 335-341

Lahiri, S. (1993). On the moving block bootstrap under long range dependence. *Statisitics & Probability Letters* **18**, 405-413

Léger, C. Politis, D. N. & Romano, J.P. (1992). Bootstrap technology and applications. *Technometrics*, **34**, 378-398.

Mathai, A.M. - Provost, Serge B. (1992) Quadratic Forms in Random Variables, Theory and Applications- Statistics: Texbooks and Monographs (volume 126), Marcel Dekker, Inc, New York-Basel-Hong Kong (1992)- CIP 91-40165, vol. 126, 120-142.

Matheron, G. (1971). The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique de Fontaineblau **5**.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advance Applied Probability* **5**, 439-468

Maritz, J.S. (1980). Distribution-Free Statistical Methods. Chapman & Hall, London.

Mantoglue, A. (1987). Digital Simulations of Multivariate Two- and Three-Dimensional Stiochastic Processes With a Spectral Turning Bands Method. *Mathematical Geology*, **19**, 129-149.

McCullagh,P. & Nelder,J.A. (1989) Generalised Linear Models (2nd edition). Chapman & Hall: London.

Mejía, J. & Rodriguez Iturbe,I (1974). On the synthesis of random fields from the spectrum: An application to the generation of hydrologic spatial processes. *Water Resources Researches*, **10**, 705-711.

Müller,H.G., Stadmüller,U. (1987) Estimation of Heteroscedasticity in Regression Analysis, *Annals Statist.*, **15** (2), 610-625

Müller,H.G., Stadmüller,U. (1993) On variance function estimation with quadratic forms, *Journal of Statistical Planning and Inference* **35**, 213-231.

Müller,H.G.& Zhao, P.H. (1995) On a semiparametric variance function model and a test for heteroscedasticity, *Ann. Statist.* (to appear).

Omre, H. (1984). The variogram and its estimation. *Geostatistics for natural resources characterization*, Part 1, G.Verly, M.David, A.Journel, and Marechal, eds. Reidel, Dordrecht, 107-125.

Pennington, M. (1983). Efficient Estimators of Abundance, for Fish and Plankton Surveys. *Biometrics*, **39**, 281-286

Politis, D. N. & Romano, J.P. (1992a). A circular block-resampling procedure for stationary data. *Exploring the limit of Bootstrap.*, R.LePage and L.Billard eds., 263-270, Wiley, New York.

Politis, D. N. & Romano, J.P. (1992b). A general resampling scheme for triangular arrays of α-mixing random variables with the application to the problem of spectral density estimation. *Ann. Statist.*, **20**, 1985-2007.

Politis, D. N. & Romano, J.P. (1993a). Estimationg the distribution of a studentized statistic by subsampling, *Bull. Int.Statist.Inst.,49th Session*, **2**, 301-328.

Politis, D. N. & Romano, J.P. (1994). The stationary bootstrap. *J.Amer.Statist.Assoc.*, **89**, 1303-1313.

Politis, D. N. & Romano, J.P. (1995). A general theory for large sample confidence regions based on subsamples under minimal assumptions. *Ann.of Statist.*, **23**, 1313-1335.

Pope, J. G. & Woolner, L. (1985). Improving the sample design of the western mackerel egg survey. ICES CM 1985/D:11.

Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomisation approach. *J.Amer.Statist.Assoc.* **85**, 132–138.

Richardson, J. (1991). Testing for no effect when estimating a smooth function by nonparametric regression: a randomisation approach. *J.Amer.Statist.Assoc.* **85**, 132–138.

Robinson, G.K. (1990). A role for variograms. *Australian Journal of Staitistics*, **32**(3), 327-335.

Ruppert, D., Sheather, S.J. & Wand, M.P. (1996). An effective bandwidth selector for local least squares regression. *J.Amer.Statist.Assoc.*, to appear.

Sabourin, R. (1976). Application of two methods for the interpretation of the undelying variogram. *Advanced Geostatistics in the Mining Industry, M.Guarascio, M.David, and C. Huijbregts, eds. Reidel, Dordrecht*, 101-109.

Shinozuka, M & Jan,C.M. (1972). Digital simulation of random processes and its applications. *Journal of sound and vibration*, **25** 111-128.

Silverman, K. (1982). Kernel Density estimation using the fast Fourier transform. *Appl. Statist.*, **31** , 93-99.

Silverman, B.W. & Bernard, W.. (1982) Density estimation for statistics and data analysis. Chapman and Hall: London, New York, Tokyo, Melbourne, Madras.

Silverman, B.W. (1984) Spline smoothing: the equivalent variable kernel method. *Ann. Statist.*, **12**, 898-916.

Silverman, B.W. & Young,G.A. (1987) The bootstrap: To smooth or not to smooth. *Biometrika*, **74**, 469-479.

Sing, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, **9** (6), 1187-1195.

Simonoff,J.S.& Tsai,C.L. (1994) Use of Modified Profile Likelihood for Improved Tests of Constancy of Variance in Regression. *Appl. Statist.*, **43**, N_0 2, 357-

Sneddon, I.N.(1972). The use of integral transform. McGraw-Hill, New York.

Swanepoel, J.W.H. & van Wyk,J.W.J. (1986). The bootstrap applied to

spectral density function estimation. *Biometrika*, **73** 135-142.

Venables, W.N. & Ripley, B.D. (1994). Modern Applied Statistics with S-Plus. Springer-Verlag: New York.

Wand, M.P. (1994) Fast computation of multivariate kernel estimators. Submitted to publication.

Wand, M. & Jones, M.C. (1995). introduction to kernel smoothing. Chapman & Hall: London.

Wand, M.P. & Schucany,W.R. (1990) Gaussian-based kernels. *Canad. J. Statist.* **18** 197-204.

Weisberg S. (1985) Applied Linear Regression, Second Edition, John Wiley & sons, New York,Chichester,Brisbane,Toronto,Singapore.

Wilson,E.B. & Hilferty,M.M. (1931). The distribution of chi-square. *Proc.Nat.Acad.Sci.* **17**, 684-8. Wiley: New York.

Wood, T.A. & Chan, G. (1994) Simulation of Stationary Gaussian Processes in $[0,1]^d$ *Journal of Computational and Graphical Statistics*, **3** (4), 409-432.