Generalized Linear Models for Large Dependent Data Sets

Steven M. Bate

Department of Statistical Science
University College London

Thesis submitted for the degree of Doctor of Philosophy
University of London

August 2004

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

Abstract

Generalized linear models (GLMs) were originally used to build regression models for independent responses. In recent years, however, effort has focused on extending the original GLM theory to enable it to be applied to data which exhibit dependence in the responses. This thesis focuses on some specific extensions of the GLM theory for dependent responses.

A new hypothesis testing technique is proposed for the application of GLMs to cluster dependent data. The test is based on an adjustment to the 'independence' likelihood ratio test, which allows for the within cluster dependence. The performance of the new test, in comparison to established techniques, is explored.

The application of the generalized estimating equations (GEE) methodology to model space-time data is also investigated. The approach allows for the temporal dependence via the covariates and models the spatial dependence using techniques from geostatistics.

The application area of climatology has been used to motivate much of the work undertaken. A key attribute of climate data sets, in addition to exhibiting dependence both spatially and temporally, is that they are typically large in size, often running into millions of observations. Therefore, throughout the thesis, particular attention has focused on computational issues, to enable analysis to be undertaken in a feasible time frame. For example, we investigate the use of the GEE one-step estimator in situations where the application of the full algorithm is impractical.

The final chapter of this thesis presents a climate case study. This involves wind speeds over northwestern Europe, which we analyse using the techniques developed.

Acknowledgements

I would like to express my gratitude to my supervisors, Dr. Richard E. Chandler and Professor Valerie S. Isham, for their continuous support, guidance, encouragement, numerous helpful comments and stimulating discussions throughout the undertaking of this work. They have always been there for me and I am deeply indebted to both.

I wish to thank my fellow students and staff members at the Department of Statistical Science, University College London, for their help and friendship in recent years. Unfortunately, there are too many of you to mention individually.

I am also grateful to the Engineering and Physical Science Research Council for providing financial support. This work would not have been possible without their backing.

Finally, I would like to thank my family and close friends back home for their constant support. Most of all, I would like to thank my wife, Clair, who has been a tower of strength and very understanding, especially during the periods of extensive work. A special thanks also goes to Sam.

Contents

List of figures 8

List of tables 11

1 Introduction 12

1.1 Overview 12

1.2 Outline of thesis 14

1.3 Notation and abbreviations 15

2 Univariate generalized linear models 17

2.1 Introduction 17

2.2 Exponential family 18

2.3 Likelihood inference 19

2.3.1 Maximum likelihood estimation 20

2.3.2 Tests of hypotheses 22

2.4 Parameter estimation 24

2.4.1 Estimation of the regression parameters 24

2.4.2 Estimation of the dispersion parameter 26

2.5 Hypothesis testing 26

2.5.1 Likelihood ratio test 27

2.5.2 Wald test 27

2.5.3 Score test 28

2.6 Analysis of residuals 29

2.7 Extensions 29
2.7.1 Distributions close to exponential family form 30
2.7.2 Quasi-likelihood 31
2.8 Summary 33

3 Generalized linear models for cluster correlated data 34
3.1 Generalized estimating equations 35
3.1.1 Introduction 35
3.1.2 Independence estimating equations 37
3.1.3 Generalized estimating equations 40
3.1.4 Parameter estimation 41
3.1.5 Hypothesis testing 44
3.1.6 Longitudinal example 47
3.2 Alternatives to conventional generalized estimating equations . . . 49
3.2.1 Second order generalized estimating equations 51
3.2.2 One-step generalized estimating equations 52
3.2.3 Generalized linear mixed models 53

4 Hypothesis testing for generalized linear models applied to clustered data 56
4.1 Introduction 56
4.2 New adjusted likelihood ratio test 57
4.2.1 Motivation 57
4.2.2 General theory and derivation of test 58
4.3 Geometry of the new test 62
4.3.1 Single parameter case 62
4.3.2 Two parameter case 63
4.4 Simulation studies 66
4.4.1 Performance assessment criteria 66
4.4.2 Binary simulations 69

4.4.3 Gamma simulations 76
4.5 Summary 82

5 Generalized estimating equations for large space-time data sets 84
5.1 Alternative approaches for space-time data 85
5.2 Overview of new generalized estimating equations approach for space-time data 87
5.3 Modelling temporal structure 88
5.3.1 Autoregressive approach 88
5.3.2 Checking the autoregressive representation 90
5.3.3 Alternative approaches 91
5.4 Modelling spatial structure 92
5.4.1 Isotropic structures 92
5.4.2 Anisotropic structures 95
5.5 The one-step estimator 97
5.6 Summary 98

6 Climate case study 99
6.1 Introduction 100
6.1.1 The data set 100
6.1.2 Aim of study 101
6.2 Preliminary analysis 101
6.2.1 Site specific properties 101
6.2.2 Seasonality 103
6.2.3 Trend 106
6.3 Generalized linear modelling approach 111
6.3.1 Gamma model 113
6.3.2 Weibull model 126
6.4 Generalized estimating equations approach 130
6.4.1 Allowing for temporal dependence 131

6.4.2 Allowing for spatial dependence 131
6.4.3 Results 137
6.4.4 The one-step estimator 138
6.5 Comparison of approaches 141
6.6 Summary 144

7 Conclusions and further work **147**

Bibliography **150**

APPENDICES **157**

A Gamma GLM coefficient estimates **157**

B Spatial GEE coefficient estimates **163**

List of Figures

Chapter 4: Hypothesis testing for generalized linear models applied to clustered data

4.1 Geometry of new method when testing a single parameter. 64

4.2 Comparison of new method with Rotnitzky and Jewell's method when testing a single parameter. 64

4.3 Comparison of new method with Rotnitzky and Jewell's method when testing two parameters. 67

4.4 Theoretical power curves for testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. 69

4.5 Simulated p-value cdf for each of the four competing tests under $H_0 : \beta_2 = 0$. 72

4.6 Power curves for each of the four competing tests under $H_0 : \beta_2 = 0$, $\alpha = 0.05$. 73

4.7 Power curves for the new method and Rotnitzky and Jewell's method under $H_0 : \beta_1 = \beta_2 = 0$, $\alpha = 0.05$. 75

4.8 Power curves for each of the four competing tests under $H_0 : \beta_1 = \beta_2 = 0$, $\alpha = 0.05$. 75

4.9 Power curves for new method and robust Wald test under $H_0 : \beta_2 = 0$, $\alpha = 0.05$, x_1 and x_2 correlated. 77

4.10 Power curves for each of the four competing tests under $H_0 : \beta_2 = 0$, $\alpha = 0.05$, x_1 and x_2 correlated. 77

4.11 Power curves for each of the four competing tests under $H_0 : \beta_1 = \beta_2 = 0$, $\alpha = 0.05$, x_1 and x_2 correlated. 78

4.12 Simulated p-value cdf for each of the four competing tests under false $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. 83
4.13 Simulated p-value cdf for each of the four competing tests under true $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. 83

Chapter 6: Climate case study

6.1 Map of study area, with NCEP grid overlaid. 100
6.2 Histograms of DMWS values over all locations and three specific locations, 1958-1998. 102
6.3 Mean DMWS values (ms^{-1}) over NCEP grid, 1958–1998. 104
6.4 Standard deviations of DMWS values (ms^{-1}) over NCEP grid, 1958–1998. 104
6.5 Coefficients of variation of DMWS values over NCEP grid, 1958–1998. 105
6.6 Maximum DMWS values (ms^{-1}) over NCEP grid, 1958–1998. . . 105
6.7 Monthly mean DMWS values (ms^{-1}) over NCEP grid, 1958–1998. 107
6.8 Monthly standard deviations of DMWS values (ms^{-1}) over NCEP grid, 1958–1998. 108
6.9 Monthly coefficients of variation of DMWS values over NCEP grid, 1958–1998. 109
6.10 Monthly maximum DMWS values (ms^{-1}) over NCEP grid, 1958–1998. 110
6.11 Daily mean and standard deviation DMWS for most north-westerly and most south-easterly locations. 111
6.12 Time series of annual mean DMWS values (ms^{-1}) taken over all locations, 1958-1998. 112
6.13 Decadal trends in annual mean DMWS values at each NCEP grid point, 1958–1998. 112
6.14 Average seasonal and regional variation in DMWS, according to the gamma GLM. 119

6.15 Neighbourhood structure for autoregressive covariates. 121
6.16 Plot of monthly Pearson residuals for gamma GLM. 124
6.17 Plot of annual Pearson residuals for gamma GLM. 124
6.18 Normal quantile plot of residuals from gamma GLM. 125
6.19 Comparison of the shape of the fitted gamma and Weibull distributions. 129
6.20 Plot of sample autocorrelation function of Pearson residuals from gamma GLM. 131
6.21 Powered exponential correlation function fit to the pairwise correlations in Pearson residuals from the gamma GLM. 134
6.22 Spherical correlation function fit to the pairwise correlations in Pearson residuals from the gamma GLM. 134
6.23 Matérn correlation function fits to the pairwise correlations in Pearson residuals from the gamma GLM. 135
6.24 Powered exponential correlation function fits to the pairwise correlations in Pearson residuals from the gamma GLM, for directions 0°, 45°, 90° and 135°. 136
6.25 Powered exponential correlation function fits to the pairwise correlations in Pearson residuals from the gamma GLM, all four directions 136
6.26 Comparison of coefficient estimates obtained from the full GEE algorithm and the one step GEE. 142
6.27 Comparison of t-values obtained from the full GEE algorithm and the one step GEE. 142
6.28 Comparison of coefficient estimates obtained from the gamma GLM and the spatial GEE. 145
6.29 Comparison of t-values obtained from gamma GLM and spatial GEE. 145
6.30 Plot of difference in gamma GLM and spatial GEE fitted values. . 146

List of Tables

Chapter 1: Introduction

1.1 Table of abbreviations 16

Chapter 3: Generalized linear models for cluster correlated data

3.1 Results from fitting a Poisson GLM, overdispersed Poisson quasi-model and Poisson IEE to the epilepsy data. 49

3.2 Results from fitting three separate GEEs to the epilepsy data. . . 50

3.3 Estimated GEE working correlation matrices for the epilepsy data. 50

Chapter 6: Climate case study

6.1 Summary of external effects considered. 116

6.2 Neighbourhood weighting scheme adopted for autoregressive covariates. 121

6.3 Grid search for the weight w_{SS} in the autoregressive neighbourhood.122

6.4 Calculated test statistics for testing the NHT effects. 126

6.5 Performance of the correlation functions in terms of sum of squared errors. 137

6.6 Convergence properties of the full spatial GEE algorithm. 140

6.7 Mahalanobis distance for the spatial GEE 140

6.8 Comparison of gamma GLM and spatial GEE in terms of R^2. . . 143

6.9 Comparison of gamma GLM and spatial GEE in terms of QIC. . 144

Chapter 1

Introduction

1.1 Overview

Generalized linear models (GLMs) were originally proposed by Nelder and Wedderburn (1972) in an attempt to unify a vast array of statistical models. Subsequently, GLMs have been applied extensively to build regression models within a wide range of application areas. One of the main, and potentially restrictive, characteristics of the univariate GLM is that responses are assumed independent, given the covariates in the model. In recent years, however, a vast amount of material has been published on extending univariate GLMs so that they may be applied to data sets that exhibit dependence in the response. Examples of such techniques include generalized estimating equations (GEEs) and generalized linear mixed models (GLMMs). This subject of extending GLMs for dependent response data is the focus of this thesis.

The application area of climatology has been used to motivate much of the work undertaken. Data sets taken from this field typically possess two key attributes, which are highly influential in determining the nature of the statistical analysis undertaken. Firstly, data sets exhibit dependence both spatially and tem-

porally, hence the need for methods which allow for dependence. And secondly, data sets are typically large in size, often running into millions of observations. Methods applied therefore need to be computationally efficient and for this reason particular attention is given to computational issues throughout this thesis.

When modelling climate data sets, typically interest lies in explaining how the various components of the climate system interact or affect one another. For example, if rainfall at a network of sites is the variable under consideration, the effect of factors such as spatial location and seasonality upon rainfall is likely to be of interest. Generalized linear models provide a natural framework within which to explore such relationships. In the above example, rainfall amounts would be classed as the response variable and factors such as seasonality and spatial location would be potential explanatory variables. The first work to be published on the application of GLMs to model climate variables was Coe and Stern (1982), and later Stern and Coe (1984), who considered the modelling of rainfall data. In recent years, these ideas have been extended extensively by Chandler and Wheater (2002) and Yan et al. (2002). The main challenge faced when applying GLMs to climate data is allowing for the dependence in the response variable. This dependence usually exists both temporally and spatially, as a result of parallel time series being collected at a network of neighbouring sites. For the single site problem, accounting for the temporal dependence is possible via covariates. However, for the multi-site case, allowing for the spatial dependence at neighbouring sites, in addition to the temporal dependence, is problematic. Within this thesis we explore some methods of extending the GLM framework to allow for this dependence.

Broadly speaking, when applying GLMs to dependent response data, one of two approaches can be adopted. Either the GLM fitting routine can be extended to allow parameters to be estimated while accounting for the dependence or, alternatively, a univariate GLM can be fitted and then the subsequent inference

adjusted to allow for the dependence within the responses. Within this thesis both approaches are considered. A new hypothesis testing technique is proposed for the application of univariate GLMs to the modelling of cluster dependent data. This technique uses theory embedded within the independence estimating equations (IEE) approach to adjust the 'independence' likelihood ratio test to allow for the dependence. Within a climate context, this technique can be applied to allow for inter-site dependence. This technique also has computational benefits since the efficient GLM fitting routine does not need to be adjusted to obtain the parameter estimates. Instead, only a small extra step is required to make inference on these parameters. As an alternative to the above approach, we consider the application of GEEs to model climate data. This approach allows for the temporal dependence via the covariates and uses ideas from geostatistics to suggest appropriate spatial dependence structures.

Even though many of the techniques presented have been motivated within a climate context, they are applicable to a much wider field. For example, the work undertaken on hypothesis testing can be applied to any cluster based data set, such as a longitudinal study. Also the work undertaken on the application of the GEE methodology can be applied to other space-time settings.

1.2 Outline of thesis

This thesis is organised in the following manner. Chapter 2 provides the necessary background required for subsequent chapters. Univariate GLM theory is introduced, along with a few possible extensions. Chapter 3 focuses on extending univariate GLMs so that dependent responses can be modelled. This material predominantly focuses on the GEE approach, but GLMMs are also considered. In Chapter 4 we propose a new hypothesis testing technique for GLMs, when applied to cluster correlated data. Using simulations we investigate the per-

formance of the new method in comparison with other established techniques. Chapter 5 moves on to consider how GEEs can be applied to model climate or, more generally, space-time data. Computational aspects are considered throughout. In Chapter 6 the analysis of a specific climate data set, involving wind speeds over Northern Europe, is undertaken. This enables us to demonstrate the GLM methodology which has been developed. Finally, Chapter 7 summarises and concludes the work undertaken, and considers a few possible extensions.

1.3 Notation and abbreviations

Realizations of random variables are denoted by lower case italic letters, while the random variables themselves are represented by the corresponding upper case italic letter. For example, y represents a realization of the random variable Y. Generally speaking, unknown parameters will be represented by lower case Greek letters and their corresponding estimators by the hat notation. For example, an estimator of the parameter α is represented by $\hat{\alpha}$. Matrices and vectors are represented in bold font and scalars in normal font. With respect to matrices and vectors, the notation $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$. For square matrices, the notations $\mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}$ denote the inverse and matrix square root of the matrix $\mathbf{A}$ respectively.

Abbreviations that are commonly used throughout the thesis are given in Table 1.1.

GLM	Generalized linear model
IEE	Independence estimating equations
GEE	Generalized estimating equations
GLMM	Generalized linear mixed model
pdf	Probability density function
cdf	Cumulative density function
mle	Maximum likelihood estimator
df	Degrees of freedom
AR	Autoregressive
ACF	Autocorrelation function

Table 1.1: Abbreviations commonly used throughout thesis.

Chapter 2

Univariate generalized linear models

Within this chapter we review univariate generalized linear models (GLMs), which underlie many of the techniques discussed throughout this thesis. Due to limitations on space, only the material needed for reference in subsequent chapters has been included. For a fuller account of univariate GLM theory the reader is referred to Dobson (2002) for an introductory account, and McCullagh and Nelder (1989) for a more advanced treatment.

2.1 Introduction

Univariate GLMs are regression models which enable us to explore the relationship between a single response variable and several explanatory variables. They extend the classical linear model from normal response data to the much wider class of response data belonging to the exponential family of distributions. A transformation of the linear predictor is also accommodated, for example, to ensure that all fitted values lie within the permitted range of the response. As with

classical linear normal models, a fundamental assumption of univariate GLMs is that individual responses are independent given the explanatory variables in the model.

More formally, a univariate GLM can be defined as follows. Let $\mathbf{Y}^T = (Y_1, \ldots, Y_n)$ denote a response vector of random variables, whose n independent elements (given the explanatory variables in the model) share the same form of parametric distribution from the exponential family. Corresponding to each Y_i are the values $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{ip})$ of p explanatory variables, which can be combined across the n observations to form a $n \times p$ design matrix $\mathbf{X}$, whose ith row is given by $\mathbf{x}_i^T$. The linear predictor $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ is related to the mean of $\mathbf{Y}$ through the model equation

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \tag{2.1}$$

where $\boldsymbol{\mu} = \mathrm{E}(\mathbf{Y})$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters, and $g(\cdot)$ is a monotonic and differentiable link function. By $g(\boldsymbol{\mu})$ we mean the $n \times 1$ vector whose ith element is $g(\mu_i)$.

2.2 Exponential family

When fitting a GLM we assume that the distribution of the response variables Y_i belongs to the exponential family of distributions, such that the probability density function (pdf) or probability mass function (pmf) can be written in the following standard form

$$f(y; \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right], \tag{2.2}$$

where θ is called the natural parameter, ϕ is called the dispersion parameter, and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are specific functions corresponding to the type of exponential family.

The natural parameter θ is a function of the mean, $\theta = \theta(\mu)$. Moreover,

applying the standard results $\mathrm{E}(\partial\ell/\partial\theta) = 0$ and $\mathrm{E}[(\partial\ell/\partial\theta)^2] = -\mathrm{E}(\partial^2\ell/\partial\theta^2)$, where $\ell = \ell(\theta, \phi; y) = \log f(y; \theta, \phi)$ denotes the log-likelihood function, it follows that $\mu = \mathrm{E}(Y) = \partial b(\theta)/\partial\theta$ and $\mathrm{var}(Y) = a(\phi)\partial^2 b(\theta)/\partial\theta^2$. The variance of Y is often written $\mathrm{var}(Y) = a(\phi)v(\mu)$, where $v(\mu) = \partial\mu/\partial\theta$ denotes the variance function, to emphasis the dependence of the variance on the mean.

Well-known distributions belonging to the exponential family include the normal, Poisson, binomial and gamma. For example, consider the gamma pdf given by

$$f(y; \mu, \nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1} \exp\left(-\frac{\nu y}{\mu}\right), \quad 0 \leq y < \infty, \quad \mu, \nu > 0, \tag{2.3}$$

where

$$\mathrm{E}(Y) = \mu \text{ and } \mathrm{var}(Y) = \frac{\mu^2}{\nu}.$$

Rewriting (2.3) in the form

$$f(y; \mu, \nu) = \exp\left[\left(-\frac{y}{\mu} - \log\mu\right)\nu - \log\Gamma(\nu) + \nu\log\nu + (\nu - 1)\log y\right]$$

it can be seen that the gamma density is a member of the exponential family as defined by (2.2), with $\theta = -1/\mu$, $b(\theta) = -\log(-\theta)$, $a(\phi) = 1/\nu$ and $c(y, \phi) = -\log\Gamma(\nu) + \nu\log\nu + (\nu - 1)\log y$. Hence, $\mathrm{E}(Y) = b'(\theta) = -\theta^{-1} = \mu$ and $\mathrm{var}(Y) = a(\phi)b''(\theta) = \theta^{-2}/\nu = \mu^2/\nu$, as above.

2.3 Likelihood inference

For GLMs, the parameter vector $\boldsymbol{\beta}$ is estimated by maximum likelihood. Before proceeding to discuss the details of this, however, the more general topic of likelihood inference will be reviewed as this material is used extensively throughout the thesis.

2.3.1 Maximum likelihood estimation

Let $\mathbf{y} = (y_1, \ldots, y_n)$ be a realisation of the n independent random variables $\mathbf{Y} = (Y_1, \ldots, Y_n)$ whose pdf $f(y; \boldsymbol{\theta})$ depends on a $p \times 1$ vector of parameters $\boldsymbol{\theta}$. The likelihood function, which is a function of the unknown $\boldsymbol{\theta}$ given the data $\mathbf{y}$, is defined as

$$L(\boldsymbol{\theta}; \mathbf{y}) = f(\mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}), \tag{2.4}$$

where the product follows from the fact that the y_i's are independent. Typically, it is more convenient to work with the log-likelihood function, which is given by

$$\log L(\boldsymbol{\theta}; \mathbf{y}) = \ell(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\theta}). \tag{2.5}$$

The maximum likelihood estimator (mle) of $\boldsymbol{\theta}$, is the value which maximizes the log-likelihood function. More formally, the mle of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, is defined by

$$\ell(\hat{\boldsymbol{\theta}}; \mathbf{y}) \geq \ell(\boldsymbol{\theta}; \mathbf{y}) \quad \forall \boldsymbol{\theta} \in \Omega,$$

where Ω is the parameter space. The mle is obtained by solving the p equations

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{2.6}$$

simultaneously. The mle $\hat{\boldsymbol{\theta}}$ is a consistent estimator of the 'true' value $\boldsymbol{\theta}$ under mild regularity conditions (see below) (Cox and Hinkley, 1974).

Define the score $\mathbf{U}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}; \mathbf{y}) / \partial \boldsymbol{\theta}$. This is a random vector, which evaluated at the true $\boldsymbol{\theta}$ has the following properties: $\mathrm{E}[\mathbf{U}(\boldsymbol{\theta})] = \mathbf{0}$ and $\mathrm{var}[\mathbf{U}(\boldsymbol{\theta})] = \mathrm{E}[\mathbf{U}(\boldsymbol{\theta})\mathbf{U}^T(\boldsymbol{\theta})] = \mathbf{I}(\boldsymbol{\theta})$, where $\mathbf{I}(\boldsymbol{\theta})$ is known as the Fisher information matrix. Under mild regularity conditions, relating to the ability to interchange the order of differentiation and integration, the information matrix is also given by

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right) = -\mathrm{E}\left[\mathbf{U}'(\boldsymbol{\theta}) \right], \tag{2.7}$$

where $\mathbf{U}'(\boldsymbol{\theta})$ is the derivative of the score vector $\mathbf{U}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Asymptotically, the score vector is distributed as

$$\mathbf{U}(\boldsymbol{\theta}) \overset{a}{\sim} N(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})), \tag{2.8}$$

where $\overset{a}{\sim}$ denotes 'is asymptotically distributed as'. Result (2.8) is obtained by application of the central limit theorem, which holds because the score vector and information are sums of n independent contributions.

A first order Taylor series approximation of $\mathbf{U}(\hat{\boldsymbol{\theta}})$ about the true parameter value $\boldsymbol{\theta}$ is given by

$$\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{U}(\boldsymbol{\theta}) + \mathbf{U}'(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(1). \tag{2.9}$$

Now, given that $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, as $\hat{\boldsymbol{\theta}}$ satisfies (2.6), we obtain

$$\mathbf{U}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(1),$$

where the observed information matrix $-\mathbf{U}'(\boldsymbol{\theta})$ has been replaced by its expected value (which is asymptotically equivalent). Thus,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathbf{I}^{-1}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta}) + O_p(n^{-1}). \tag{2.10}$$

Using (2.10), and since $\mathrm{E}[\mathbf{U}(\boldsymbol{\theta})] = \mathbf{0}$ and $\mathrm{var}[\mathbf{U}(\boldsymbol{\theta})] = \mathbf{I}(\boldsymbol{\theta})$, we obtain $\mathrm{E}(\hat{\boldsymbol{\theta}}) \overset{a}{=} \boldsymbol{\theta}$ and $\mathrm{var}(\hat{\boldsymbol{\theta}}) \overset{a}{=} \mathbf{I}^{-1}(\boldsymbol{\theta})\mathrm{var}[\mathbf{U}(\boldsymbol{\theta})]\mathbf{I}^{-1}(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta})$, where $\overset{a}{=}$ denotes 'is asymptotically equal to'. Moreover, application of the weak law of large numbers and the central limit theorem as detailed in Cox and Hinkley (1974, p. 294), yields the following asymptotic result

$$\hat{\boldsymbol{\theta}} \overset{a}{\sim} N(\boldsymbol{\theta}, \mathbf{I}^{-1}(\boldsymbol{\theta})). \tag{2.11}$$

For (2.11) to hold several regularity conditions must be satisfied, these include: a) the true parameter $\boldsymbol{\theta}$ must be interior to the parameter space Ω, b) the parameter vector $\boldsymbol{\theta}$ is identifiable, i.e. if $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ then $f(y;\boldsymbol{\theta}) \neq f(y;\boldsymbol{\theta}^*)$, c) the log-likelihood function is three times differentiable with respect to $\boldsymbol{\theta}$ and the third derivative is bounded. For further details on the above regularity conditions see Cox and Hinkley (1974, p. 281).

2.3.2 Tests of hypotheses

Having calculated the mle $\hat{\boldsymbol{\theta}}$, consider testing the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, for fixed value $\boldsymbol{\theta}_0$. By definition, the mle $\hat{\boldsymbol{\theta}}$ maximizes the likelihood function, and therefore a sensible quantity to consider for testing H_0 is the following ratio of likelihoods

$$\lambda = \frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})}. \tag{2.12}$$

Naturally, the quantity λ is bounded by 0 and 1, with values closer to one supporting H_0.

The Neyman-Pearson lemma (Casella and Berger, 2002, p. 388) states that the best test of size α that can be constructed for testing H_0 is based on (2.12). We now show how, under H_0, $W_L = -2\log\lambda$ follows an asymptotic chi-squared distribution with p degrees of freedom (d.f.), where $p = \dim(\hat{\boldsymbol{\theta}})$. This result forms the basis for testing H_0.

Consider the second order Taylor series approximation of $\ell(\boldsymbol{\theta}_0)$ about the consistent mle $\hat{\boldsymbol{\theta}}$

$$\ell(\boldsymbol{\theta}_0) = \ell(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})^T\mathbf{U}(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\mathbf{U}'(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(n^{-1/2}).$$

Since $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, and again replacing $\mathbf{U}'(\hat{\boldsymbol{\theta}})$ by its expected value we obtain

$$\ell(\boldsymbol{\theta}_0) = \ell(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(n^{-1/2}).$$

Finally, replacing $\mathbf{I}(\hat{\boldsymbol{\theta}})$ by its asymptotic equivalent $\mathbf{I}(\boldsymbol{\theta}_0)$, and rearranging we obtain

$$-2\left[\ell(\boldsymbol{\theta}_0) - \ell(\hat{\boldsymbol{\theta}})\right] = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\mathbf{I}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(n^{-1/2}). \tag{2.13}$$

Due to the asymptotic normality of $\hat{\boldsymbol{\theta}}$ given by (2.11), under H_0, the right hand side of (2.13) follows an asymptotic chi-squared distribution with p d.f. It follows that the quantity on the left hand side of (2.13), $-2\log\lambda$, also has a limiting chi-squared distribution with p d.f. Thus, the test statistic

$$W_L = -2\left[\ell(\boldsymbol{\theta}_0) - \ell(\hat{\boldsymbol{\theta}})\right]$$

can be used to test H_0. This test is known as the likelihood ratio test.

An alternative to the likelihood ratio test, for testing H_0, can be obtained by making direct use of the asymptotic normality of the mle. Under H_0, the Wald test statistic

$$W_T = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

follows an asymptotic chi-squared distribution with p d.f., by (2.11). In the case of a scalar θ, the signed square root of W_T has a standard normal distribution.

A further alternative for testing H_0 can be obtained by making use of the asymptotic normality of the score vector. Under H_0, the score test statistic

$$W_S = \mathbf{U}(\boldsymbol{\theta}_0)^T \mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{U}(\boldsymbol{\theta}_0)$$

follows an asymptotic chi-squared distribution with p d.f., by (2.8).

The Wald and score tests are quadratic approximations to the likelihood ratio test. Asymptotically, all three tests are equivalent (Cox and Hinkley, 1974, Section 9.3). The asymptotic equivalence of the Wald and likelihood ratio test statistics can be seen from (2.13), while to show the asymptotic equivalence of the score and likelihood ratio test statistics we make use of (2.13) and (2.10). Substituting the right hand side of (2.10) into the right hand side of (2.13) we obtain $W_L = \mathbf{U}(\boldsymbol{\theta}_0)^T \mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{U}(\boldsymbol{\theta}_0) + O_p(n^{-1/2}) = W_S + O_p(n^{-1/2})$. For large samples, therefore, all three tests should provide similar results. For small and medium sized samples, however, the methods can differ (Cox and Hinkley, 1974, Section 9.3). Likelihood ratio tests and score tests are invariant under parameter transformation, while Wald tests do not possess this appealing property (Rotnitzky and Jewell, 1990). For a more extensive comparison of the three tests see Buse (1982).

2.4 Parameter estimation

Having reviewed the general topic of likelihood inference in the previous section, we now focus on the more specific case of parameter estimation for GLMs. When fitting a GLM, two quantities must be estimated, these being the regression parameter vector $\boldsymbol{\beta}$ and the dispersion parameter ϕ (assuming it is unknown). In this section, the estimation of each of these is considered in turn.

2.4.1 Estimation of the regression parameters

The parameter vector $\boldsymbol{\beta}$, consisting of p elements, can be estimated by the method of maximum likelihood. The log-likelihood function for n independent responses Y_i, of the form (2.2), is given by

$$\ell(\theta; \mathbf{y}) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi). \tag{2.14}$$

To maximize this log-likelihood function with respect to $\boldsymbol{\beta}$, the p score equations $\mathbf{U}(\boldsymbol{\beta}) = \partial\ell/\partial\boldsymbol{\beta} = \mathbf{0}$ are solved for $\boldsymbol{\beta}$, where the jth element of the score vector is given by

$$\frac{\partial\ell}{\partial\beta_j} = U_j = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\text{var}(Y_i)} \left(\frac{\partial\mu_i}{\partial\eta_i}\right), \quad j = 1, \ldots, p. \tag{2.15}$$

The derivation of (2.15) involves the application of the chain rule $\partial\ell_i/\partial\beta_j = (\partial\ell_i/\partial\theta_i)\,(\partial\theta_i/\partial\mu_i)(\partial\mu_i/\partial\eta_i)(\partial\eta_i/\partial\beta_j)$, see McCullagh and Nelder (1989, Chapter 2). In matrix notation the p score equations are given by

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{X}^T\mathbf{D}(\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta})\mathbf{S}(\boldsymbol{\beta}) = \mathbf{0}, \tag{2.16}$$

where $\mathbf{D}(\boldsymbol{\beta})$ is an $n \times n$ diagonal matrix with elements $(\partial\mu_1/\partial\eta_1, \ldots, \partial\mu_n/\partial\eta_n)$, $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ is an $n \times n$ diagonal variance-covariance matrix for $\mathbf{Y}$ with elements $\{\text{var}(Y_1), \ldots, \text{var}(Y_n)\}$ and $\mathbf{S}(\boldsymbol{\beta})$ is an $n \times 1$ vector with elements $\{y_1 - \mu_1(\boldsymbol{\beta}), \ldots, y_n - \mu_n(\boldsymbol{\beta})\}^T$. These score equations, consisting of p simultaneous equations in p unknowns, are often non-linear and thus a numerical technique is usually adopted to solve them.

The most commonly used technique is known as the method of scoring, which is a modification of the Newton-Raphson method, where the matrix of second derivatives are replaced by their expected value. The $(t+1)$th iteration of the method of scoring is given by

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \left[\mathbf{I}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)\right]^{-1} \mathbf{U}\left(\hat{\boldsymbol{\beta}}^{(t)}\right), \tag{2.17}$$

where $\mathbf{I}(\boldsymbol{\beta})$ denotes the Fisher information matrix given by

$$\mathbf{I}(\boldsymbol{\beta}) = \text{cov}\left[\mathbf{U}(\boldsymbol{\beta})\right] = \text{E}\left[\mathbf{U}(\boldsymbol{\beta})\mathbf{U}(\boldsymbol{\beta})^T\right] = -\text{E}\left[\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right] = \mathbf{X}^T\mathbf{W}\mathbf{X}, \tag{2.18}$$

and $\mathbf{W}$ is an $n \times n$ diagonal matrix whose ith diagonal element is

$$W_{ii} = \frac{1}{\text{var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

Equation (2.17) can be re-expressed in iterative weighted least squares (IWLS) form

$$(\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})\hat{\boldsymbol{\beta}}^{(t+1)} = \mathbf{X}^T\mathbf{W}^{(t)}\mathbf{z}^{(t)}, \tag{2.19}$$

where $\mathbf{W}^{(t)}$ and $\mathbf{z}^{(t)}$ are evaluated at $\boldsymbol{\beta}^{(t)}$, and $\mathbf{z}^{(t)}$ is an $n \times 1$ vector with ith element

$$z_i^{(t)} = \eta_i^{(t)} + \left(y_i - \mu_i^{(t)}\right)\left(\left.\frac{\partial \eta}{\partial \mu}\right|_{\mu=\mu_i^{(t)}}\right).$$

For details of expressing (2.17) in IWLS form see McCullagh and Nelder (1989, Chapter 2).

The fitting process is iterative as both W_{ii} and z_i depend on the current fitted value μ_i, which in turn depends on the current estimate of $\boldsymbol{\beta}$ through the link function $g(\mu_i) = \eta_i$. Therefore, to obtain $\hat{\boldsymbol{\beta}}$ we begin by calculating $\mathbf{W}^{(0)}$ and $\mathbf{z}^{(0)}$, using an initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$. $\mathbf{W}^{(0)}$ and $\mathbf{z}^{(0)}$ are then substituted into (2.19), which is solved for $\hat{\boldsymbol{\beta}}^{(1)}$. This new estimate of $\boldsymbol{\beta}$ is then used to obtain new estimates $\mathbf{W}^{(1)}$ and $\mathbf{z}^{(1)}$, which in turn are used in (2.19) to solve for $\hat{\boldsymbol{\beta}}^{(2)}$. This iterative procedure continues until the successive estimates $\hat{\boldsymbol{\beta}}^{(t)}$ and $\hat{\boldsymbol{\beta}}^{(t+1)}$ agree to within a specified tolerance.

As the appropriate regularity conditions stated in Section 2.3.1 are satisfied for GLMs, from (2.11) we obtain the result that $\hat{\boldsymbol{\beta}}$ has an asymptotic multivariate normal distribution with mean equal to the true parameter value $\boldsymbol{\beta}$ and variance-covariance matrix given by $\mathbf{I}^{-1}(\boldsymbol{\beta})$. Thus,

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} N(\boldsymbol{\beta}, \mathbf{I}^{-1}(\boldsymbol{\beta})), \tag{2.20}$$

where the matrix $\mathbf{I}(\boldsymbol{\beta})$ can be consistently estimated by $\mathbf{I}(\hat{\boldsymbol{\beta}})$.

2.4.2 Estimation of the dispersion parameter

The dispersion parameter ϕ remains constant across observations and must be estimated when unknown. The following method of moments estimator provides a consistent estimate of ϕ (Fahrmeir and Tutz, 2001, p. 47)

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{v(\mu_i)}. \tag{2.21}$$

The above estimator is generally advocated, in favour of its maximum likelihood (ML) counterpart, as it is considered more robust. For example, for gamma response variables, if a zero observation is recorded, the ML approach does not yield an estimate of ϕ (McCullagh and Nelder, 1989, p. 295-296).

By inspecting (2.15) we see that ϕ only enters into the score equations via var(Y), and therefore ϕ does not affect the estimation of $\hat{\boldsymbol{\beta}}$. The covariance of $\hat{\boldsymbol{\beta}}$ is a function of ϕ, however, but provided ϕ is replaced by the consistent estimate $\hat{\phi}$, then the asymptotic result (2.20) remains valid for $\hat{\boldsymbol{\beta}}$.

2.5 Hypothesis testing

For GLMs, hypothesis testing is usually undertaken via one of the three general tests outlined in Section 2.3.2, these being the likelihood ratio, Wald and score

tests. Within this section we outline the specific form these tests take for GLMs.

For the remainder of this section, assume that the parameter vector $\boldsymbol{\beta}$ is partitioned into two subvectors, such that $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have p_1 and p_2 elements respectively. The null hypothesis is of the form

$$H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0, \tag{2.22}$$

where $\boldsymbol{\beta}_2^0$ denotes a specific value of $\boldsymbol{\beta}_2$.

2.5.1 Likelihood ratio test

The likelihood ratio test, also known as a scaled deviance test within the GLM context, is given by

$$W_L = -2\left[\ell\left(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2^0\right) - \ell\left(\hat{\boldsymbol{\beta}}\right)\right], \tag{2.23}$$

where $\hat{\boldsymbol{\beta}}_1$ is the restricted mle of $\boldsymbol{\beta}_1$ under H_0, and $\hat{\boldsymbol{\beta}}$ is the unrestricted mle of $\boldsymbol{\beta}$. It can be shown, using arguments similar to those used in Section 2.3.2, that W_L has in large samples an approximate χ^2 distribution with p_2 d.f., under H_0 (Cox and Hinkley, 1974, p. 321-323). Thus, H_0 is rejected at significance level α if $W_L > \chi^2_{p_2,\alpha}$, where $\chi^2_{p_2,\alpha}$ is the upper $100\alpha\%$ point of a χ^2 distribution with p_2 d.f.

If the dispersion parameter ϕ is unknown and has to be estimated, it is common to eliminate it from the calculated test statistic by replacing the above χ^2 test with an F-test (Venables and Ripley, 1994). This corresponds to the conventional F-test for model comparison in normal theory linear models.

2.5.2 Wald test

The Wald statistic is given by

$$W_T = (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)^T \left[\mathbf{C}_{22}(\hat{\boldsymbol{\beta}})\right]^{-1} (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0), \tag{2.24}$$

where $\mathbf{C}(\boldsymbol{\beta}) = \mathbf{I}^{-1}(\boldsymbol{\beta})$, and $\mathbf{C}_{22}(\hat{\boldsymbol{\beta}})$ is the $p_2 \times p_2$ submatrix of $\mathbf{C}(\hat{\boldsymbol{\beta}})$ corresponding to $\boldsymbol{\beta}_2$. It can be shown, using arguments similar to those used in Section 2.3.2, that W_T has in large samples an approximate χ^2 distribution with p_2 d.f., under H_0. Thus, H_0 is rejected at significance level α if $W_T > \chi^2_{p_2,\alpha}$, where $\chi^2_{p_2,\alpha}$ is the upper $100\alpha\%$ point of a χ^2 distribution with p_2 d.f.

When the null hypothesis (2.22) takes the specific form $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, the Wald statistic of (2.24) simplifies to $W_T = \hat{\boldsymbol{\beta}}_2^T \left[\mathbf{C}_{22}(\hat{\boldsymbol{\beta}})\right]^{-1} \hat{\boldsymbol{\beta}}_2$.

2.5.3 Score test

The score test statistic is given by

$$W_S = \mathbf{U}\left(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2^0\right)^T \mathbf{I}^{-1}\left(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2^0\right) \mathbf{U}\left(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2^0\right), \tag{2.25}$$

where $\hat{\boldsymbol{\beta}}_1$ is the restricted mle of $\boldsymbol{\beta}_1$ under H_0. It can be shown, using arguments similar to those used in Section 2.3.2, that W_S has in large samples an approximate χ^2 distribution with p_2 d.f., under H_0. Thus, H_0 is rejected at significance level α if $W_S > \chi^2_{p_2,\alpha}$, where $\chi^2_{p_2,\alpha}$ is the upper $100\alpha\%$ point of a χ^2 distribution with p_2 d.f.

When the null hypothesis (2.22) takes the specific form $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, the score test statistic of (2.25) simplifies to $W_S = \mathbf{U}_{22}(\hat{\boldsymbol{\beta}}_1)^T \mathbf{I}_{22}^{-1}(\hat{\boldsymbol{\beta}}_1) \mathbf{U}_{22}(\hat{\boldsymbol{\beta}}_1)$, where $\mathbf{U}_{22}(\hat{\boldsymbol{\beta}}_1)$ is the $p_2 \times 1$ subvector of $\mathbf{U}(\hat{\boldsymbol{\beta}}_1)$ and $\mathbf{I}_{22}^{-1}(\hat{\boldsymbol{\beta}}_1)$ is the $p_2 \times p_2$ submatrix of $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_1)$, both corresponding to $\boldsymbol{\beta}_2$.

The score test has a computational advantage over both the Wald and likelihood ratio test, since only the restricted model needs to be fitted. In turn, the Wald test has a computational advantage over the likelihood ratio test in that while the likelihood ratio test requires both the restricted and unrestricted models to be fitted, the Wald test only requires the fitting of the unrestricted model.

2.6 Analysis of residuals

An examination of the model residuals can provide a valuable insight into how well specific aspects of the model are performing. For GLMs there are several different types of residual that are commonly used. For example, the Pearson residual for observation y_i is defined as

$$r_i^{(P)} = \frac{y_i - \hat{\mu}_i}{(v(\hat{\mu}_i))^{1/2}}, \tag{2.26}$$

which can be viewed as a standardised residual since the raw residual is divided through by the square root of the variance function. If the fitted model is correct then the Pearson residuals follow distributions with mean 0 and variance ϕ. In later chapters, we see how Pearson residuals play a central role in estimating the correlation in the $Y_i's$, when the restrictive assumption of independent $Y_i's$ is relaxed.

Other types of residual frequently used for GLMs include Anscombe and deviance residuals. For a detailed account of residuals in GLMs, see Pierce and Schafer (1986).

2.7 Extensions

Two specific ways in which the univariate GLM theory can be extended to a wider class of response data are examined in this section. In Section 2.7.1 we consider how GLMs can be fitted for distributions close in form to the exponential family, such as the Weibull distribution. This material will be used in Chapter 6 when a Weibull GLM is applied to model a climate data set. Then, in Section 2.7.2 we consider how the GLM theory can provide an estimation technique when only the first two moments of $\mathbf{Y}$ are specified. This technique is developed further in Chapter 3, when generalized estimating equations are used to model dependent response data.

2.7.1 Distributions close to exponential family form

The GLM fitting routine can be extended to accommodate some distributions which are not members of the exponential family, as defined by (2.2), but are close in form. More specifically, if the distribution of the response Y is not of exponential family form, but the distribution of $V = \tau(Y; \boldsymbol{\alpha})$ is of the required form, then the GLM algorithm can be extended for Y, where τ is a known monotonic function and $\boldsymbol{\alpha}$ is a parameter vector. The basic idea is to iterate back and forth between fitting a GLM to V for fixed $\boldsymbol{\alpha}$, and estimating $\boldsymbol{\alpha}$ given the fitted values for V. For example, consider the Weibull distribution, with pdf given by

$$f(y; \lambda, \alpha) = \frac{\alpha}{\lambda} y^{\alpha-1} \exp\left(-\frac{y^\alpha}{\lambda}\right), \quad 0 \leq y < \infty, \quad \lambda, \alpha > 0, \tag{2.27}$$

where

$$\mathrm{E}(Y) = \lambda^{1/\alpha} \Gamma\left(1 + \frac{1}{\alpha}\right) \text{ and } \mathrm{Var}(Y) = \lambda^{2/\alpha} \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right].$$

Rewriting (2.27) in the form

$$f(y; \lambda, \alpha) = \exp\left[\log \alpha + (\alpha - 1) \log y - \log \lambda - \frac{y^\alpha}{\lambda}\right]$$

it can be seen that the Weibull density can not be expressed in the standard exponential family form of (2.2). However, for fixed shape parameter α, the density of $V = Y^\alpha$ is in standard exponential family form, since V follows an exponential distribution. Thus, a GLM can be fitted to Weibull response data via the fitting routine for the exponential distribution, with an extra iteration level added for the estimation of α.

Under the iterative scheme outlined above, the natural way to fit the Weibull model is as follows:

1. Fix α to an initial value. One possible choice would be $\alpha = 1$, corresponding to an exponential distribution.

2. With α fixed, estimate the parameter vector $\boldsymbol{\beta}$ via the fitting of an exponential GLM for the transformed response variable $\mathbf{V} = \mathbf{Y}^{\alpha}$. The exponential model takes the form

$$\log \boldsymbol{\lambda} = \boldsymbol{\eta},$$

where $\mathrm{E}(\mathbf{V}) = \boldsymbol{\lambda}$.

3. Re-estimate α using the method of maximum likelihood, by solving the equation

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log y_i - \frac{1}{\lambda_i} y_i^{\alpha} \log y_i = 0$$

numerically, where the λ_i's are the exponential fitted values from step 2.

4. Repeat steps 2-3 until successive estimates of α converge to the specified tolerance.

An alternative implementation of the Weibull model is given by Aitkin and Clayton (1980), who consider the modelling of censored survival data. They implement the Weibull model by fitting a series of Poisson log-linear models to the censoring indicator, for fixed shape parameter α. The algorithm oscillates between the fitting of a Poisson model and the estimation of the shape parameter until convergence is achieved. The presentation of their approach for Weibull response data, is standard within the GLM literature, as the Weibull distribution is usually presented within a survival analysis context. However, while Aitkin and Clayton's method can still be implemented for non-survival data, the method detailed above has the advantage that it eliminates the redundant censoring indicator.

2.7.2 Quasi-likelihood

Wedderburn (1974) identified that the GLM score equations given in (2.16) only depend on the first two moments of $\mathbf{Y}$. Therefore, in situations where it is not

possible to specify the full distribution of $\mathbf{Y}$, but it is reasonable to specify the first two moments, the GLM score equations can be used as a set of estimating equations to obtain an estimate of the regression parameters $\boldsymbol{\beta}$. Any relationship between the mean and variance of $\mathbf{Y}$ can be assumed, with the choice no longer being restricted to correspond to a specific distribution. This method is known as quasi-likelihood estimation, because integrating the score equations with respect to $\boldsymbol{\beta}$ does not necessarily produce a true log-likelihood function, instead it corresponds to what Wedderburn (1974) called a 'quasi-likelihood'.

The implementation of quasi-likelihood estimation proceeds along similar lines to that of univariate GLMs. The individual elements of the response vector $\mathbf{Y}$ are assumed independent given the covariates in the model, and the mean and variance of $\mathbf{Y}$ are specified as $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu}$ and $\mathrm{var}(\mathbf{Y}) = \phi v(\boldsymbol{\mu})$. For the regression equation $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$, the GLM score equations (2.16), referred to as the quasi-score equations, are solved for $\boldsymbol{\beta}$ using (2.19). McCullagh (1983) proved that the quasi-likelihood estimate of $\boldsymbol{\beta}$ follows an asymptotic multivariate normal distribution with mean equal to the true parameter value $\boldsymbol{\beta}$ and variance given by the inverse of (2.18). Hypothesis testing for quasi-likelihood estimates can be undertaken via Wald and quasi-score tests.

Quasi-likelihood estimation provides a simple method of accounting for overdispersion in count data. The natural choice of distribution for the modelling of count data is the Poisson, which possesses the property $\mathrm{E}(Y) = \mathrm{var}(Y) = \mu$. Many data sets, however, exhibit greater variability than that permitted by the Poisson distribution, and for these situations quasi-likelihood estimation can be implemented with $\mathrm{E}(Y) = \mu$ and $\mathrm{var}(Y) = \phi\mu$. The inclusion of the additional dispersion parameter ϕ, enables the variance to be inflated by an appropriate factor. The quasi-likelihood estimate obtained for $\boldsymbol{\beta}$ is identical to that obtained under the Poisson GLM, while the standard errors of $\hat{\boldsymbol{\beta}}$ are inflated by the factor $\sqrt{\phi}$. This theory is implemented in Section 3.1.6 when a longitudinal data set is

analysed.

2.8 Summary

Within this chapter we have introduced univariate generalized linear models for independent response data, since they provide a suitable starting point for the more complicated dependent data case, which is the subject of the remainder of this thesis. A discussion of standard likelihood theory has been undertaken, as it plays a prominent role in subsequent chapters. Hypothesis testing theory has also been reviewed, since this provides the foundation for the work undertaken in Chapter 4 on hypothesis testing for cluster correlated data. Finally, in Section 2.7, two ways in which the univariate GLM theory can be extended were outlined. This provides a preview of some extensions of the basic GLM theory that will be presented in later chapters. In particular, the ideas of quasi-likelihood estimation are extended further in Chapter 3 to account for dependent data.

Chapter 3

Generalized linear models for cluster correlated data

The term 'cluster correlated data' refers to data with multiple observations on the same sampling unit or cluster. A key feature of such data is that observations taken from the same cluster are correlated. An example is provided by a longitudinal study (Diggle et al., 2002) which involves a set of individuals, on each of whom a response variable and set of covariates are measured on several occasions over time. Responses for different individuals are assumed independent, whereas responses measured on the same individual are correlated. Thus, each individual forms a cluster of correlated measurements. Clustered data arise naturally in many other settings, such as in family studies and dentistry.

Regression analysis for cluster correlated data is complicated by the correlation present within cluster. The direct application of the theory of Chapter 2 is inappropriate, as it assumes that all responses are independent given the covariates. Cluster correlated data therefore necessitate an extension of the univariate GLM theory to account for dependent responses and, in recent years, a wide range of techniques have been developed. In this chapter we explore some of

these techniques. Section 3.1 is dedicated solely to generalized estimating equations, as they form the basis for many of the techniques covered in subsequent chapters. Section 3.2 then outlines some extensions of the GEE methodology, along with the alternative technique of generalized linear mixed models.

3.1 Generalized estimating equations

3.1.1 Introduction

Generalized estimating equations (GEEs) were originally proposed by Liang and Zeger (1986), Zeger and Liang (1986) as an extension of GLMs to longitudinal data. More recently, the generality of the GEE methodology has led to them being applied to many more situations in which the responses fall naturally into clusters. Within the GEE framework, responses within clusters are assumed to be correlated, while those from different clusters are considered independent after allowing for the covariates. GEEs are designed for situations in which the primary research interest lies in the dependence of the response variable on the covariates, with the association between responses being regarded as a nuisance.

A natural way to account for within cluster dependence is to fit a multivariate distribution to each cluster. In general, however, this approach is impractical for non-normal responses due to the lack of availability of appropriate multivariate distributions (Fitzmaurice, 1995). Liang and Zeger (1986) avoid this problem by not specifying a multivariate density for each cluster, but specifying instead the marginal distribution for each response, along with a working covariance structure for the responses within the same cluster, to allow for the dependence. In Liang and Zeger's (1986) original presentation the marginal distribution of the response variable was assumed to belong to the exponential family of distributions (see Section 2.2). However, in the spirit of quasi-likelihood (see Section 2.7.2), the

specification of a marginal distribution for the response variable may be relaxed, such that only a functional relationship between the marginal mean and variance is specified.

We now formalise the GEE methodology. Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$ be a response vector of random variables for the ith cluster, where $i = 1, \ldots, k$. In general, the cluster size m_i may vary by cluster, however to simplify presentation we assume that the cluster size m is the same for all clusters. Thus, $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$, and the total number of responses is $n = k \times m$. Corresponding to each response Y_{ij} are the values $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ of p covariates, where x_{ij1} is set to 1 to allow for an intercept. Thus, for each cluster there is an $m \times p$ matrix of covariates $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im})^T$, where each covariate can either vary or remain fixed within cluster.

As with GLMs, the response variable is related to the covariates through the relationship

$$g(\mu_{ij}) = \eta_{ij},$$

where

$$\mathrm{E}(Y_{ij}) = \mu_{ij}, \qquad \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta},$$

$\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and $g(\cdot)$ is a monotonic and differentiable link function. The variance of Y_{ij} depends on the mean of Y_{ij} as follows

$$\mathrm{var}(Y_{ij}) = \phi v(\mu_{ij}),$$

where $v(\mu_{ij})$ and ϕ denote the variance function and the dispersion parameter, respectively. In addition to specifying the marginal mean and variance structure, the within cluster dependence structure is also specified. The covariance between Y_{ij} and Y_{ik}, within cluster i, is assumed to depend on their fitted means and an association parameter vector $\boldsymbol{\alpha}$, such that for a known function τ

$$\mathrm{cov}(Y_{ij}, Y_{ik}) = \tau(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha}). \tag{3.1}$$

GEEs can be thought of as a multivariate extension of the quasi-likelihood technique introduced in Section 2.7.2. They amend the GLM score equations to incorporate the above information regarding the first two moments of the response, including the within cluster covariance structure defined in (3.1). A key feature of GEEs is that the within cluster covariance structure given in (3.1), and assumed during the estimation process, is only a working assumption which may not necessarily correspond to the true covariance structure. For this reason the covariance structure assumed during the fitting process is known as the 'working covariance structure'. In Section 3.1.2 we introduce independence estimating equations, which are the simplest example of a GEE, as an independence working covariance assumption is used. Then in Section 3.1.3 we outline the more general class of GEEs where non-independent working covariance structures are permitted.

3.1.2 Independence estimating equations

Independence estimating equations (IEEs) use the GLM score equations (as given by (2.16) in Section 2.4.1) as a set of estimating equations for estimating the regression parameter vector $\boldsymbol{\beta}$. To emphasis the clustered nature of the data, the GLM score equations are re-expressed such that all matrices are redimensioned to the size of each cluster and a sum is formed over the k (conditionally independent) clusters. Thus,

$$\mathbf{U}_q(\boldsymbol{\beta}) = \sum_{i=1}^{k} \mathbf{X}_i^T \mathbf{D}_i(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \mathbf{S}_i(\boldsymbol{\beta}) = \mathbf{0}, \tag{3.2}$$

where $\mathbf{X}_i$ is a $m \times p$ design matrix, $\mathbf{D}_i(\boldsymbol{\beta})$ is a $m \times m$ diagonal matrix with elements $(\partial \mu_{i1}/\partial \eta_{i1}, \ldots, \partial \mu_{im}/\partial \eta_{im})$, $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$ is a $m \times m$ diagonal variance-covariance matrix of the responses with elements $\{\text{var}(Y_{i1}), \ldots, \text{var}(Y_{im})\}$ and $\mathbf{S}_i(\boldsymbol{\beta})$ is a $m \times 1$ residual vector with elements $\{y_{i1} - \mu_{i1}(\boldsymbol{\beta}), \ldots, y_{im} - \mu_{im}(\boldsymbol{\beta})\}^T$. Notice that a subscript q has been introduced for the vector $\mathbf{U}(\boldsymbol{\beta})$ to emphasis that a multivariate distribution has not been specified for $\mathbf{Y}_i$ and thus $\mathbf{U}_q(\boldsymbol{\beta})$ represents what

is termed a 'quasi-score' vector.

For GLMs, the n independent responses are assumed to follow a specific distribution from the exponential family and thus the matrices $\mathbf{\Sigma}_i(\boldsymbol{\beta})$ are diagonal and assumed to correspond to the 'true' covariance of $\mathbf{Y}_i$, which we denote by $\text{cov}(\mathbf{Y}_i)$. For cluster correlated data the true covariance matrix $\text{cov}(\mathbf{Y}_i)$, will be non-diagonal, reflecting the within cluster dependence. For IEEs, however, the assumption that $\mathbf{\Sigma}_i(\boldsymbol{\beta})$ is diagonal is maintained, but it is recognised that in general $\mathbf{\Sigma}_i(\boldsymbol{\beta}) \neq \text{cov}(\mathbf{Y}_i)$, and thus $\mathbf{\Sigma}_i(\boldsymbol{\beta})$ is now labelled a working covariance matrix.

The IEEs estimate of the true parameter vector $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_I$, is obtained by solving (3.2) for $\boldsymbol{\beta}$. Liang and Zeger (1986) showed that under appropriate regularity conditions $\hat{\boldsymbol{\beta}}_I$ is consistent and asymptotically normal

$$\hat{\boldsymbol{\beta}}_I \overset{a}{\sim} N(\boldsymbol{\beta}, \mathbf{F}^{-1}(\boldsymbol{\beta})\mathbf{V}(\boldsymbol{\beta})\mathbf{F}^{-1}(\boldsymbol{\beta})), \tag{3.3}$$

where

$$\mathbf{F}(\boldsymbol{\beta}) = -\text{E}[\mathbf{U}'_q(\boldsymbol{\beta})] = \sum_{i=1}^{k} \mathbf{X}_i^T \mathbf{D}_i(\boldsymbol{\beta}) \mathbf{\Sigma}_i^{-1}(\boldsymbol{\beta}) \mathbf{D}_i(\boldsymbol{\beta}) \mathbf{X}_i, \tag{3.4}$$

and

$$\mathbf{V}(\boldsymbol{\beta}) = \text{cov}[\mathbf{U}_q(\boldsymbol{\beta})] = \sum_{i=1}^{k} \mathbf{X}_i^T \mathbf{D}_i(\boldsymbol{\beta}) \mathbf{\Sigma}_i^{-1}(\boldsymbol{\beta}) \text{cov}(\mathbf{Y}_i) \mathbf{\Sigma}_i^{-1}(\boldsymbol{\beta}) \mathbf{D}_i(\boldsymbol{\beta}) \mathbf{X}_i. \tag{3.5}$$

The regularity conditions necessary for this result to hold are similar to those stated in Section 2.3.1 for the asymptotic normality of the mle, with additional conditions placed on the cluster size.

To see why $\text{E}(\hat{\boldsymbol{\beta}}_I) \overset{a}{=} \boldsymbol{\beta}$ and $\text{cov}(\hat{\boldsymbol{\beta}}_I) \overset{a}{=} \mathbf{F}^{-1}(\boldsymbol{\beta})\mathbf{V}(\boldsymbol{\beta})\mathbf{F}^{-1}(\boldsymbol{\beta})$ we follow a similar argument to that used in Section 2.3.1 for the derivation of the asymptotic distribution of the mle. A first order Taylor series approximation of $\mathbf{U}_q(\hat{\boldsymbol{\beta}}_I)$ about the true parameter $\boldsymbol{\beta}$ is given by

$$\mathbf{U}_q(\hat{\boldsymbol{\beta}}_I) = \mathbf{U}_q(\boldsymbol{\beta}) + \mathbf{U}'_q(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}) + O_p(1).$$

Using $\mathbf{U}_q(\hat{\boldsymbol{\beta}}_I) = \mathbf{0}$, replacing $\mathbf{U}'_q(\boldsymbol{\beta})$ by its expected value and rearranging we obtain

$$\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta} = \left\{-\mathrm{E}[\mathbf{U}'_q(\boldsymbol{\beta})]\right\}^{-1} \mathbf{U}_q(\boldsymbol{\beta}) + O_p(k^{-1}). \tag{3.6}$$

Thus, $\mathrm{E}(\hat{\boldsymbol{\beta}}_I) \stackrel{a}{=} \boldsymbol{\beta}$ since $\mathrm{E}[\mathbf{U}_q(\boldsymbol{\beta})] = \mathbf{0}$, and

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}_I) \stackrel{a}{=} \left\{-\mathrm{E}[\mathbf{U}'_q(\boldsymbol{\beta})]\right\}^{-1} \mathrm{cov}\left[\mathbf{U}_q(\boldsymbol{\beta})\right] \left\{-\mathrm{E}[\mathbf{U}'_q(\boldsymbol{\beta})]\right\}^{-1}. \tag{3.7}$$

Referring to Section 2.3 on likelihood inference, for a univariate GLM we have $-\mathrm{E}[\mathbf{U}'(\boldsymbol{\beta})] = \mathrm{cov}[\mathbf{U}(\boldsymbol{\beta})]$ and thus the right hand side of (3.7) reduces to $\{-\mathrm{E}[\mathbf{U}'(\boldsymbol{\beta})]\}^{-1}$. For IEEs, however, since in general $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}) \neq \mathrm{cov}(\mathbf{Y}_i)$ this simplification does not arise and thus $\mathrm{cov}(\hat{\boldsymbol{\beta}}_I) = \mathbf{F}^{-1}(\boldsymbol{\beta})\mathbf{V}(\boldsymbol{\beta})\mathbf{F}^{-1}(\boldsymbol{\beta})$.

The covariance matrix $\mathbf{F}^{-1}(\boldsymbol{\beta})\mathbf{V}(\boldsymbol{\beta})\mathbf{F}^{-1}(\boldsymbol{\beta})$ can be consistently estimated by

$$\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{\beta}}_I) = \hat{\mathbf{F}}^{-1}(\hat{\boldsymbol{\beta}}_I)\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_I)\hat{\mathbf{F}}^{-1}(\hat{\boldsymbol{\beta}}_I), \tag{3.8}$$

where the term $\mathrm{cov}(\mathbf{Y}_i)$ in $\mathbf{V}(\boldsymbol{\beta})$ is replaced by $\mathbf{S}_i\mathbf{S}_i^T$. The matrix $\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{\beta}}_I)$ is known as the robust variance estimator of $\mathrm{cov}(\hat{\boldsymbol{\beta}}_I)$, since it provides a consistent estimator of $\mathrm{cov}(\hat{\boldsymbol{\beta}}_I)$ even when the covariance structure has been misspecified. Another name frequently used for (3.8) is the sandwich estimator due to the matrix $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_I)$ being sandwiched between the matrix $\hat{\mathbf{F}}^{-1}(\hat{\boldsymbol{\beta}}_I)$. If the elements of $\mathbf{Y}_i$ really are independent and $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}) = \mathrm{cov}(\mathbf{Y}_i)$, the expression for $\mathrm{cov}(\hat{\boldsymbol{\beta}}_I)$ reduces to $\mathbf{F}^{-1}(\boldsymbol{\beta})$. For this reason, when the matrix

$$\boldsymbol{\mathcal{N}}(\hat{\boldsymbol{\beta}}_I) = \hat{\mathbf{F}}^{-1}(\hat{\boldsymbol{\beta}}_I) \tag{3.9}$$

is used to estimate $\mathrm{cov}(\hat{\boldsymbol{\beta}}_I)$, it is known as the naive variance estimator, as its use naively assumes that the responses are independent.

If the marginal distribution of Y_{ij} belongs to the exponential family of distributions then $\hat{\boldsymbol{\beta}}_I$ corresponds to the mle $\hat{\boldsymbol{\beta}}$ for a univariate GLM. For this reason, IEEs can be viewed as a post-estimation adjustment for dependence, as we essentially fit a univariate GLM and then adjust the subsequent precision of the estimates to account for the dependence.

3.1.3 Generalized estimating equations

IEEs are a specific example of GEEs, where the working covariance matrix for cluster i, $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$, is assumed to be diagonal. In general, however, GEEs take the GLM score equations of (3.2) and replace the diagonal matrix $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$ with a non-diagonal working covariance matrix. Again, in general $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}) \neq \text{cov}(\mathbf{Y}_i)$ since $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$ is only a working covariance assumption which is usually adopted for convenience.

The working covariance matrix $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is expressed as follows

$$\boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{A}_i^{1/2}(\boldsymbol{\beta})\mathbf{R}(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}(\boldsymbol{\beta}),$$

where $\mathbf{R}(\boldsymbol{\alpha})$ is a $m \times m$ working correlation matrix and $\mathbf{A}_i(\boldsymbol{\beta})$ is a $m \times m$ diagonal matrix with elements $\{\text{var}(Y_{i1}), \ldots, \text{var}(Y_{im})\}$. The working correlation matrix $\mathbf{R}(\boldsymbol{\alpha})$ is determined by an $s \times 1$ vector of association parameters $\boldsymbol{\alpha}$, which is estimated during the fitting process from the Pearson residuals (see Section 3.1.4 below). Various structures can be placed on $\boldsymbol{\alpha}$, some of which will be discussed in Section 3.1.4.

The GEE estimate of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_G$, is obtained by solving (3.2) for $\boldsymbol{\beta}$, where $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$ may be non-diagonal. Liang and Zeger (1986) showed that under appropriate regularity conditions and provided that ϕ and $\boldsymbol{\alpha}$ are replaced by $k^{1/2}$ consistent estimates then $\hat{\boldsymbol{\beta}}_G$ is consistent (with respect to k, the number of clusters) and follows an asymptotic normal distribution

$$\hat{\boldsymbol{\beta}}_G \overset{a}{\sim} N(\boldsymbol{\beta}, \mathbf{F}^{-1}(\boldsymbol{\beta})\mathbf{V}(\boldsymbol{\beta})\mathbf{F}^{-1}(\boldsymbol{\beta})), \tag{3.10}$$

where $\mathbf{F}(\boldsymbol{\beta})$ and $\mathbf{V}(\boldsymbol{\beta})$ are given by (3.4) and (3.5), and the matrix $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})$ may be non-diagonal. The matrix $\text{cov}(\hat{\boldsymbol{\beta}}_G)$ can be consistently estimated by the robust variance estimator given in (3.8) where $\hat{\boldsymbol{\beta}}_I$ is replaced by $\hat{\boldsymbol{\beta}}_G$. Use of the naive variance estimate $\hat{\boldsymbol{\mathcal{N}}}(\hat{\boldsymbol{\beta}}_G) = \hat{\mathbf{F}}^{-1}(\hat{\boldsymbol{\beta}}_G)$ implicitly assumes the $\text{cov}(\mathbf{Y}_i)$ has been modelled correctly.

The regularity conditions which must be met for (3.10) to hold are similar to those for IEEs, with additional conditions placed on the working covariance structure.

According to Liang and Zeger (1986), one of the appealing properties of GEEs is that provided the mean of Y_{ij} is correctly specified then consistent estimates of $\boldsymbol{\beta}_G$ and $\text{cov}(\boldsymbol{\beta}_G)$ are obtained, even if the working covariance matrix has been misspecified. In contrast, Crowder (1995) showed that for some simple cases of misspecification of $\boldsymbol{\alpha}$, the above consistency properties breakdown, due to problems with the estimation of $\boldsymbol{\alpha}$.

One further point regarding the estimation of the covariance of $\hat{\boldsymbol{\beta}}_G$ is worth highlighting. Liang and Zeger (1986) proposed estimating the $\text{cov}(\mathbf{Y}_i)$ term within the robust variance estimator by $\mathbf{S}_i\mathbf{S}_i^T$, which is based on data from cluster i only. Pan (2001b) suggests a more efficient estimator of the term $\text{cov}(\mathbf{Y}_i)$, which pools information from across the k clusters. Use of this estimator, however, requires the additional assumptions that the marginal variance of the response has been modelled correctly and that there is a common correlation structure across clusters.

3.1.4 Parameter estimation

Three quantities need to be estimated during the fitting process, these being the parameter vector $\boldsymbol{\beta}$, the association vector $\boldsymbol{\alpha}$ and the dispersion parameter ϕ. The estimation process, as detailed by Liang and Zeger (1986), involves iterating back and forth between estimating the parameter vector $\boldsymbol{\beta}$ and estimating the pair $\boldsymbol{\alpha}$ and ϕ. To begin with, all quantities are initialized to some plausible values, for example, ϕ could be set to 1 and $\mathbf{R}(\boldsymbol{\alpha})$ the identity matrix. Using these initial values, the set of non-linear quasi-score equations (3.2), are solved iteratively for

$\boldsymbol{\beta}$, using the following modified Fisher scoring method

$$\hat{\boldsymbol{\beta}}_G^{(t+1)} = \hat{\boldsymbol{\beta}}_G^{(t)} + \left[\mathbf{F}\left(\hat{\boldsymbol{\beta}}_G^{(t)}\right)\right]^{-1} \mathbf{U}_q\left(\hat{\boldsymbol{\beta}}_G^{(t)}\right), \tag{3.11}$$

where t denotes the tth iteration. Once convergence is reached in $\hat{\boldsymbol{\beta}}_G$, a function of the current Pearson residuals (as described below) is used to obtain new method of moments estimates of $\boldsymbol{\alpha}$ and ϕ. These new estimates are then used in (3.11) to obtain a new estimate of $\boldsymbol{\beta}$, which in turn is used to obtain updated Pearson residuals for re-estimating $\boldsymbol{\alpha}$ and ϕ. This oscillating between $\boldsymbol{\beta}$, and the pair $\boldsymbol{\alpha}$ and ϕ continues until successive estimates of all parameters agree to a specified tolerance.

Estimation of the dispersion parameter and association parameters

Liang and Zeger (1986) proposed estimating the dispersion parameter ϕ and association vector $\boldsymbol{\alpha}$ by the method of moments. The parameters are estimated, at each iteration, from the current Pearson residuals

$$\hat{r}_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{(v(\hat{\mu}_{ij}))^{1/2}}, \tag{3.12}$$

where $i = 1, \ldots, k$ and $j = 1, \ldots, m$. Given the set of $n(= mk)$ Pearson residuals, the dispersion parameter ϕ is estimated by

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{k} \sum_{j=1}^{m} \hat{r}_{ij}^2.$$

The exact method of estimation of $\boldsymbol{\alpha}$ is determined by the specific structure used. Some common structures are outlined below.

1. Independence - letting $\mathbf{R}(\boldsymbol{\alpha}) = \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix, the working assumption that responses within a cluster are uncorrelated is adopted, corresponding to IEEs. In this case there are no correlation parameters to estimate.

2. Exchangeable - assumes that the pairwise correlation is constant for all pairs of observations within the cluster, $\text{corr}(Y_{ij}, Y_{il}) = \alpha$, $j \neq l$. Only a single correlation parameter α needs to be estimated ($s = 1$), and this can be achieved by

$$\hat{\alpha} = \frac{1}{\hat{\phi}\left\{\left[\frac{1}{2}km(m-1)\right] - p\right\}} \sum_{i=1}^{k} \sum_{l>j} \hat{r}_{il}\hat{r}_{ij}. \tag{3.13}$$

3. AR(1) - an autoregressive structure of order one is commonly placed on α within a longitudinal setting, $\text{corr}(Y_{ij}, Y_{il}) = \alpha^{|j-l|}$. Again, only a single parameter α needs to be estimated ($s = 1$),

$$\hat{\alpha} = \frac{1}{\hat{\phi}\left\{[k(m-1)] - p\right\}} \sum_{i=1}^{k} \sum_{j=1}^{m-1} \hat{r}_{ij}\hat{r}_{i,j+1}. \tag{3.14}$$

4. Unstructured - $\boldsymbol{\alpha}$ is left completely unspecified and each of the pairwise correlations are estimated separately, $\text{corr}(Y_{ij}, Y_{il}) = \alpha_{jl}$, $j \neq l$. Under this structure, $\boldsymbol{\alpha}$ is a vector with $s = \frac{1}{2}m(m-1)$ elements, which can be estimated by

$$\hat{\alpha}_{jl} = \frac{\sum_{i=1}^{k} \hat{r}_{ij}\hat{r}_{il}}{\hat{\phi}(k-p)}. \tag{3.15}$$

An alternative approach for the estimation of $\boldsymbol{\alpha}$ involves supplementing the set of estimating equations for $\boldsymbol{\beta}$ with an additional set for $\boldsymbol{\alpha}$. This approach is discussed further in Section 3.2.1

There has been considerable debate concerning the relative merits of GEEs with a non-diagonal working covariance structure, compared to their IEEs counterparts. Liang and Zeger (1986) stated that increased efficiency results from using a non-diagonal structure which is as close as possible to the true covariance structure. Also, Fitzmaurice (1995) showed that use of IEEs can result in considerable loss of efficiency when covariates which vary within cluster are present and the correlation between responses is high. However, McDonald (1993) and Sutradhar and Das (1999) both highlight benefits of adopting an IEEs approach.

An additional advantage of IEEs, which is particularly important for large data sets, is that they are computationally more efficient to implement.

3.1.5 Hypothesis testing

Within this section we consider hypothesis testing for the regression parameters of GEEs. The tests outlined here are extensions of those tests detailed in Section 2.5 for GLMs, with appropriate adjustments for the within cluster dependence. Throughout this section we consider testing the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$, where $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$ and $\boldsymbol{\beta}_2^0$ denotes a specific value of $\boldsymbol{\beta}_2$. The dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p_1 and p_2, respectively.

Robust Wald test

By the asymptotic normality of $\hat{\boldsymbol{\beta}}_G$ given in (3.10), the robust Wald statistic

$$W_T = (\hat{\boldsymbol{\beta}}_{G2} - \boldsymbol{\beta}_2^0)^T \left[\boldsymbol{\mathcal{R}}_{22}(\hat{\boldsymbol{\beta}}_G)\right]^{-1} (\hat{\boldsymbol{\beta}}_{G2} - \boldsymbol{\beta}_2^0), \tag{3.16}$$

follows an asymptotic $\chi^2_{p_2}$ distribution under H_0, where $\boldsymbol{\mathcal{R}}_{22}(\hat{\boldsymbol{\beta}}_G)$ is the $p_2 \times p_2$ submatrix of the robust variance matrix (3.8) corresponding to $\boldsymbol{\beta}_2$.

Robust Wald tests are by far the most commonly used method of hypothesis testing for GEEs. This is largely due to convenience, as robust standard errors are routinely available from standard software packages.

Quasi-score test

The quasi-score vector defined in (3.2) is asymptotically distributed as

$$\mathbf{U}_q(\boldsymbol{\beta}) \overset{a}{\sim} N(\mathbf{0}, \mathbf{V}(\boldsymbol{\beta})), \tag{3.17}$$

where $\mathbf{V}(\boldsymbol{\beta})$ is given by (3.5). Result (3.17) follows from a similar argument to that used in Section 2.3.1 for the asymptotic distribution of the score vector. Under H_0, it follows that the quasi-score test statistic

$$W_S = \mathbf{U}_q \left[\hat{\boldsymbol{\beta}}_{G1}, \boldsymbol{\beta}_2^0\right]^T \mathbf{V}^{-1} \left[\hat{\boldsymbol{\beta}}_{G1}, \boldsymbol{\beta}_2^0\right] \mathbf{U}_q \left[\hat{\boldsymbol{\beta}}_{G1}, \boldsymbol{\beta}_2^0\right] \tag{3.18}$$

follows an asymptotic $\chi^2_{p_2}$ distribution, where $\hat{\boldsymbol{\beta}}_{G1}$ is the restricted GEE estimate of $\boldsymbol{\beta}_1$ under H_0. The term $\text{cov}(\mathbf{Y}_i)$ in $\mathbf{V}(\boldsymbol{\beta})$ must be estimated and as in (3.8) this can be achieved using $\mathbf{S}_i\mathbf{S}_i^T$, under H_0.

Likelihood ratio test

While extending the Wald and score tests of Section 2.5, such that they are appropriate for GEEs, is fairly straightforward, unfortunately this is not the case for the likelihood ratio test. The reason for this is that while the Wald and score tests involve only first and second moments of $\mathbf{Y}$, the likelihood ratio test depends on full distributional properties and since GEEs do not assume a specific form for the multivariate distribution of $\mathbf{Y}_i$ we do not have a likelihood function.

Assuming the marginal distribution of Y_{ij} belongs to the exponential family, Rotnitzky and Jewell (1990) derived the asymptotic distribution of the 'independence' likelihood ratio test statistic when applied to clustered data. This test statistic is constructed under the assumption that all responses are independent and Rotnitzky and Jewell (1990) derive its asymptotic distribution using properties of quadratic forms of non-normal random variables (see Johnson and Kotz, 1970, p.150). They propose using this theory as a basis for testing H_0, essentially by calculating the 'independence' likelihood ratio test statistic and adjusting the critical value of the test to allow for the within cluster dependence.

More formally, under H_0, the 'independence' likelihood ratio statistic is given by

$$W_L = -2\left[\ell(\hat{\boldsymbol{\beta}}_{I1}, \boldsymbol{\beta}_2^0) - \ell(\hat{\boldsymbol{\beta}}_I)\right], \tag{3.19}$$

where $\ell(\hat{\boldsymbol{\beta}}_{I1}, \boldsymbol{\beta}_2^0)$ and $\ell(\hat{\boldsymbol{\beta}}_I)$ are the restricted and unrestricted independence log-likelihood functions, respectively. Rotnitzky and Jewell (1990) show that, under H_0 and for clustered data, asymptotically,

$$W_L = \sum_{i=1}^{p_2} d_i X_i, \tag{3.20}$$

where X_i, $i = 1, \ldots, p_2$, are independent χ_1^2 random variables and d_i are the eigenvalues of $(\mathbf{F}^{-1}\mathbf{V}\mathbf{F}^{-1})_{22}\left[(\mathbf{F}^{-1})_{22}\right]^{-1}$. The matrices $(\mathbf{F}^{-1}\mathbf{V}\mathbf{F}^{-1})_{22}$ and $(\mathbf{F}^{-1})_{22}$ are the $p_2 \times p_2$ submatrices of $\mathbf{F}^{-1}\mathbf{V}\mathbf{F}^{-1}$ and $\mathbf{F}^{-1}$ corresponding to $\boldsymbol{\beta}_2$, and calculated under a working independence assumption. The matrix $(\mathbf{F}^{-1}\mathbf{V}\mathbf{F}^{-1})_{22}\left[(\mathbf{F}^{-1})_{22}\right]^{-1}$ can be consistently estimated by $\boldsymbol{\mathcal{R}}_{22}(\hat{\boldsymbol{\beta}}_I)\left[\boldsymbol{\mathcal{N}}_{22}(\hat{\boldsymbol{\beta}}_I)\right]^{-1}$, where $\boldsymbol{\mathcal{R}}_{22}(\hat{\boldsymbol{\beta}}_I)$ and $\boldsymbol{\mathcal{N}}_{22}(\hat{\boldsymbol{\beta}}_I)$ are the $p_2 \times p_2$ submatrices of the robust and naive variance estimates corresponding to $\boldsymbol{\beta}_2$, and calculated under a working independence assumption.

If the data really are independent then $\mathbf{F} = \mathbf{V}$ and hence we arrive back at the usual chi-squared distribution with p_2 d.f., since $(\mathbf{F}^{-1}\mathbf{V}\mathbf{F}^{-1})_{22}\left[(\mathbf{F}^{-1})_{22}\right]^{-1}$ is the identity matrix whose eigenvalues are all 1.

In the simple case where a single parameter is being tested, the above theory simplifies such that W_L/d follows a chi-squared distribution with 1 d.f., where d is the ratio of the robust to naive variance estimates for the parameter of interest. When more than one parameter is being tested, the asymptotic distribution of W_L follows a linear combination of χ_1^2 random variables, which in practice must be approximated. Rotnitzky and Jewell (1990) suggest adjusting the test statistic and treating $W_L/\bar{d}$ as a χ^2 random variable with p_2 d.f., where $\bar{d} = \sum_{i=1}^{p_2} d_i/p_2$. An improved approximation can be obtained, as detailed in Bowman and Azzalini (1997, p. 86-88), by using results on quadratic forms of normal random variables. This method involves approximating the distribution of W_L by a shifted and scaled chi-squared distribution of the form $a\chi_b^2 + c$, where the constants a, b and c are found by matching the moments of a $a\chi_b^2 + c$ distribution with those of W_L. Thus, the adjusted test statistic $(W_L - c)/a$ is treated as a χ^2 random variable with b d.f.

Using a similar argument to that used for the 'independence' likelihood ratio test, Rotnitzky and Jewell (1990) present what they term 'working' Wald and score tests. The 'working' Wald test replaces the robust variance estimate in (3.16) with the naive variance estimate $\mathcal{N}_{22}(\hat{\beta}_G)$, while the 'working' score test replaces the matrix $\mathbf{V}^{-1}$ in (3.18) with the naive variance estimate. Rotnitzky and Jewell (1990) show that, like the 'independence' likelihood ratio statistic, these test statistics follow a linear combination of χ^2_1 distributions. Thus, these working tests have a more complicated asymptotic distribution than their robust counterparts. However, they have the advantage that they remove the need to estimate the term $\text{cov}(\mathbf{Y}_i)$ in $\mathbf{V}$ which involves the estimation of $\frac{1}{2}m(m-1)$ values. Rotnitzky and Jewell (1990) advocate the use of these tests when there are a small number of clusters and cluster sizes are large, since the estimation of $\text{cov}(\mathbf{Y}_i)$ can become unstable in this setting. Unlike the 'independence' likelihood ratio test, these tests may be used for all GEEs and not just IEEs.

3.1.6 Longitudinal example

We now analyse a small longitudinal data set using both a univariate GLM analysis and a GEE analysis to gain an insight into how the estimates from these two approaches differ. The data set has been taken from Thall and Vail (1990) and the presentation of results is similar to Pickles (1998). The data relate to an epilepsy trial, where for each of 59 patients, the number of seizures experienced within a two week period was recorded for four successive periods. The purpose of the study was to see if a drug called Progabide was effective in reducing the number of seizures experienced. Each patient was assigned to either a placebo (treatment=0) or Progabide (treatment=1), and the covariates of baseline measurement and age were controlled in the study.

This data set has been presented on numerous occasions within the GEE literature, for example, by Pickles (1998) and Hardin and Hilbe (2002). A Poisson

distribution is assumed for the marginal distribution of the response variable, and it is argued that a GEE analysis is necessary to allow for the dependence within individual, which arises as a result of four successive measurements being taken. An alternative and more parsimonious analysis would involve modelling the totals for each subject directly, with some allowance for overdispersion. For illustrative purposes, however, we follow the approach adopted by other authors.

We now proceed to fit various models to this data set using S-Plus (Venables and Ripley, 1994). The results from fitting a Poisson GLM, overdispersed-Poisson quasi-model (see Section 2.7.2) and Poisson IEE are given in Table 3.1. The coefficient estimates under the three different approaches are the same, whereas the standard errors differ. The standard errors for the GLM are labelled 'naive' as they do not allow for the within subject correlation. The standard errors for the overdispersed model are labelled 'quasi' and allow for the dependence by scaling the naive standard errors by the constant factor $\hat{\phi}^{1/2}$, where $\hat{\phi} = 5.1$ (which is substantially bigger than the value 1 for Poisson data). Finally, the IEE standard errors are labelled 'robust', which allow for the dependence via the robust variance matrix. Ignoring the dependence and using naive standard errors for inference, leads us incorrectly to conclude that the treatment effect is significant at the 5% level. Once the dependence is accounted for via robust standard errors, the treatment effect becomes insignificant at the 5% level.

Table 3.2 shows the results of fitting three different GEEs to the data set, where the three separate sets of results correspond to three common working correlation structures: exchangeable, AR(1) and unstructured. The regression estimates and standard errors from the three separate fits are all similar and therefore the three different approaches produce similar conclusions. Moving on to compare these coefficient estimates with those from the GLM analysis in Table 3.1, we see that again the estimates are similar.

Table 3.3 shows the lower triangle of the estimated working correlation matrix

Covariate	Estimate	Naive Std Err	Quasi-Std Err	Robust Std-Error
Time	-0.0589	0.0202 (-2.92)	0.0458 (-1.29)	0.0353 (-1.67)
Treatment	-0.1544	0.0478 (-3.23)	0.1080 (-1.43)	0.1718 (-0.90)
Age	0.0226	0.0040 (5.65)	0.0091 (2.48)	0.0126 (1.79)
Baseline	0.0227	0.0005 (45.4)	0.0011 (19.7)	0.0012 (18.92)
Constant	0.7115	0.1444 (4.92)	0.3260 (2.18)	0.3482 (2.04)

Table 3.1: Results from fitting a Poisson GLM, overdispersed Poisson quasi-model and Poisson IEE to the epilepsy data. The coefficient estimates are the same for all three approaches, while the estimated standard errors differ ('naive' for GLM, 'quasi' for overdispersed quasi-model and 'robust' for IEE); t-values are given in brackets.

under each of the three GEE structures. The unstructured matrix resembles the AR(1) structure and therefore the AR(1) structure is preferred as it avoids the potential problem of estimator bias, caused by the need to estimate a large number of nuisance parameters (Liang and Zeger, 1995). In this particular case, however, this concern is not too serious since there are only six nuisance parameters in the unstructured form, which is fairly small compared with the number of clusters. This explains why, in this case, the estimates of the regression parameters are similar for all methods.

3.2 Alternatives to conventional generalized estimating equations

Within this section we consider extensions to the basic GEE algorithm, along with the alternative technique of generalized linear mixed models.

Covariate	Structure	Coefficient Estimate	Robust Std Err	t-value
Time	Exchangeable	-0.0589	0.0350	-1.68
	AR(1)	-0.0640	0.0338	-1.89
	Unstructured	-0.0516	0.0423	-1.22
Treatment	Exchangeable	-0.1495	0.1689	-0.88
	AR(1)	-0.1647	0.1602	-1.03
	Unstructured	-0.1480	0.1317	-1.12
Age	Exchangeable	0.0234	0.0118	1.98
	AR(1)	0.0260	0.0119	2.19
	Unstructured	0.0237	0.0122	1.93
Baseline	Exchangeable	0.0227	0.0012	18.27
	AR(1)	0.0232	0.0012	18.65
	Unstructured	0.0228	0.0012	19.44
Constant	Exchangeable	0.6802	0.3547	1.92
	AR(1)	0.5974	0.3510	1.70
	Unstructured	0.6249	0.3780	1.65

Table 3.2: Results from fitting three separate GEEs to the epilepsy data.

Exchangeable				AR(1)				Unstructured			
1.00				1.00				1.00			
0.40	1.00			0.51	1.00			0.24	1.00		
0.40	0.40	1.00		0.26	0.51	1.00		0.42	0.68	1.00	
0.40	0.40	0.40	1.00	0.13	0.26	0.51	1.00	0.21	0.29	0.59	1.00

Table 3.3: Estimated GEE working correlation matrices for the epilepsy data.

3.2.1 Second order generalized estimating equations

GEEs are usually applied in situations where explaining the relationship between the response and the covariates is the main aim, with the association between responses being a nuisance. If the association is important, however, then the GEE methodology can be extended to allow additional emphasis to be placed on the estimation of the association. One such approach involves supplementing the set of estimating equations for $\boldsymbol{\beta}$ with an additional set of estimating equations for $\boldsymbol{\alpha}$ (Prentice, 1988) and solving these sets of equations jointly. This is known as second order generalized estimating equations or GEE2 for short. We now provide a brief overview of this method.

The set of estimating equations for $\boldsymbol{\beta}$ given in (3.2) can be written in a slightly different form as follows

$$\mathbf{U}_{\beta}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \left[\mathbf{y}_i - \boldsymbol{\mu}_i\right] = \mathbf{0}, \tag{3.21}$$

where $\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)$ is an $m \times p$ matrix with its jkth element equal to $\frac{\partial \mu_{ij}}{\partial \beta_k}$.

Similarly for $\boldsymbol{\alpha}$, we can define a set of estimating equations of the form

$$\mathbf{U}_{\alpha}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\alpha}}\right)^T \mathbf{W}_i^{-1} \left[\mathbf{t}_i - \boldsymbol{\theta}_i\right] = \mathbf{0}, \tag{3.22}$$

where $\mathbf{t}_i = (t_{i12}, t_{i13}, \ldots, t_{i,m-1,m})$, $t_{ijk} = (y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})$, $\boldsymbol{\theta}_i = E(\mathbf{T}_i)$ and $\mathbf{W}_i$ is the working covariance matrix of $\mathbf{t}_i$. $\mathbf{W}_i$ has dimensions $m(m-1)/2 \times m(m-1)/2$, and a convenient choice is $\mathbf{W}_i = \text{diag}\{\text{var}(t_{i12}), \text{var}(t_{i13}), \ldots, \text{var}(t_{i,m-1,m})\}$.

If $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are treated as if they are orthogonal, then equations (3.21) and (3.22) can be solved for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ using separate modified Fisher scoring algorithms. This approach is known as first order generalized estimating equations (GEE1). Alternatively, both sets of estimating equations can be solved

jointly

$$\mathbf{U}_{\alpha\beta}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^{n} \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\alpha}} \\ \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\alpha}} \end{pmatrix}^T \begin{pmatrix} \mathrm{cov}(\mathbf{Y}_i) & \mathrm{cov}(\mathbf{Y}_i, \mathbf{T}_i) \\ \mathrm{cov}(\mathbf{Y}_i, \mathbf{T}_i) & \mathrm{cov}(\mathbf{T}_i) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{t}_i - \boldsymbol{\theta}_i \end{pmatrix} = \mathbf{0}, \tag{3.23}$$

and this is known as second order generalized estimating equations (GEE2).

In contrast to GEE1, GEE2 does not provide consistent estimates of $\boldsymbol{\beta}$ when the working correlation structure has been misspecified. The GEE2 approach can, however, provide more efficient estimates. Generally speaking, these gains in efficiency, while potentially significant for $\boldsymbol{\alpha}$, are small for $\boldsymbol{\beta}$ (Liang et al., 1992). Thus, GEE2 is only recommended if significant interest lies in the estimation of the association.

Within this thesis we are concerned with modelling large data sets and therefore the GEE2 algorithm has the significant disadvantage that it is computationally more expensive to implement than conventional GEEs. Furthermore, our primary aim is in explaining the relationship between the covariates and the response, with the association between the responses being less important. For these reasons, the GEE2 algorithm is less appealing than the conventional GEE algorithm and we will not consider GEE2 further.

3.2.2 One-step generalized estimating equations

One drawback of the GEE method is that it is computationally more intensive than the univariate GLM algorithm, due to the extra level of iteration involved in estimating the working correlation matrix. Therefore, it would be beneficial if we could speed up or adjust the GEE algorithm in some way to make it more practical for modelling large data sets, such as those typically experienced in climatology. Lipsitz et al. (1994) consider a completely different problem relating to the full GEE algorithm failing to converge, which can be quite common when sample sizes

are small and the correlation is high (Lipsitz et al., 1994). The authors recommend using a one-step GEE algorithm when the full GEE algorithm fails to converge. This one-step method involves starting with the $\boldsymbol{\beta}$ vector corresponding to a GLM fit. From this independence fit, $\boldsymbol{\alpha}$ and ϕ are estimated from the Pearson residuals and finally one further iteration of the GEE algorithm for $\boldsymbol{\beta}$ is undertaken. The authors carry out simulations for the binary case and show that the performance of the one-step estimator is qualitatively similar to that of the full GEE method in terms of bias and power.

We propose applying the one-step estimator to large data sets, which would otherwise require an excessive amount of computing time for implementation of the full GEE algorithm. In Chapter 6 we apply the one-step estimator to a climate case study and consider its performance.

3.2.3 Generalized linear mixed models

Generalized linear mixed models (GLMMs) are an alternative technique to GEEs to account for within cluster dependence. They extend univariate GLMs by assuming that the responses are conditionally independent given a vector of cluster-specific random effects, and it is the introduction of these random effects which induces correlation within cluster. The conditional distribution of the responses belongs to the exponential family, while typically the distribution of the random effects is assumed to be multivariate normal.

Estimation of the regression parameters $\boldsymbol{\beta}$ can be undertaken by the method of maximum likelihood. To achieve this, however, a marginal likelihood for the responses must be obtained, which involves integrating out the random effects. In addition to obtaining estimates of the regression parameters, estimates of the variance components of the random effects are also of interest and these are usually obtained using maximum likelihood (ML) or restricted maximum

likelihood (REML). REML estimates are often preferred as they avoid some of the bias problems that arise with ML estimation.

In recent years, the topic of parameter estimation for GLMMs has been considered at great length from various perspectives and there are now many competing techniques available. One of the original treatments was due to Schall (1991) who used a penalised-likelihood approach, avoiding the need for integration. Similar approaches were adopted by Breslow and Clayton (1993), and Wolfinger and O'Connell (1993). Alternative techniques include approximating the integrals by numerical Gauss-Hermite quadrature, Gibbs sampling (Zeger and Karim, 1991), and application of the EM-algorithm (McCulloch, 1997) which treats the random effects as missing data. A simplification also arises when the distribution chosen for the random effects is conjugate to the conditional distribution of the responses, as analytical solutions are then available.

Distinction between GEEs and GLMMs

GEEs can be viewed as marginal methods since parameter estimates are obtained by averaging over clusters. In contrast, GLMMs are called subject-specific models, since all parameter estimates obtained are conditional on the realised random effects. Thus, the parameter estimates obtained from GLMMs should be interpreted on a subject by subject basis. When analysing a specific data set, the choice of technique will depend largely on the purpose of the study. If explaining the average overall influence of the covariates on the response is the primary aim then the GEE methodology is likely to be the most appropriate. If, however, identifying risk on an individual basis is most important, then the subject-specific approach of GLMMs is likely to be preferable. The distinction between parameter interpretation for GEEs and GLMMs is covered at length in Zeger et al. (1988).

As outlined in Chapter 1, climatology has been used to motivate this work,

where we are typically interested in explaining the whole system under study rather than individual aspects. Therefore the GEE methodology is most appropriate since we attempt to explain the average overall dependence of the response on the covariates. Moreover, as climate data sets are typically large, the GEE methodology has the additional advantage that it is, in general, computationally more efficient to implement when compared with GLMMs. For these reasons, we focus on the GEE methodology for the remainder of this thesis.

Chapter 4

Hypothesis testing for generalized linear models applied to clustered data

4.1 Introduction

One approach to modelling clustered data is to fit a univariate GLM and then adjust the subsequent inference to allow for the within cluster dependence. This approach is adopted by Liang and Zeger's (1986) IEEs (see Section 3.1.2), which allow for the dependence by using the robust variance matrix for inference. Within this chapter we propose a new hypothesis testing technique for this setting, which involves adjusting the 'independence' likelihood ratio test statistic to allow for the within cluster dependence. This approach to modelling clustered data has considerable computational advantages over other techniques such as GEEs and GLMMs.

This chapter is organised as follows. Section 4.2 outlines the theory of the new approach to hypothesis testing. In Section 4.3, the geometry of the new method

is explored, when testing both a single unknown parameter and two unknown parameters. Then in Section 4.4, using simulations, the new method is compared with the established hypothesis testing techniques outlined in Section 3.1.5.

4.2 New adjusted likelihood ratio test

4.2.1 Motivation

In Section 3.1.5 various hypothesis testing techniques were introduced for the regression parameters of IEEs, which included the robust Wald test, the quasi-score test and Rotnitzky and Jewell's likelihood ratio test. All of these techniques have some drawbacks, for example, the performance of the robust Wald test can be adversely affected when correlated predictors are present (Chandler, 1998). Rotnitzky and Jewell's likelihood ratio test statistic, on the other hand, follows a complicated asymptotic distribution which in practice must be approximated. Their test also uses the independence log-likelihood function to construct confidence regions, which are therefore likely to be of the 'wrong' shape (see Section 4.3).

Within this section we propose a new hypothesis testing technique as an alternative to the established techniques outlined above. The approach is similar to Rotnitzky and Jewell's, in that an adjustment is made to the 'independence likelihood ratio' test to allow for the within cluster dependence. The two methods differ, however, in that Rotnitzky and Jewell use the same test statistic and adjust the critical value of the test, whereas we propose adjusting the test statistic itself and maintain the same critical value for the test.

4.2.2 General theory and derivation of test

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ be a response vector of random variables for the ith cluster ($i = 1, \ldots, k$), where the marginal distribution of Y_{ij} belongs to the exponential family of distributions. Corresponding to each Y_{ij} are the values $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ of p covariates, which are related to the response via the usual GLM relationship

$$g\left(\mathrm{E}[Y_{ij}]\right) = \mathbf{x}_{ij}^T \boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters and $g(\cdot)$ is a link function.

Interest lies in testing the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$, where the parameter vector $\boldsymbol{\beta}$ has been partitioned $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$ and $\boldsymbol{\beta}_2^0$ denotes a specific value of $\boldsymbol{\beta}_2$. The dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p_1 and p_2, respectively.

If the responses were independent then the log-likelihood function would be given by

$$\ell_{IND}(\boldsymbol{\beta}) = \sum_{i=1}^{k} \sum_{j=1}^{m} \log\left[f(y_{ij}; \boldsymbol{\beta})\right], \tag{4.1}$$

where $f(\cdot)$ denotes the pdf of Y_{ij}. Then to test H_0 we could calculate the 'independence' likelihood ratio test statistic

$$W_L = -2\left[\ell_{IND}(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2^0) - \ell_{IND}(\hat{\boldsymbol{\beta}})\right], \tag{4.2}$$

where under H_0, $\ell_{IND}(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_2^0)$ and $\ell_{IND}(\hat{\boldsymbol{\beta}})$ are the maximized restricted and unrestricted 'independence' log-likelihood functions respectively, and W_L follows an asymptotic $\chi^2_{p_2}$ distribution.

For clustered data, responses are not independent and therefore the 'independence' log-likelihood function of (4.1) does not hold. Moreover, since we do not assume a joint distribution for the within cluster responses we are unable to write down a likelihood function. Therefore, no standard likelihood ratio test can

be applied. Thus, we propose adjusting the independence likelihood ratio test statistic of (4.2) to allow for the within cluster dependence.

The new method constructs a new 'dependence adjusted' inference function $\ell(\beta)$, for carrying out inference on β. The function is constructed about $\hat{\beta}_I$ (the mle from the independence model) and is designed in such a way that the profile of the independence log-likelihood function is maintained, since we believe this contains valuable information. To achieve this a linear transformation is defined between β^* on the independence log-likelihood surface and β on the new dependence adjusted surface. The transformation is designed in such a way that $\ell(\beta) = \ell_{IND}(\beta^*)$ is accurate to second order (in the sense that Wald tests are preserved) in the neighbourhood of $\hat{\beta}_I$. Below we formalise these ideas.

In general terms, a linear transformation from β to β^* can be defined as follows

$$\beta^* = \mathcal{T}(\beta) = \hat{\beta}_I + \mathbf{C}\left(\beta - \hat{\beta}_I\right), \tag{4.3}$$

where $\mathbf{C}$ is a $p \times p$ square matrix. Thus we may define

$$\ell(\beta) = \ell_{IND}(\beta^*) = \ell_{IND}\left(\hat{\beta}_I + \mathbf{C}[\beta - \hat{\beta}_I]\right),$$

and a second order Taylor series expansion of $\ell_{IND}\left(\hat{\beta}_I + \mathbf{C}[\beta - \hat{\beta}_I]\right)$ in the neighbourhood of $\hat{\beta}_I$ is given by

$$\begin{aligned}
\ell_{IND}\left(\hat{\beta}_I + \mathbf{C}[\beta - \hat{\beta}_I]\right) &= \ell_{IND}(\hat{\beta}_I) + \left(\mathbf{C}[\beta - \hat{\beta}_I]\right)\ell'_{IND}(\hat{\beta}_I) \\
&+ \frac{1}{2}\left(\mathbf{C}[\beta - \hat{\beta}_I]\right)^T \ell''_{IND}(\hat{\beta}_I)\left(\mathbf{C}[\beta - \hat{\beta}_I]\right) + O_p(k^{-1/2}) \\
&= \ell_{IND}(\hat{\beta}_I) - \frac{1}{2}(\beta - \hat{\beta}_I)^T\mathbf{C}^T\mathbf{I}_{IND}(\hat{\beta}_I)\mathbf{C}(\beta - \hat{\beta}_I) + O_p(k^{-1/2}),
\end{aligned} \tag{4.4}$$

where $\mathbf{I}_{IND}(\beta)$ denotes the information matrix under the independence model. Also, a second order Taylor series expansion of $\ell(\beta)$ about $\hat{\beta}_I$ is given by

$$\ell(\beta) = \ell(\hat{\beta}_I) - \frac{1}{2}(\beta - \hat{\beta}_I)^T\mathbf{I}(\hat{\beta}_I)(\beta - \hat{\beta}_I) + O_p(k^{-1/2}), \tag{4.5}$$

where $\mathbf{I}(\boldsymbol{\beta})$ denotes the information matrix under dependence. Equating the second order approximation of $\ell(\boldsymbol{\beta})$ given by (4.4) with the one given in (4.5), and since $\ell(\hat{\boldsymbol{\beta}}_I) = \ell_{IND}(\hat{\boldsymbol{\beta}}_I)$ by construction, we obtain

$$\mathbf{I}(\hat{\boldsymbol{\beta}}_I) = \mathbf{C}^T \mathbf{I}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C}. \tag{4.6}$$

Now, defining $\mathbf{M}_{IND}(\boldsymbol{\beta})$ to be a matrix square root of $\mathbf{I}_{IND}(\boldsymbol{\beta})$, such that

$$\mathbf{I}_{IND}(\boldsymbol{\beta}) = \mathbf{M}_{IND}^T(\boldsymbol{\beta})\mathbf{M}_{IND}(\boldsymbol{\beta})$$

and, similarly

$$\mathbf{I}(\boldsymbol{\beta}) = \mathbf{M}^T(\boldsymbol{\beta})\mathbf{M}(\boldsymbol{\beta}),$$

(4.6) can be written as

$$\mathbf{I}(\hat{\boldsymbol{\beta}}_I) = \mathbf{C}^T \mathbf{I}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C} = \mathbf{C}^T \mathbf{M}_{IND}^T(\hat{\boldsymbol{\beta}}_I)\mathbf{M}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C} = [\mathbf{M}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C}]^T \mathbf{M}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C}.$$

Thus,

$$\mathbf{M}^T(\hat{\boldsymbol{\beta}}_I)\mathbf{M}(\hat{\boldsymbol{\beta}}_I) = [\mathbf{M}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C}]^T \mathbf{M}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C},$$

which is satisfied by

$$\mathbf{M}(\hat{\boldsymbol{\beta}}_I) = \mathbf{M}_{IND}(\hat{\boldsymbol{\beta}}_I)\mathbf{C}.$$

Therefore, an appropriate transformation can be achieved by taking

$$\mathbf{C} = \mathbf{M}_{IND}^{-1}(\hat{\boldsymbol{\beta}}_I)\mathbf{M}(\hat{\boldsymbol{\beta}}_I). \tag{4.7}$$

The matrix $\mathbf{C}$ is not unique because $\mathbf{M}$ and $\mathbf{M}_{IND}$ are not uniquely determined. Thus, the transformation defined by (4.3) is also not unique, however, the second order approximation (4.5) is unique.

The matrices $\mathbf{I}(\boldsymbol{\beta})$ and $\mathbf{I}_{IND}(\boldsymbol{\beta})$ must be estimated and this can be consistently achieved by the inverse of the robust and naive variance estimates (see (3.8) and (3.9) of Section 3.1.2), evaluated at $\hat{\boldsymbol{\beta}}_I$, respectively. Thus, the matrix $\mathbf{C}$ can be estimated by

$$\mathbf{C} = \left\{ \left[\boldsymbol{\mathcal{N}}(\hat{\boldsymbol{\beta}}_I)^{-1} \right]^{1/2} \right\}^{-1} \left[\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{\beta}}_I)^{-1} \right]^{1/2}, \tag{4.8}$$

where the notation $\mathbf{A}^{1/2}$ denotes the matrix square root of $\mathbf{A}$ obtained from the Choleski factorization. In the special case when only a single parameter is being tested C is a scalar, which simplifies to the square root of the ratio of the naive to robust variance estimates for the parameter of interest.

In applying the above theory to test H_0, two GLMs are fitted to clustered data, one nested within the other. For the restricted model we then apply the transformation defined in (4.3) to the vector $(\hat{\boldsymbol{\beta}}_{I1}^T, \boldsymbol{\beta}_2^{0T})$ to obtain the corresponding vector $(\hat{\boldsymbol{\beta}}_{I1}^{*T}, \boldsymbol{\beta}_2^{0*T})$. The usual 'independence' likelihood ratio test is then carried out using this transformed parameter vector. Thus, the null hypothesis is tested by calculating the following test statistic

$$W_{NL} := -2\left[\ell_{IND}(\hat{\boldsymbol{\beta}}_{I1}^*, \boldsymbol{\beta}_2^{0*}) - \ell_{IND}(\hat{\boldsymbol{\beta}}_I)\right], \tag{4.9}$$

where asymptotically W_{NL} follows a $\chi^2_{p_2}$ distribution under H_0. This asymptotic result follows since the adjustment has been designed in such a way that the asymptotic theory outlined in Section 2.3 carries over directly.

Note that when the data really are independent, $\mathbf{I}(\boldsymbol{\beta})$ and $\mathbf{I}_{IND}(\boldsymbol{\beta})$ are identical, the matrix $\mathbf{C}$ is the identity matrix, $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and hence we obtain the usual 'independence' likelihood ratio test, as expected.

In some special cases, the transformation defined by (4.3) may result in $\boldsymbol{\beta}^*$ taking on a value outside of its allowable range. In this instance we recommend reparameterizing, for example, if $\boldsymbol{\beta}$ must lie in the interval [0,1] then a logistic transformation may be applied.

4.3 Geometry of the new test

4.3.1 Single parameter case

Within this section we investigate the geometry of the new method when testing a single unknown parameter. To achieve this we focus on a simple example where the data, y_i $(i = 1, \ldots, n)$, consist of n=20 observations from a geometric distribution with pmf

$$\Pr(Y_i = y_i) = \theta(1-\theta)^{y_i}, \quad y_i = 0, 1, \ldots. \tag{4.10}$$

Interest lies in carrying out inference on the single unknown parameter θ.

The independence log-likelihood function for θ can be written as

$$\ell(\theta, \mathbf{y}) = n\left\{\bar{y}\log(1-\theta) + \log\theta\right\}, \tag{4.11}$$

where $\bar{y}$ denotes the sample mean of the data, which in this case is taken to be 3. The mle is given by $\hat{\theta} = 1/(1+\bar{y}) = 0.25$. The independence log-likelihood function for θ has been plotted in Figure 4.1 and based on this function, a 95% acceptance region for the null hypothesis $H_0 : \theta = \theta_0$, has been constructed and is also shown.

For illustration, suppose that the 20 observations were actually obtained from k clusters of equal size, where the observations within each cluster are correlated, and that the ratio of the robust to naive variance estimates is 2. As a result of this correlation the independence log-likelihood function is incorrect and the acceptance region for H_0 based on this function is too narrow. To allow for the dependence, the new method constructs a new inference function, by applying the inverse transformation of (4.3) to each value θ^* on the independence surface. For the single parameter case this transformation simplifies to

$$\theta = \hat{\theta} + \sqrt{Q}(\theta^* - \hat{\theta}), \tag{4.12}$$

where $Q = C^{-2}$ is a scalar, which can be estimated by the ratio of the robust to naive variance estimate for θ. The transformed inference function, used to carry out inference on θ, is shown in Figure 4.1. A 'dependence adjusted' 95% acceptance region for H_0, based on the new inference function, is also shown.

It is interesting to compare our adjustment to that of Rotnitzky and Jewell (see Section 3.1.5). They use the independence log-likelihood function and make an adjustment to the critical value to allow for the dependence. As discussed earlier, this corresponds to dropping the critical value line by a factor equal to Q, the ratio of the robust to naive variance estimates, and this is shown in Figure 4.2. We, on the other hand, essentially stretch out the independence log-likelihood function by the square root of the same ratio. This implies that when a single parameter is being tested and the independence log-likelihood function is quadratic, the two methods are equivalent. Thus, when testing a single parameter the new method and Rotnitzky and Jewell's method are asymptotically equivalent.

4.3.2 Two parameter case

Having compared the new method with Rotnitzky and Jewell's method when testing a single unknown parameter, we now consider a simple example when testing two unknown parameters. Suppose that a pair of correlated measurements are recorded on each of k individuals; suppose also that each pair of measurements follows a bivariate normal distribution with known variance-covariance matrix, where both variances are 1 and the correlation ρ is 0.7. Interest lies in carrying out inference on the unknown mean $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ by testing the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. For illustration, we present calculations for 10 data pairs that have been simulated from a bivariate normal distribution with $\boldsymbol{\mu} = (2, 3)^T$.

The 'true' bivariate normal log-likelihood function for $\boldsymbol{\mu}$, given the data, can

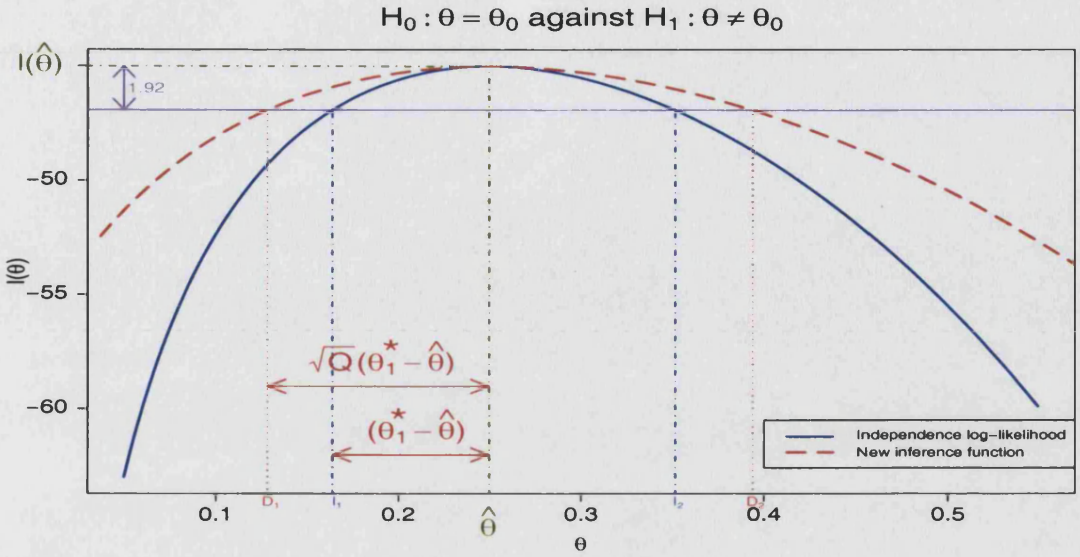

Figure 4.1: Geometry of new method when testing a single parameter. 95% acceptance regions for H_0, based on the independence log-likelihood function and the new inference function are given by $[I_1,I_2]$ and $[D_1,D_2]$ respectively.

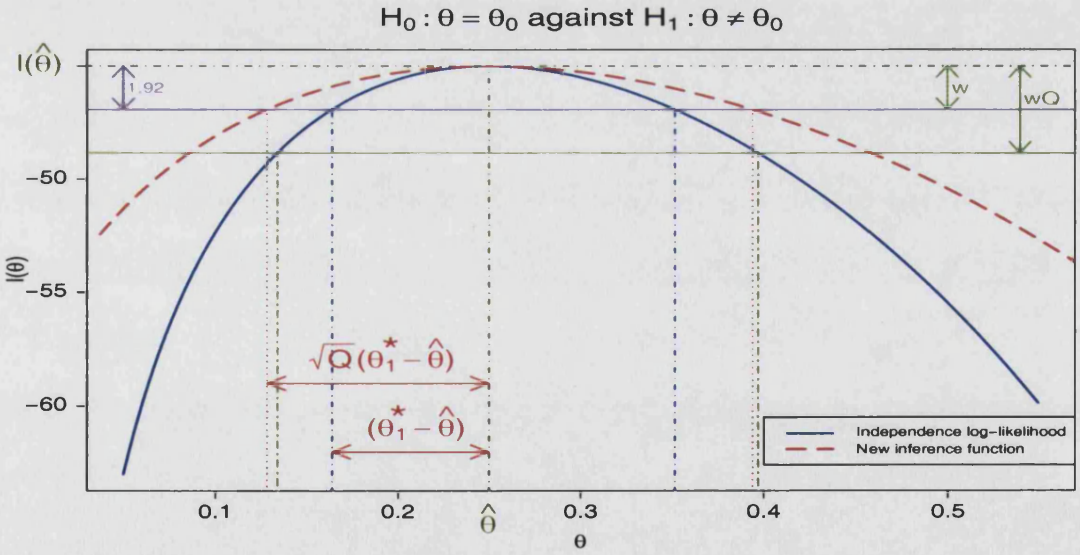

Figure 4.2: Comparison of new method with Rotnitzky and Jewell's method when testing a single parameter. 95% acceptance regions for H_0, based on the new method and Rotnitzky and Jewell's method are shown by line types $(\cdots)$ and $(-\cdot)$ respectively.

be written as

$$\begin{aligned} \ell_{TRUE}(\boldsymbol{\mu};\mathbf{y}) &= -k\log\left(2\pi\sqrt{1-\rho^2}\right) \\ &- \frac{1}{2(1-\rho^2)}\sum_{i=1}^{k}\left[(y_{i1}-\mu_1)^2 - 2\rho(y_{i1}-\mu_1)(y_{i2}-\mu_2) + (y_{i2}-\mu_2)^2\right]. \end{aligned} \tag{4.13}$$

This log-likelihood function for $\boldsymbol{\mu}$ has been plotted for the simulated data in the top left hand plot of Figure 4.3. Acceptance regions for H_0 based on the true bivariate normal log-likelihood function follow the elliptical contours shown.

To obtain an acceptance region for H_0, Rotnitzky and Jewell work with the independence log-likelihood function, which assumes all observations are independent. The independence log-likelihood function is given by

$$\ell_{IND}(\boldsymbol{\mu};\mathbf{y}) = -k\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\left[(y_{i1}-\mu_1)^2 + (y_{i2}-\mu_2)^2\right] \tag{4.14}$$

and this function has been plotted for the simulated data in the top right hand plot of Figure 4.3. When constructing a confidence region for $\boldsymbol{\mu}$ they allow for the dependence by slicing this surface at a different level to that used for truly independent data. Nevertheless, acceptance regions for H_0 are still circular and not elliptical as given by the true bivariate normal log-likelihood function.

We, on the other hand, define an inference function $\ell(\boldsymbol{\mu})$ on which acceptance regions for H_0 are constructed. Our inference function is defined as $\ell(\boldsymbol{\mu}) = \ell_{IND}(\boldsymbol{\mu}^*)$, with

$$\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}}_I + \mathbf{C}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_I).$$

This is equivalent to applying the inverse transformation, given by

$$\boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{pmatrix}(\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}), \tag{4.15}$$

to every point on the independence log-likelihood surface in the top right hand plot of Figure 4.3. Since the transformation applied is a function of the correlation between pairs of observations, the transformed surface obtained, which is shown

in the bottom left hand plot of Figure 4.3, has the same elliptical contours as the true bivariate normal log-likelihood function. Thus, an acceptance region for H_0 based on the new inference function will be of the correct elliptical shape. This provides a good illustration of the claim made in Section 4.2, that the new method aims to produce an inference surface with the right shape.

Based on the above it is hoped that when testing more than one parameter the new method will outperform Rotnitzky and Jewell's method, and this will be explored further in the next section. In the multiparameter case the new method also has the additional advantage of being computationally more efficient, since Rotnitzky and Jewell's method involves computing both an eigenvalue analysis and an approximation to the asymptotic distribution of the test statistic (see Section 3.1.5).

4.4 Simulation studies

We now investigate the performance of the new method relative to the established techniques of the robust Wald test, the quasi-score test and Rotnitzky and Jewell's likelihood ratio test. This is achieved via simulation, where the exact mechanism for generating the data is known. Two contrasting simulation environments are considered. Sections 4.4.2 and 4.4.3 present simulations based upon binary and gamma response variables respectively. Before presenting the results, we outline in Section 4.4.1 how we intend to use simulations to compare and contrast the various tests.

4.4.1 Performance assessment criteria

How good a particular test is, relative to other competing tests, is usually measured in terms of its type I and type II error rates. Under a given H_0, a test

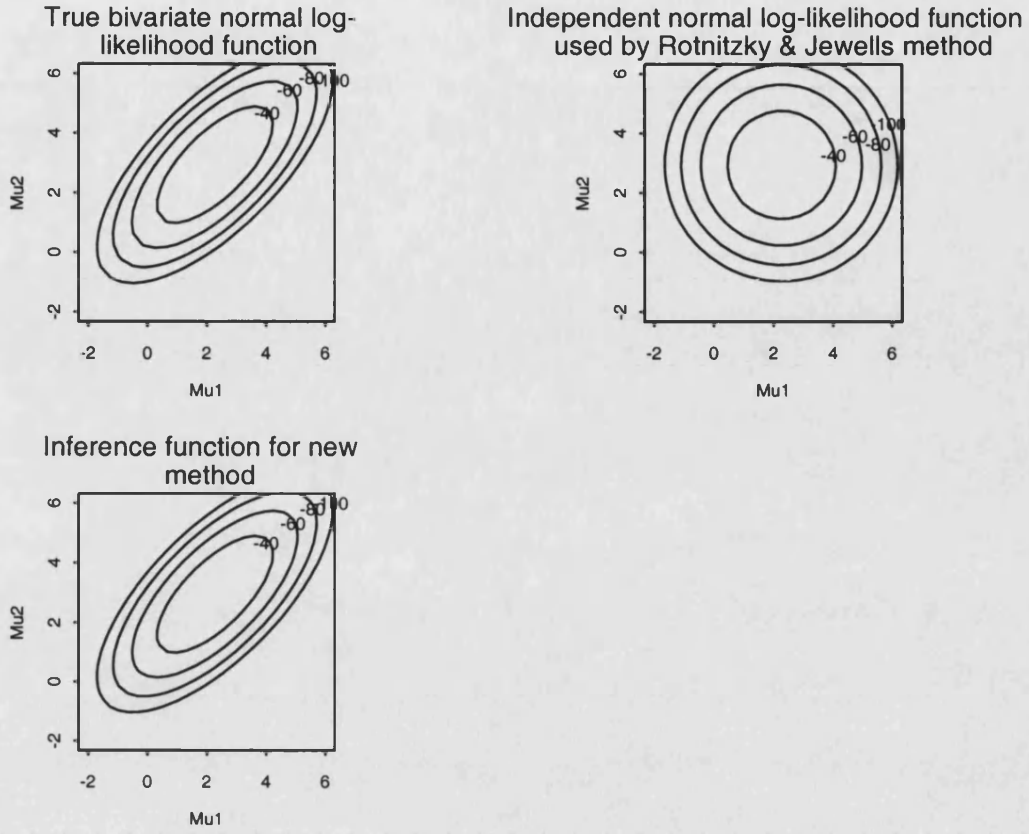

Figure 4.3: Comparison of new method with Rotnitzky and Jewell's method when testing two parameters.

can result in one of two possible errors. A type I error is made if H_0 is rejected when it is true and conversely, a type II error occurs if H_0 is accepted when it is false. The probability of making type I and type II errors are denoted by α and γ respectively. Naturally, it is desirable to have α and γ as small as possible. In practice, however, for a fixed sample size, decreasing α results in an increase in γ, and vice versa. Therefore when comparing tests it is common to set the value of α and then prefer the test with the smallest value of γ. The quantity α is also known as the size or significance level of the test and is typically set to a value of 0.05 or 0.01.

A related quantity to the type II error, is the power of the test. The power of a test is defined as the probability of rejecting H_0 when it is false. Clearly, the power of the test is equal to $1-\gamma$ and therefore when tests are compared with α fixed, the test with the greatest power is usually preferred.

Power curves can be constructed to compare tests under a specified null hypothesis. Figure 4.4, which is based on Figures 10.13 and 10.14 of Wackerley et al. (2002), shows some theoretical power curves, when testing the null hypothesis $H_0 : \beta = \beta_0$. The quantities along the horizontal and vertical axes represent the true value of β and the power of the test, respectively. Ideally, we would like to reject H_0 with probability 1 when the true value β is not equal to β_0 ($\beta \neq \beta_0$), and accept H_0 with probability 1 when β is equal to β_0. This is represented in Figure 4.4 by the 'perfect' power curve, however, in reality such a curve is not attainable. A 'typical' power curve is also shown in Figure 4.4, where the further the true value moves away from β_0, the greater the power of the test. Notice also that at the point $\beta = \beta_0$ the power is equal to α, the size of the test. Finally, in Figure 4.4 a 'preferred typical' power curve has also been plotted, where this power curve is preferred to the 'typical' power curve as it has greater power, being closer to the 'perfect' power curve.

Using simulations, we plan to construct power curves, similar to those in Figure 4.4, in an attempt to compare the various competing tests. A wide range of scenarios and null hypotheses will be considered to enable as much evidence as possible to be collected. Using the results obtained, we hope to be able to answer many questions, some of which are as follows:

1. Does the new test have the correct coverage? Thus, if we apply the new test at size α, do we reject a true null hypothesis $100\alpha\%$ of the time?

2. How does the new test compare to Rotnitzky and Jewell's test? Does the answer to this question depend on the number of parameters being tested?

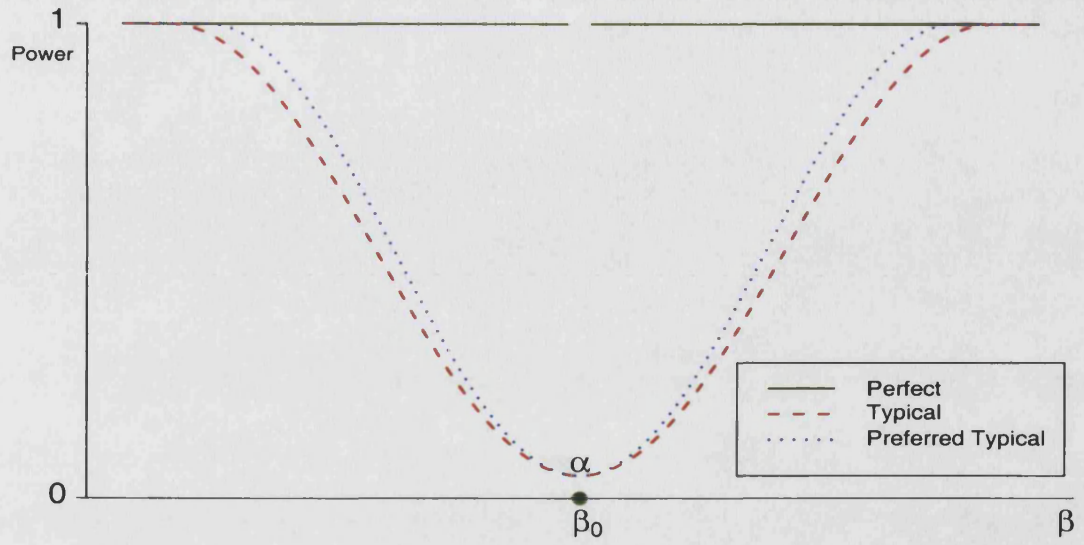

Figure 4.4: Theoretical power curves for testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.

3. How does the new test perform relative to the robust Wald and quasi-score tests? In particular, does the new test outperform the robust Wald test when dealing with correlated predictors?

4.4.2 Binary simulations

In this section the performance of the new test is compared with the established hypothesis testing techniques introduced in Section 3.1.5, through simulations of binary variables. Using a marginal logistic regression model, the ability of the test to identify significant predictors under specified null hypotheses is considered.

Simulation environment

The simulation environment adopted is the same as that used by Fitzmaurice (1995) and more recently by Pan (2001a). There are 100 clusters of data and within each cluster three repeated measurements are recorded over time. The response variable Y_{ij} is binary, and its marginal mean μ_{ij} is modelled by a logistic regression model of the form

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2}, \tag{4.16}$$

where $i = 1, \ldots, 100$, $j = 1, 2, 3$, x_{ij1} is a Bernoulli variable which is fixed within each cluster (representing group membership) and takes the values of 0 and 1 with equal probabilities of 0.5, and $x_{ij2} = (j - 1)$ representing a linear trend within each cluster. The true parameter values are $\beta_0 = 0.25$ and $\beta_1 = \beta_2 = -0.25$.

The joint distribution of $\mathbf{Y}_i$ is simulated using the Bahadur (1961) representation which can be written as

$$f(\mathbf{y}_i|\boldsymbol{\mu}_i, \boldsymbol{\rho}_i) = \prod_{j=1}^{3} \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1-y_{ij}} \cdot \left(1 + \sum_{j<k} \rho_{ijk} w_{ij} w_{ik}\right), \tag{4.17}$$

where W_{ij} is the standardised variable $W_{ij} = (Y_{ij} - \mu_{ij})/\sqrt{\mu_{ij}(1 - \mu_{ij})}$ and $\rho_{ijk} = \text{E}(W_{ij}W_{ik})$ is the correlation between Y_{ij} and Y_{ik}. The second term in (4.17) represents the extent of the dependence in $\mathbf{Y}_i$. All pairwise correlations ρ_{ijk} $(j \neq k)$ are set to 0.5, thus an exchangeable correlation structure is assumed. For more details on Bahadur's representation see Cox (1972).

The implementation of the simulation process can be summarised by the following steps:

1. Generate x_{ij1} for $i = 1, \ldots, 100$ and $j = 1, 2, 3$.

2. Using the true $\boldsymbol{\beta}$ and (4.16), calculate the marginal means μ_{ij} for $i = 1, \ldots, 100$ and $j = 1, 2, 3$.

3. For each of the 100 clusters, use the marginal means obtained in step 2 above to simulate a correlated $\mathbf{Y}_i$ vector using Bahadur's representation, with $\rho_{ijk} = 0.5$.

4. Use these simulated $\mathbf{Y}_i$'s to fit the following logistic regression models:

   (a) regress $\mathbf{Y}$ on a constant only,

   (b) regress $\mathbf{Y}$ on a constant and x_1,

   (c) regress $\mathbf{Y}$ on a constant, x_1 and x_2,

   and obtain estimates of $\boldsymbol{\beta}$.

5. Using the estimates of $\boldsymbol{\beta}$ obtained in step 4, test the null hypotheses $H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$ for all test procedures under consideration. Since the true values are $\beta_1 = \beta_2 = -0.25$, these null hypotheses are incorrect and should be rejected. For each procedure under each hypothesis obtain a p-value.

6. Repeat steps 1 to 5, 1,000 times to obtain 1,000 independently simulated p-values for each test procedure under each hypothesis; use these to produce a simulated p-values cdf. These simulated distribution functions can then be compared across tests to assess the relative performance of each test.

Figure 4.5 shows the simulated p-values cdf, based on 1,000 simulations, for each of the four competing tests, under the null hypothesis $H_0 : \beta_2 = 0$. Since H_0 is false, the method that corresponds to the greatest number of small p-values is preferred. It can be seen, however, that all four test procedures perform comparably as the four simulated cdf's overlay each other.

When producing Figure 4.5, the 1,000 simulations were undertaken with the true values of β_1 and β_2 both set equal to -0.25. To produce power curves similar to those in Figure 4.4, the above process must be repeated for many different true values of β_1 and β_2. This has been undertaken, where the true values of

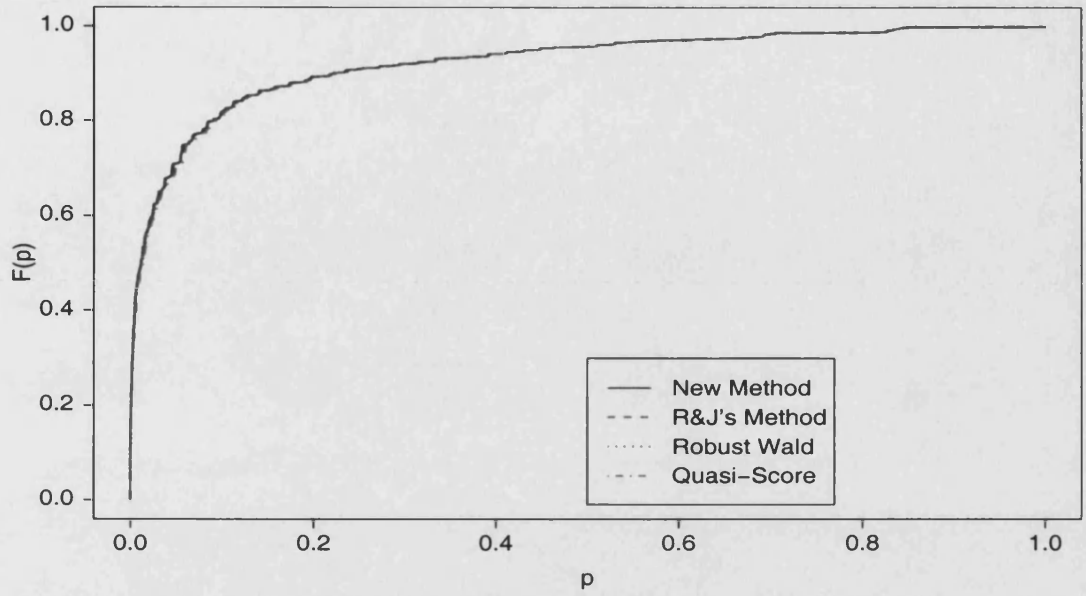

Figure 4.5: Simulated p-value cdf for each of the four competing tests under $H_0 : \beta_2 = 0$. All four lines overlay each other and are therefore indistinguishable.

β_1 and β_2 were varied together within the range -1 to 1 in steps of 0.05. Thus 41 simulations, each of size 1,000 were performed. Using this computationally intensive method, simulated power curves were produced for the various tests under the specified hypotheses, for a chosen size of test.

Power curve results

a) Testing $H_0 : \beta_2 = 0$

The first set of power curves considered relate to the testing of the null hypothesis $H_0 : \beta_2 = 0$. In Figure 4.6 the power curves for each of the four methods are displayed, where the size of the test is fixed at $\alpha = 0.05$ in all cases. This plot shows that all four tests perform comparably, as the four power curves overlay

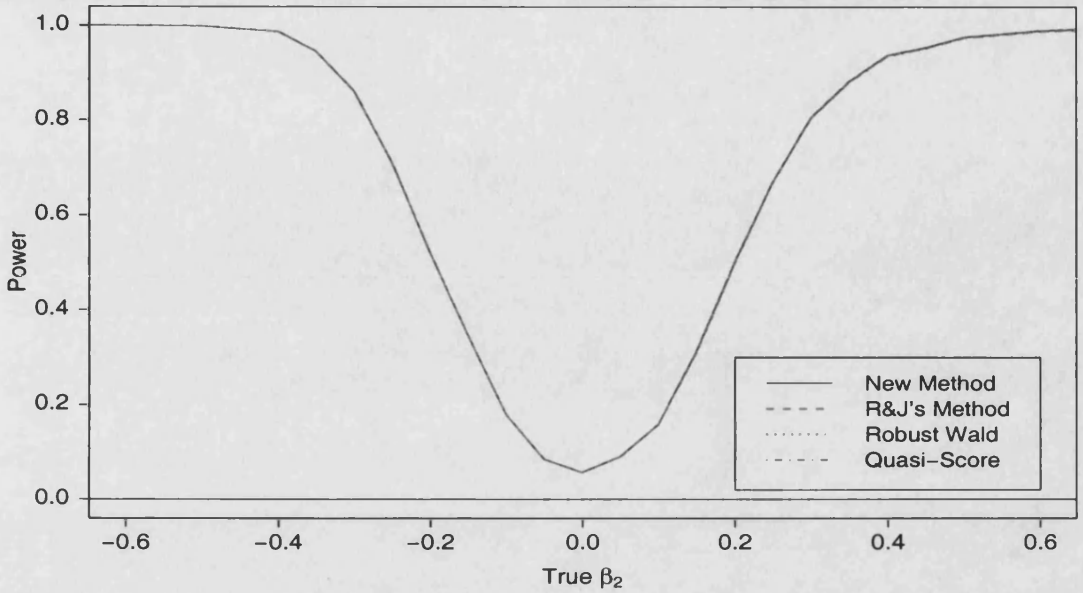

Figure 4.6: Power curves for each of the four competing tests under $H_0 : \beta_2 = 0$, $\alpha = 0.05$. All four power curves overlay each other and are therefore indistinguishable.

each other. These results are fairly uninteresting, although they do highlight two important features of the new method. Firstly, the method produces sensible results, as in this case the results are comparable with the other established techniques. Secondly, the new test has the correct coverage properties, since when the true value of β_2 is zero, the null hypothesis is correct and the new method does indeed reject H_0 5% of the time.

b) Testing $H_0 : \beta_1 = \beta_2 = 0$

In Figure 4.7 the simulated power curves under $H_0 : \beta_1 = \beta_2 = 0$ are shown for the two competing likelihood ratio tests, where $\alpha = 0.05$. From this plot it is clear that the new method has greater power than Rotnitzky and Jewell's method. This result is consistent with the theory discussed earlier, which sug-

gested that the new method may be more powerful when testing more than a single parameter. Figure 4.8 reproduces Figure 4.7 with the additional power curves for the robust Wald test and quasi-score test added. It can be seen that the new method performs comparably with the robust Wald and quasi-score tests in terms of power.

Correlated covariates

All of the above simulations have been carried out with the two covariates x_1 and x_2 uncorrelated. We now consider introducing correlation between the two covariates. The definition of covariate x_1 is modified, such that it is no longer a Bernoulli variable which is fixed within clusters, but instead it is allowed to vary within each cluster and is defined as $x_{ij1} = \text{Bernoulli}(0.3j - 0.1)$. Thus x_{ij1} is a 0/1 variable, with a probability of being 1 that increases over time. The definition of x_2 remains unchanged, $x_{ij2} = j - 1$. Thus, as both covariates on average, increase over time, this induces correlation between them. The above analysis has been repeated with the newly defined correlated covariates.

Power curve results for correlated covariates

a) Testing $H_0 : \beta_2 = 0$, x_1 and x_2 correlated

Figure 4.9 shows the simulated power curves for the new method and the robust Wald test, for correlated covariates, under $H_0 : \beta_2 = 0$. It can be seen that the new method has greater power than the robust Wald test. This result is expected from the theory discussed earlier, since robust Wald tests are known not to perform particularly well when only a subset of the correlated covariates present, are tested. This is because a robust Wald test only considers the regression coefficients and their estimated variances for those covariates being tested, while none

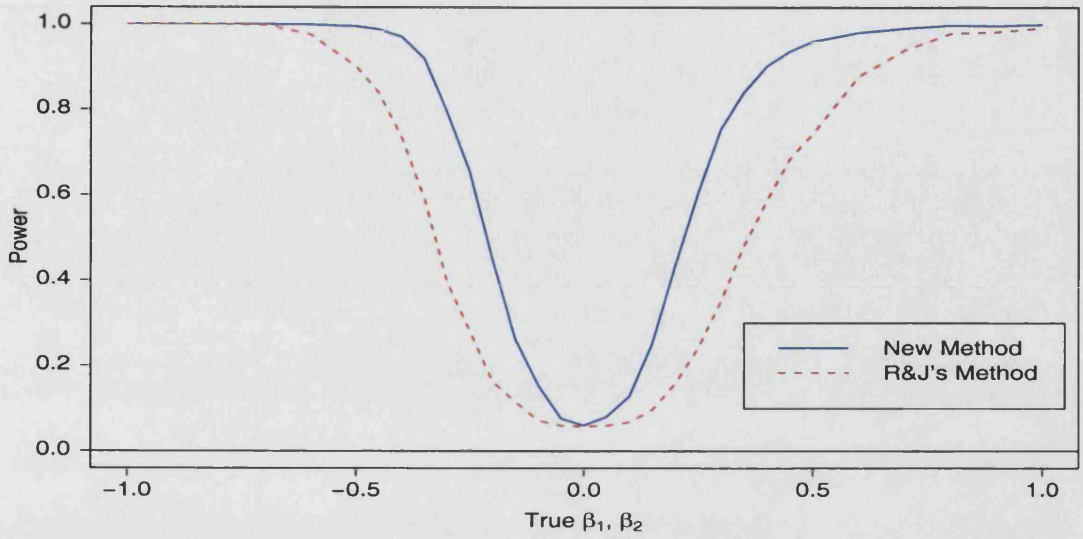

Figure 4.7: Power curves for new method and Rotnitzky and Jewell's method under H_0 : $\beta_1 = \beta_2 = 0$, $\alpha = 0.05$.

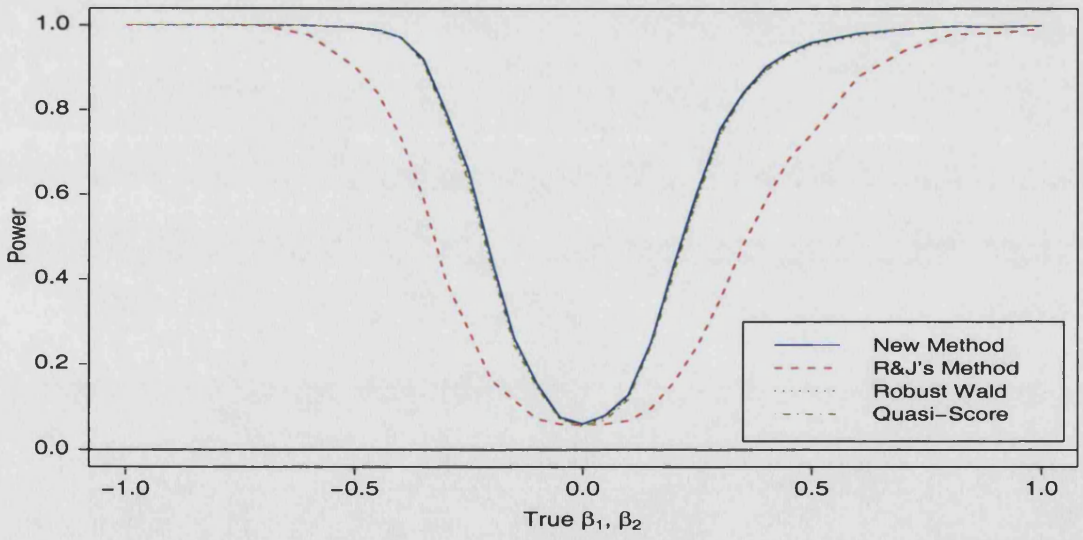

Figure 4.8: Power curves for each of the four competing tests under H_0 : $\beta_1 = \beta_2 = 0$, $\alpha = 0.05$. All power curves, except that for Rotnitzky and Jewell's method, are very similar and therefore difficult to distinguish.

of the other estimated effects are considered.

Figure 4.10 reproduces Figure 4.9 with the additional power curves for the quasi-score test and Rotnitzky and Jewell's likelihood ratio test added. From this we can see that the new method also outperforms these methods very slightly.

b) Testing $H_0 : \beta_1 = \beta_2 = 0$, x_1 and x_2 correlated

Finally, Figure 4.11 displays all four simulated power curves under $H_0 : \beta_1 = \beta_2 = 0$, where x_1 and x_2 are correlated. The results obtained here are very similar to those obtained under the same null hypothesis when x_1 and x_2 were uncorrelated. Thus, the new method outperforms Rotnitzky and Jewell's method, and is comparable with the robust Wald and quasi-score tests.

Overall, when all correlated covariates are tested together, the new method appears to provide no advantage over the robust Wald test in terms of power. However, when correlated covariates are present and only a subset of these are tested together, the new method appears to provide an advantage over the robust Wald test, since the robust Wald test is unable to account for the correlation in the covariates which are not being tested. The new test therefore is expected to be preferable when applied to climate data, where many of the covariates are correlated and it is not feasible to identify and then test all of the correlated covariates together.

4.4.3 Gamma simulations

In this section the performance of the new test is compared with the established hypothesis testing techniques introduced in Section 3.1.5, through simulations of gamma variables.

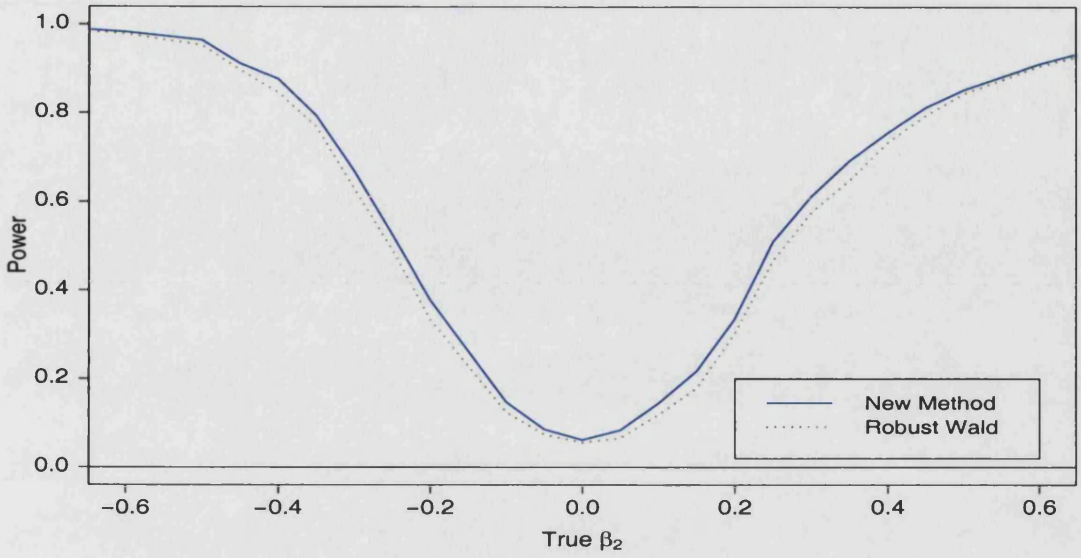

Figure 4.9: Power curves for new method and robust Wald test under $H_0 : \beta_2 = 0$, $\alpha = 0.05$, x_1 and x_2 correlated.

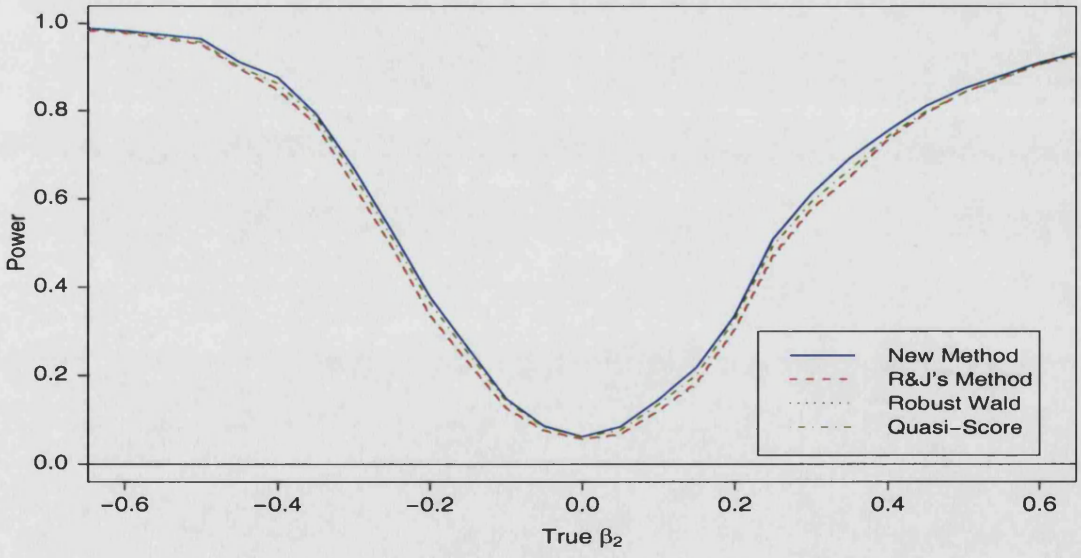

Figure 4.10: Power curves for each of the four competing tests under $H_0 : \beta_2 = 0$, $\alpha = 0.05$, x_1 and x_2 correlated.

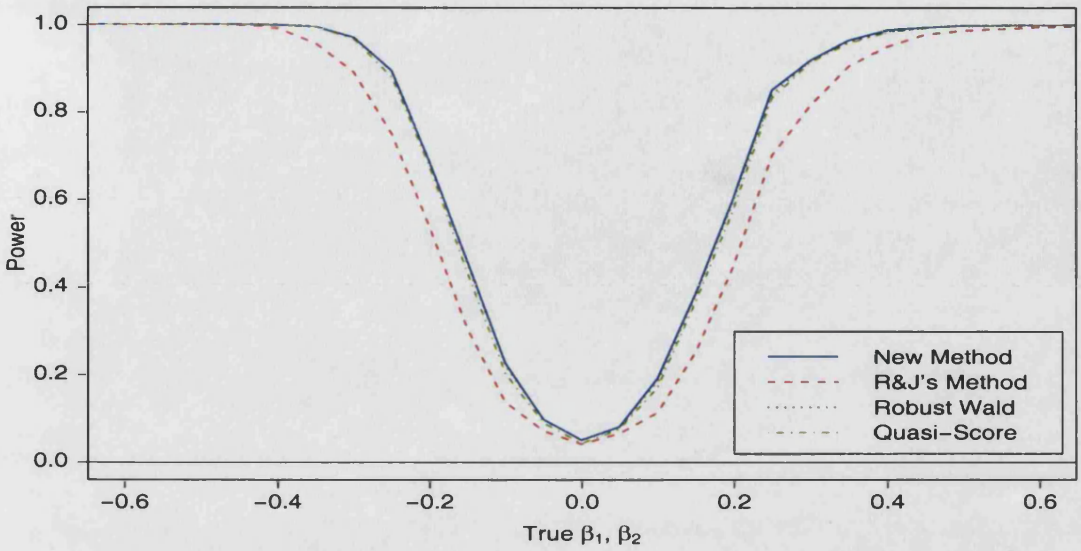

Figure 4.11: Power curves for each of the four competing tests under $H_0 : \beta_1 = \beta_2 = 0$, $\alpha = 0.05$, x_1 and x_2 correlated. All power curves, except Rotnitzky and Jewell's method, are very similar and therefore difficult to distinguish.

Simulation environment

One of the main objectives of the thesis, is to develop methods that are suitable for space-time data. Therefore, we now consider a space-time simulation environment. A simple gamma GLM is fitted to some actual space-time data, and the fitted model is then used to generate simulations representing realistic spatial-temporal structure. Using these simulations, various null hypotheses are then tested using the various competing hypothesis testing techniques, in an attempt to compare the performance of the tests.

The data relate to daily maximum wind speeds over northwestern Europe. This data set is analysed extensively in Chapter 6 and therefore the reader is referred forward for more details. Here we consider only a subset of the data, namely, six specific sites for the seven year period 1983-1989. The six sites are

made up of two sub-groups; one sub-group comprises of two neighbouring sites located in the ocean just north of Ireland and the second sub-group consists of four neighbouring sites located over land in Germany. In total, there are 2,557 days and 15,342 (=2,557×6) observations.

To formulate the above data set within a cluster correlated data framework, we incorporate autoregressive based covariates into the model and then assume that time points are independent, conditional on the covariates. Thus, time points correspond to conditionally independent clusters, and within cluster there are six spatially correlated measurements. Chapter 5 provides further details of formulating space-time data within a cluster correlated framework.

We assume that the daily maximum wind speed Y_{ts}, at time point t and spatial location s, is gamma distributed, and its marginal mean μ_{ts} is modelled by a gamma GLM of the form

$$\log(\mu_{ts}) = \beta_0 + \beta_1 x_{ts1} + \beta_2 x_{ts2} + \beta_3 x_{ts3} + \beta_4 x_{ts4} + \beta_5 x_{ts5} \qquad (4.18)$$

where $t = 1, \ldots, 2557$, $s = 1, \ldots, 6$, x_{ts1} and x_{ts2} are autoregressive terms of the form $\log(1 + y_{t-1,s})$ and $\log(1 + y_{t-2,s})$ respectively, x_{ts3} is a cosine component used to capture seasonality, and x_{ts4} and x_{ts5} are annualized climate indices for the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) respectively. The NAO and AO are large scale circulation patterns that are believed to impact upon climate within the region under study.

Model (4.18) was fitted to the data and the following parameter estimates were obtained $\beta_0 = 0.31$, $\beta_1 = 0.67$, $\beta_2 = 0.04$, $\beta_3 = 0.06$, $\beta_4 = -0.01$ and $\beta_5 = 0.02$. In addition to these parameter estimates, the correlations in Anscombe residuals for each pair of sites were calculated. In summary, the pairwise correlations for pairs of sites in different sub-groups was approximately zero, whereas the pairwise correlations for sites within the same sub-group ranged from 0.4-0.8.

The above fitted model was used to generate simulated data. Each simulated

data set comprised of daily values for the period 1983-1989 at the six chosen sites, this being equivalent in form to the original data set. Five hundred simulated data sets were generated in total. Note that this is only half the number of simulated data sets used in the binary case of Section 4.4.2, however, in that case there were 300 (=100×3) observations per data set, whereas here we have 15,342 (=2,557×6) observations. Thus, due to computational constraints we decided to produce 500 simulations only.

The implementation of the simulation process can be summarised by the following steps:

1. Initialize the simulated values for the first two time points Y_{ts} $(t = 1, 2, s = 1, \ldots, 6)$ to some observed wind speed values.

2. Using (4.18) and the estimate of β obtained from the data fit, calculate the marginal means μ_{ts} $(t = 3, s = 1, \ldots, 6)$. This is possible since all covariate values are known.

3. Use the marginal means obtained in step 2 to simulate a correlated $\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{t6})$, $t = 3$, consistent with the fitted correlations in Anscombe residuals. Anscombe residuals are approximately normally distributed, and for the gamma distribution are defined by $r_{ts}^{(A)} = (y_{ts}/\mu_{ts})^{1/3}$. Therefore the simulation process can proceed by simulating a vector of Anscombe residuals $\mathbf{r}_t^{(A)} = r_{t1}^{(A)}, \ldots, r_{t6}^{(A)}$ from a multivariate normal distribution with the fitted correlation structure. Then each simulated value can be obtained by rearranging the above relationship for the Anscombe residuals i.e. $y_{ts} = \mu_{ts}(r_{ts}^{(A)})^3$. This approach preserves the spatial structure within each time point.

4. Repeat steps 2 and 3 for all remaining time points $t = 4, \ldots, 2557$.

5. Use these simulated $\mathbf{Y}_t$'s to fit a series of gamma GLMs, for example

(a) regress $\mathbf{Y}$ on x_1 and x_2

(b) regress $\mathbf{Y}$ on x_1, x_2, x_3, x_4 and x_5

and obtain estimates of $\boldsymbol{\beta}$.

6. Using the estimates of $\boldsymbol{\beta}$ obtained from step 5, test, for example, the null hypothesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ for all test procedures under consideration. Since the true values of β_3, β_4 and β_5 are all non-zero, the null hypothesis is false and should be rejected. For each procedure under each hypothesis obtain a p-value.

7. Repeat steps 1 to 6, 500 times to obtain 500 independently simulated p-values for each test procedure under each hypothesis; use these to produce a simulated p-values cdf. These simulated distribution functions can be compared across tests to assess the relative performance of each test.

Results

Figure 4.12 shows the simulated p-values cdf, based on 500 simulations, for each of the four competing tests, under the null hypothesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. Here we are testing the effects of autocorrelation at lag 2, NAO and AO, which are all non-zero ($\beta_3 = 0.06$, $\beta_4 = -0.01$, $\beta_5 = 0.02$) in the model, and therefore the null hypothesis is clearly false. Thus, the method that corresponds to the greatest number of small p-values is preferred. It can be seen, therefore, that the new method outperforms Rotnitzky and Jewell's method. Also, the performance of the new method is comparable to that of the robust Wald and quasi-score tests.

Figure 4.12 suggests that in this case the new method is more powerful than Rotnitzky and Jewell's method, however, it does not provide us with any information as to whether or not the new method has the correct coverage properties i.e. if we apply the new test at size α, do we reject a true null hypothesis $100\alpha\%$ of the time. To achieve this we calculate a new set of 500 simulations, but this

time with $\beta_3 = \beta_4 = \beta_5 = 0$. Thus, the null hypothesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ is true. Figure 4.13 shows the cdf of the simulated p-values under each technique. This plot suggests that all methods have the correct coverage properties, since the p-values follow an approximate uniform distribution, which is to be expected under a true H_0. Thus, in this case the new method appears to have the correct coverage properties and is more powerful than Rotnitzky and Jewell's method.

Various other null hypotheses were tested using the same simulation procedure as outlined above. The results obtained were consistent with the previous results presented. Thus, in all cases the new method performed at least as well as all the other methods, and in some instances the performance of the new method was better. For this reason we do not present any further results.

4.5 Summary

We have developed a hypothesis testing technique which is appropriate for the modelling of clustered data with GLMs. The method adjusts the 'independence' likelihood ratio test to allow for the within cluster dependence. Using simulations we have shown that the power of the new test is comparable with other methods and in some instances better. In particular, it outperforms Rotnitzky and Jewell's likelihood ratio test when testing more than one parameter. There is also evidence to suggest that the new method is more powerful than the robust Wald test when dealing with correlated covariates. These results are in line with what would be expected theoretically.

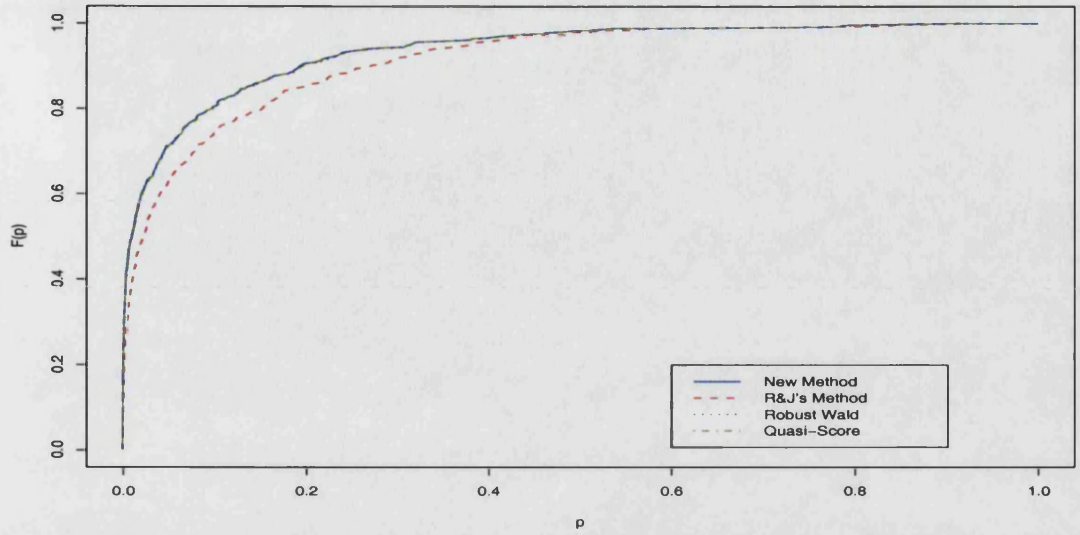

Figure 4.12: Simulated p-value cdf for each of the four competing tests under false $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. All simulated cdf's, except that corresponding to Rotnitzky and Jewell's method, are very similar and therefore difficult to distinguish.

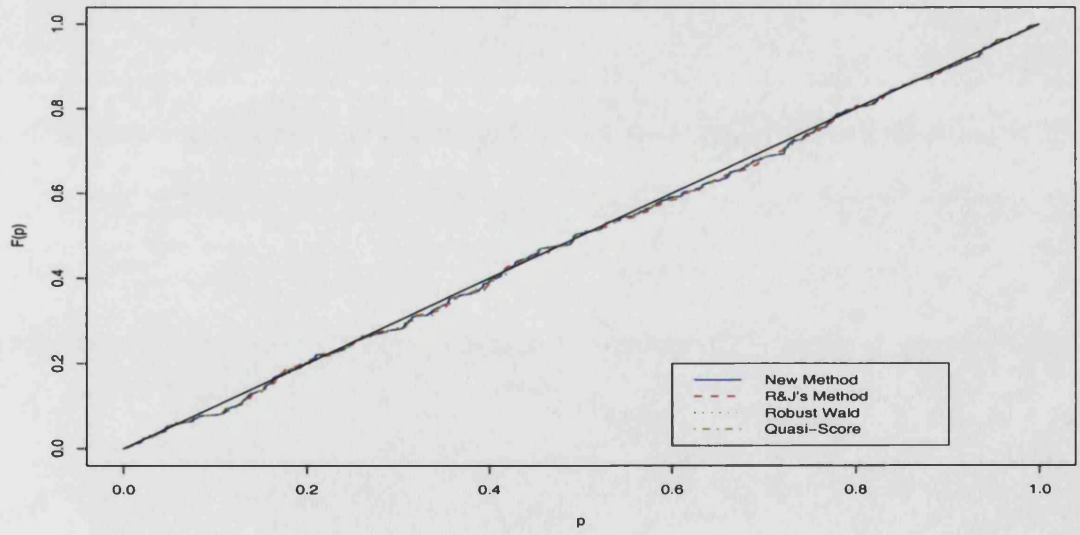

Figure 4.13: Simulated p-value cdf for each of the four competing tests under true $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. All four lines overlay each other and are therefore indistinguishable.

Chapter 5

Generalized estimating equations for large space-time data sets

Within this chapter we consider applying the generalized estimating equations (GEEs) methodology to large space-time data sets. We focus on the specific case where data are collected at a series of spatial locations and these spatial readings are repeated over time. Typically, under this setting, dependence exists both temporally and spatially. The temporal dependence arises from successive measurements being taken over time, while the spatial dependence results from neighbouring sites being subjected to similar conditions.

Section 5.1 reviews some of the existing techniques available for space-time data. In Section 5.2, we provide an overview of the GEE approach we adopt, and introduce the notation used throughout the remainder of the chapter. Sections 5.3 and 5.4 detail how we propose accounting respectively for the temporal and spatial dependence. Finally, Section 5.5 considers the use of the one-step estimator, which offers considerable computational advantages over the full GEE algorithm.

5.1 Alternative approaches for space-time data

In recent years, a significant amount of research has been undertaken on the development of statistical models for space-time data. This activity, in part, has been driven by the desire to analyse a wealth of data from the environmental sciences, which includes fields such as hydrology, meteorology and air pollution. Data sets from these application areas are typically very large in size and therefore advances in computing power have played a vital role in the development of these models.

Space-time data sets can take many forms. For example, spatial locations may form a regular grid, or they may be irregularly spaced. There may be few spatial locations and many time points, or conversely, many spatial locations and few time points. Typically, the analysis objectives will also vary across different data sets. For example, for one data set the identification of a new spatial location with specific characteristics may be the aim, whereas for another the prediction of future values at existing spatial locations may be the goal. Consequently, a vast array of space-time models have been developed to accommodate the various data set-ups and analysis objectives. Below we briefly highlight some of the approaches taken.

Haslett and Raftery (1989) considered the modelling of daily wind speeds in Ireland; their objective being to quantify the wind energy at a potential new site, while making use of long term records from other sites across Ireland. They applied a square root transformation to obtain approximately normal data, and deseasonalized the data to remove the seasonal component. Spatial correlation, short-memory temporal dependence and long-memory temporal dependence were allowed for through kriging, autoregressive moving average modelling and fractional differencing respectively. They claimed that their method was capable of capturing the main features of the data, despite the fact that some of the as-

sumptions they made are unlikely to hold in practice. For example, their model assumed constant seasonality across sites, the same univariate time series structure for each site, and an isotropic process in space.

An alternative approach for space-time data is detailed in Hughes et al. (1999). Here the authors aimed to investigate the relationship between precipitation occurrences and atmospheric circulation patterns, and produce simulations of precipitation occurrences. To achieve this, a nonhomogeneous hidden Markov model for precipitation occurrences was proposed. Temporal dependence was captured indirectly by defining several unobservable 'weather states' for the atmospheric processes that drive precipitation. A Markov assumption was implemented for the weather states, where the transition matrix was dependent upon a set of covariates, which were derived from atmospheric data obtained from a general circulation model (GCM). Conditional upon these weather states, precipitation was assumed to be temporally independent. Spatial dependence, on the other hand, was allowed for using an autologistic anisotropic model for the precipitation occurrences. The method presented did not attempt to model precipitation amounts, though specific extensions for this case were discussed. Estimation was undertaken via a modified EM algorithm, which used the technique of Monte Carlo maximum likelihood and assumed that the hidden states were missing data.

A different approach was taken by Stroud et al. (2001), who allowed for spatial variability through the use of a locally weighted mixture of regression surfaces, while temporal variability was modelled within a Gaussian state space modelling framework. This approach enabled seasonality, autoregressive terms and temporal trends to be incorporated into their model. Goodall and Mardia (1994) also considered a state space modelling approach and jointly employed the techniques of kriging and the Kalman filter to allow for spatial and temporal dependence. Brown et al. (2000) considered a stationary model based upon Gaussian 'blur-

ring'. This method used a non-separable space-time covariance function, and was recommended for the modelling of processes, such as air pollution, which disperse over time. Wikle et al. (1999) adopted a hierarchical Bayesian approach; like most Bayesian approaches to space-time data, this method is highly computationally intensive, making it inappropriate for large data sets. Finally, Brix and Diggle (2001) considered a class of models suitable for space-time point processes.

For the remainder of this chapter we focus on a GEE-based approach for space-time data. The method allows for temporal dependence via autoregressive based covariates and models the spatial dependence using techniques from geostatistics. Spatial non-stationarity and temporal trends can also be accounted for via spatial and time-varying covariates. Other key attributes such as seasonality can be incorporated into the covariates, and through the use of interactions, seasonality can easily be allowed to vary in space, for example. The method is computationally efficient and has considerable computational advantages over many of the techniques outlined above.

5.2 Overview of new generalized estimating equations approach for space-time data

GEEs are used to build regression models for clustered data (Chapter 3). Responses within a cluster are correlated, while clusters are assumed independent given the covariates. GEEs were originally applied to longitudinal data, where it is natural to view each individual as forming a cluster of correlated responses, with clusters assumed independent across individuals. For space-time data, however, dependence exists both temporally and spatially, and therefore the application of the GEE approach does not seem so natural. We propose a two stage approach to this problem. Firstly, the temporal dependence is accounted for via the covariates, using autoregressive terms. Then, conditional on the covariates, time

points correspond to independent clusters, and the spatial dependence can then be allowed for by adopting a spatial within cluster working correlation structure. Techniques from geostatistics (Journel and Huijbregts, 1978) are used to suggest working correlation structures of a spatial nature.

As we are now operating within a space-time setting, slightly different notation is adopted from that used previously in Chapter 3. Let Y_{ts} be a response random variable for time point t at spatial location s, where $t = 1, \ldots, T$ and $s = 1, \ldots, S$. Thus, the total number of responses is $N = T \times S$. Corresponding to each response Y_{ts} are the values $\mathbf{x}_{ts} = (x_{ts1}, \ldots, x_{tsp})^T$ of p covariates. Let $\mathbf{Y}_t = (\mathbf{Y}_{t1}, \ldots, \mathbf{Y}_{tS})^T$ denote the response vector of random variables for all spatial locations at time point t. Correlation exists between the elements of $\mathbf{Y}_t$, and also in general between $\mathbf{Y}_t$ and $\mathbf{Y}_l$ $(t \neq l)$. Further notation is the same as that used previously in Section 3.1.

5.3 Modelling temporal structure

5.3.1 Autoregressive approach

For simplicity, we begin by focusing on the single site case. If the time series of observations $\mathbf{y} = (y_1 \ldots, y_T)$ consisted of T independent elements, then the joint density could be expressed as a product of the independent densities $f(y_i)$, $i = 1 \ldots T$. Statistical analysis could then proceed in the standard manner detailed in Section 2.3. Unfortunately, due to the presence of temporal dependence, observations are correlated and this prevents us from expressing the joint density as a product of independent densities. Standard statistical analysis must therefore be adjusted to allow for this dependence. On the other hand, if we are able to obtain a factorization of the joint density which enables it to be expressed as a product of independent terms then analysis could proceed in the manner detailed

in Section 2.3.

Now the joint density can, as always, be factorized as a product of conditional densities

$$f(\mathbf{y}) = \prod_{t=1}^{T} f(y_t|\mathbf{P}_t), \tag{5.1}$$

where $\mathbf{P}_t = (y_{t-1}, y_{t-2}, \ldots, y_1)$ represents the history of past responses. Within a time series setting, it is common to simplify (5.1) by adopting a Markovian structure. Under a Markov model of order ψ, $\mathbf{P}_t$ in (5.1) reduces to $\mathbf{P}_t = (y_{t-1}, \ldots, y_{t-\psi})$. Thus, the joint density can be represented by a product of conditionally independent densities, where conditioning is undertaken on the recent past.

When fitting a GLM, one possible way of accounting for the temporal dependence structure described above, is to include some function of the ψ previous time point responses in the linear predictor. Thus, an autoregressive structure of order ψ, denoted by AR(ψ), is included in the covariates. By adopting this approach we can account for the temporal dependence at a specific site. In the multi-site case, the same theory can be applied to individual sites. Thus, conditional on the covariates, we are able to obtain independent time points, which correspond to clusters in the GEE framework. This idea of using an AR representation as a convenient and flexible approximation is well-established in other areas e.g. spectral estimation (Priestley, 1981). Also this approach naturally results in the loss of the first ψ responses at each spatial location. However, as we are dealing with large data sets this should be of little concern.

Within a GLM framework, inference for AR based covariates can be undertaken in the same manner as that used for other types of covariate (Fahrmeir and Tutz, 2001, Chapter 6). Therefore we propose the following approach to select the order ψ. Begin by fitting a GLM with an AR(1) structure included in the covariates. Test the significance of this effect, for example, by using a robust Wald test or the new likelihood ratio test outlined in Section 4.2.2, where an allowance

is made for the spatial dependence. Assuming the AR(1) effect is significant, fit a second GLM with $\psi = 2$, and test the significance of the second AR effect. Continue in this manner until additional AR effects are insignificant or until they are too small to make any practical difference. As part of this model selection process, various transformations of the AR terms may also be considered. In practice, using the same transformation as that used for the link function can work well, since the covariates are then on the same scale as the response.

5.3.2 Checking the autoregressive representation

A key assumption of GEEs is that the contributions from distinct clusters to the estimating equations (3.2) are uncorrelated. These contributions are expressed in terms of model residuals. It is imperative therefore, that having allowed for the temporal dependence in the manner described above, the assumption of independence of residuals across time points (or clusters) is checked. Since Liang and Zeger (1986) originally proposed estimating the within cluster correlations via Pearson residuals (see Section 3.1.4), it seems sensible to base a check for this independence on some function of the Pearson residuals. We propose producing a sample autocorrelation function plot of the Pearson residuals, obtained from a GLM fit which includes the chosen AR structure. Any temporal dependence remaining in the residuals should be evident from this plot.

When there are a large number of spatial locations, a single ACF plot should be produced, calculated over all spatial locations. The sample ACF at lag k can be calculated by

$$\tau_k = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T-k}(r_{ts} - \bar{r})(r_{t+k,s} - \bar{r})}{\sum_{s=1}^{S}\sum_{t=1}^{T}(r_{ts} - \bar{r})^2} \tag{5.2}$$

where r_{ts} is the Pearson residual corresponding to observation y_{ts}, and $\bar{r} = (\sum_{s=1}^{S}\sum_{t=1}^{T} r_{ts})/N$.

If the ACF plot shows no evidence of remaining temporal dependence, then a

GEE approach can be adopted, where time points represent clusters and spatial readings correspond to the within cluster measurements. If, however, temporal dependence is identified by the ACF plot, then an alternative representation of the temporal dependence needs to be found before a GEE approach can be taken.

5.3.3 Alternative approaches

Space-time processes, such as climate, by definition evolve over time. Therefore it is likely that observations taken at neighbouring sites at previous time points will also be correlated with the current observation at a particular site. To account for this additional temporal dependence a more sophisticated approach to that outlined above is needed. We propose replacing the standard AR terms with 'refined' AR terms which incorporate a neighbourhood structure. Weights can be allocated to each of the neighbours, reflecting the relative importance of each neighbour.

Up until now, autoregressive based covariates have been used to account for the temporal dependence. An alternative approach would be to adopt a moving average (MA) based structure. This would involve using covariates which are formed from transformations of past residuals instead of transformations of past observations. This approach, however, is more complicated to implement and computationally more intensive to fit. Besides, a linear MA process can be represented as an infinite order linear AR process (Chatfield, 2003, p. 43), which is essentially the approach we take, where all insignificant effects are set to zero. For these reasons we favour the AR representation.

5.4 Modelling spatial structure

Having allowed for the temporal dependence, we now focus on modelling the spatial dependence. Under the approach we adopt, the spatial dependence is allowed for via a within cluster working correlation structure. Therefore, within this section we investigate GEE working correlation structures which are applicable to spatially correlated data.

In Section 3.1.4 we outlined some of the most common working correlation structures assumed within GEEs. Of the three structures considered, the AR(1) and exchangeable are not appropriate, and the unstructured is parameter intensive, especially when there are many spatial locations. Therefore, below we consider alternatives that exploit the spatial nature of the data.

5.4.1 Isotropic structures

Typically, when modelling spatial data, observations recorded at nearby locations are more highly correlated than those observations taken further apart. For example, when modelling rainfall at a network of sites, typically observations taken from nearby sites are highly correlated because the sites are subjected to the same weather systems at the same time. It seems sensible therefore, to propose a spatial working correlation structure based on the distance between pairs of sites, such that pairs of sites the same distance apart are assigned the same working correlation value.

The first stage in the implementation of this working correlation structure is to calculate the Euclidean distances (u) between all pairs of sites:

$$u = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \tag{5.3}$$

where Δx and Δy are respectively the x- and y- separations of the sites.

The correlation between Y_{t,s_1} and Y_{t,s_2}, where sites s_1 and s_2 are located u spatial units apart, can be estimated as follows

$$\hat{\alpha}_u = \frac{1}{\hat{\phi}(GT-p)} \sum_{t=1}^{T} \sum \hat{r}_{t,s_1} \hat{r}_{t,s_2}, \tag{5.4}$$

where $\hat{r}_{t,s}$ denotes the Pearson residual for time point t at spatial location s (corresponding to $Y_{t,s}$), where the inner summation is over the set of all pairs of sites s_1, s_2 located u spatial units apart, and G denotes the number of pairs in this set. Also the estimation of the dispersion parameter ϕ is obtained by $\hat{\phi} = \{1/(N-p)\} \sum_{t=1}^{T} \sum_{s=1}^{S} \hat{r}_{ts}^2$. The working correlation matrix $\mathbf{R}(\hat{\boldsymbol{\alpha}})$ is determined by the vector $\hat{\boldsymbol{\alpha}}$, which consists of the elements $\hat{\alpha}_u$ and is of dimension equal to the number of distinct pairwise distances.

For large cluster sizes, particularly when sites are irregularly spaced, the above approach can result in the dimension of $\boldsymbol{\alpha}$ being substantial. One way to reduce the dimension of $\boldsymbol{\alpha}$ is to group together pairs of sites with similar distances as opposed to grouping together pairs of sites with identical distances (Venables and Ripley, 1994). Groups of equal width can be set up to cover the entire range of distances and each pair of sites is then allocated to a group. All pairs of sites in the same group are then used to calculate a single $\hat{\alpha}_u$ using (5.4) (where a tolerance is now placed on u). This technique can reduce the dimension of $\boldsymbol{\alpha}$ substantially, the extent of which naturally depends on the width chosen for each group.

As mentioned earlier, inter-site correlations will typically decrease with distance. Therefore, an alternative way of reducing the number of parameters needed for specification of $\boldsymbol{\alpha}$ is to smooth the individual elements $\hat{\alpha}_u$ using a spatial correlation function (Albert and McShane, 1995). Various families of correlation function have been proposed in the spatial statistics literature (Cressie, 1991). Three of the most common are outlined below.

a) Powered Exponential: The powered exponential family of correlation func-

tions is given by

$$\rho(u) = \exp\left\{-(u/\varphi)^{\xi}\right\}, \tag{5.5}$$

where u represents distance and φ and ξ are parameters to be estimated. This correlation function is defined for $\varphi > 0$ and $0 < \xi \leq 2$.

b) Matérn: The Matérn family of correlation functions is defined by

$$\rho(u) = \left\{2^{\xi-1}\Gamma(\xi)\right\}^{-1} (u/\varphi)^{\xi} K_{\xi}(u/\varphi), \tag{5.6}$$

where φ and ξ are parameters to be estimated, $K_{\xi}(.)$ denotes the modified Bessel K function of fractional order ξ, and $\Gamma(.)$ is the gamma function.

c) Spherical: The spherical family of correlation functions is defined by

$$\rho(u) = \begin{cases} 1 - \frac{3}{2}(u/\varphi) + \frac{1}{2}(u/\varphi)^3 & \text{for} \quad 0 \leq u \leq \varphi \\ 0 & \text{for} \quad u > \varphi \end{cases} \tag{5.7}$$

where φ is the only parameter to be estimated ($\varphi > 0$). This function differs from the previous two functions in that it reaches zero within a finite range ($u = \varphi$). This family of functions is also less flexible than the previous two as it is only a function of a single parameter.

For some processes, the spatial correlation observed at arbitrarily small distances will be less than one, due mainly to measurement error and very small scale effects. In these instances, each of the above correlation functions can be extended to include what is termed a 'nugget effect', to capture this characteristic.

There are several advantages to smoothing the working correlation matrix with one of the correlation functions outlined above. Firstly, in general, convergence of the GEE algorithm is faster. This is because only the one or two parameters which determine the correlation function need to converge, as opposed to each of the individual elements $\hat{\alpha}_u$. This increased efficiency is particularly important when modelling large data sets. A second convergence related benefit

is that smoothing alleviates the problem of non-convergence associated with the estimation of many parameters (Dobson, 2002). A further advantage of smoothing is that the potential problems of bias associated with the estimation of many nuisance parameters is avoided (Liang and Zeger, 1995).

Selecting a structure

If one of the correlation functions outlined above is to be used to parameterize the working correlation structure, then a specific structure needs to be selected from the available options. To achieve this, all candidate correlation functions should be fitted to the Pearson residual correlations obtained from a GLM fit, using a non-linear least squares algorithm. The various fits can then be represented graphically to ascertain whether any of the candidate functions provides a good fit to the correlations. When more than one correlation function provides an adequate fit, the function which minimizes the sum of squared errors between the fitted correlations and the individual estimates for each pair of sites, should be selected. Note that, for the Matérn family, estimation of ξ is difficult and it is usual therefore to carry out a grid search for this parameter, where typical values are $\xi = 1, 1.5, 2$.

Once a structure has been selected, the full GEE algorithm can be implemented, where the non-linear least squares fitting technique can be adopted to estimate the spatial correlation function parameters, at each iteration.

5.4.2 Anisotropic structures

The above isotropic structures assume that the correlation between pairs of sites is a function of distance only. For many spatial processes, however, the correlation between pairs of sites will not only depend on the distance which separates them but also upon their orientation. Such processes are known as anisotropic. Within

this section we focus on the special form of anisotropy known as 'geometric'.

In the case of an isotropic process the interpair correlations decay at the same rate in all directions and therefore the correlation contours are represented by circles. In geometric anisotropy, however, as the correlations persist to a greater extent in one direction, the correlation contours form ellipses. The spatial process is most highly correlated in the direction of the major axis of the ellipse, while the correlation is least in the direction of the minor axis.

A geometrically anisotropic process can be transformed into an isotropic one by applying two co-ordinate transformations; the isotropic structures above can then be applied under the transformed co-ordinate system. The first transformation rotates the set of axes such that they are aligned along the major and minor axis of the ellipse. The second then rescales the minor axis to equal the length of the major axis. Using these transformations and (5.3) we find that the transformed distance between pairs of sites is given by

$$u' = \sqrt{(\Delta x')^2 + (\Delta y')^2}, \tag{5.8}$$

where

$$(\Delta x', \Delta y') = (\Delta x, \Delta y) \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & r^{-1} \end{pmatrix}, \tag{5.9}$$

r denotes the anisotropy ratio of the ellipse which is defined to be the length of minor axis divided through by the length of major axis ($0 < r \leq 1$) and ω denotes the anisotropy angle of the major axis of the ellipse, measured in a anti-clockwise direction from the East.

The correlation functions defined in (5.5) to (5.7) can be extended to allow for anisotropy. For example, the powered exponential function (5.5) can be extended as follows

$$\rho(u') = \exp\left\{-\left(\frac{u'(\omega, r)}{\varphi}\right)^{\xi}\right\}, \tag{5.10}$$

which is now a function of the four parameters $(\varphi, \xi, r, \omega)$, where φ and ξ are the powered exponential smoothing parameters, and r and ω are the anisotropy

transformation parameters. All four parameters can be estimated simultaneously from the individual inter-site residual correlations during the fitting process, using non-linear least squares.

Detecting an anisotropic process

An anisotropic correlation function, such as the one given in (5.10) is clearly more complicated than its isotropic counterpart (5.5). Therefore fitting an anisotropic structure is only recommended if there is clear evidence against an isotropic structure. One way of detecting an anisotropic process was outlined by Cressie (1991). For each pair of sites calculate their orientation and their distance. Place each pair of sites into one of n groups depending on their orientation. For example, if $n = 4$ then there are four groups each covering 45°, centered on 0°, 45°, 90° and 135° (where 0° corresponds to the north and angles are standard compass bearings). Within groups, pairs of sites are then grouped according to their distance. Correlations in the Pearson residuals, obtained from a GLM fit, are then calculated. A separate isotropic correlation function, such as the powered exponential (5.5), is then fitted for each of the four orientation groups. Plots of the correlations against distance should also be produced for each orientation group. If the estimated correlation parameters differ significantly across the four orientation groups, and the plots show clear evidence of a different correlation decay rate in the various directions, an anisotropic structure such as (5.10) should be adopted.

5.5 The one-step estimator

As an alternative to the spatial GEE working correlation structures outlined in Section 5.4, a working independence structure could be adopted, corresponding to IEEs. The spatial GEE approach has the advantage over IEEs that it ex-

plicitly allows for the spatial correlation during the fitting process. One of the drawbacks of the spatial GEE approach, however, is that it is computationally more expensive to fit than its IEE counterpart, which is a very important consideration since large data sets are the focus. A compromise between the two approaches is provided by the spatial GEE one-step estimator, which initially fits a GLM and then carries out one iteration of the GEE algorithm, using a spatial working correlation structure. Thus, it allows for the spatial dependence during the fitting process, but is computationally less intensive than the full spatial GEE method.

5.6 Summary

Within this chapter we have considered modelling space-time data within a generalized estimating equations framework. The method proposed has many appealing properties, for example, it can be used to model both non-stationary and geometrically anisotropic processes. The method can also be applied to non-lattice data, and is computational efficient to implement.

The theory developed in this chapter is applied to a climate data set in Chapter 6. Both the full GEE algorithm and the one-step estimator are considered.

Chapter 6

Climate case study

In this chapter we analyse a climate data set within a generalized linear modelling framework. This enables us to bring together and demonstrate many of the techniques discussed throughout the thesis. In particular, we are able to apply the new hypothesis testing technique of Chapter 4 and the generalized estimating equations (GEE) methodology of Chapter 5, to a space-time data set. In addition, some of the broader issues relating to the application of the GLM methodology to climate data are discussed.

Section 6.1 introduces the data set and outlines the aim of the study. Some preliminary analysis of the data is undertaken in Section 6.2 and a GLM approach to modelling the data is adopted in Section 6.3. An alternative analysis is undertaken in Section 6.4, which is based on a GEE approach. The two modelling approaches are compared in Section 6.5 and finally, a summary of the work undertaken is provided in Section 6.6.

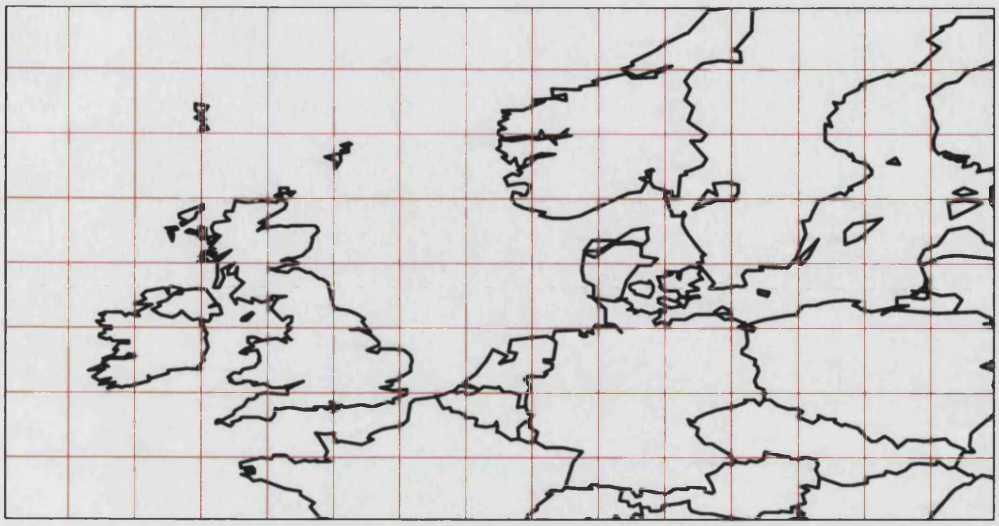

Figure 6.1: Map of study area, with NCEP grid overlaid.

6.1 Introduction

6.1.1 The data set

The data relate to wind speeds and have been taken from the US National Center for Environmental Prediction's (NCEP) reanalysis project (Kalnay et al., 1996). The study area is northwestern Europe, which is divided into 120 grid nodes within the region 47.5°-65°N and 12.5°W-22.5°E, as can be seen in Figure 6.1. Each grid node covers an area of 2.5° latitude × 2.5° longitude and for each grid node, instantaneous wind speeds are provided every 6 hours. Here we study the maximum of the four daily readings, which is labelled 'daily maximum wind speed' or DMWS for short. Data are available for the 41 year period 1958-1998. This corresponds to 14,975 observations per location and almost 1.8 million observations in total.

6.1.2 Aim of study

Extreme wind speeds have led to great devastation in recent years, resulting in substantial human and economic losses. Insurance companies are particularly interested in being able to model wind speeds as they can incur substantial payouts under extreme conditions. This case study has been motivated as a direct result of these concerns.

Interest lies in identifying the important factors which affect the wind speed process, and understanding how the various physical components interact with one another. Generalized linear models provide a framework for building models to explain such processes. Also, once a suitable GLM has been identified, it can then be used to study extreme events via simulation. Within this chapter, we focus on the identification of a GLM which is capable of explaining the important factors affecting wind speeds within the area under study.

6.2 Preliminary analysis

We begin by carrying out some preliminary graphical analysis on the data. This provides an overview of the data, which will help guide the model building process later on.

6.2.1 Site specific properties

Figure 6.2 shows histograms of the DMWS values at various locations. The top left histogram has been produced for the DMWS values at all locations. The three other histograms represent the DMWS values at specific locations. All of the histograms are positively skewed. Possible candidate distributions for modelling such data include the gamma, Weibull and log-normal.

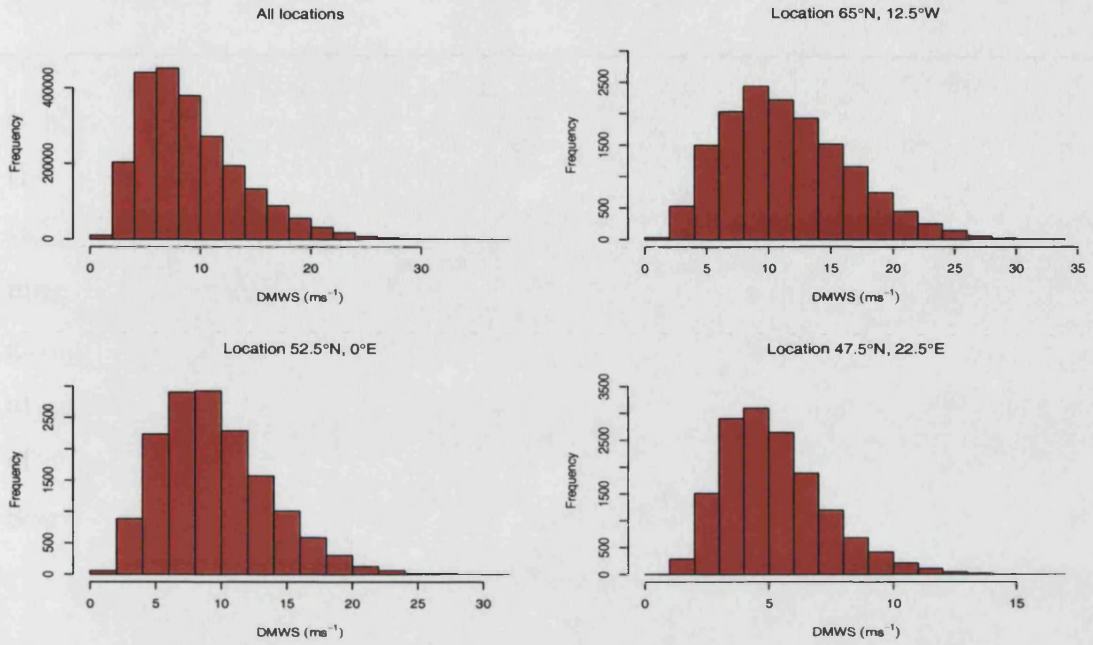

Figure 6.2: Histograms of DMWS values over all locations and three specific locations, 1958-1998.

Figure 6.3 shows the mean of the 14,975 DMWS values for each location. There is clear evidence of a land-sea effect, with the values over the sea generally being higher than those over land. This is nicely captured in the Baltic Sea where the higher values in the sea are surrounded by lower values on the surrounding land. The average values also tend to increase as we move in a north-westerly direction. However, as most of the north-west locations fall over the sea and most of the south-east locations fall over land, it is difficult to separate the land-sea effect from the north-west south-east effect. The greatest mean DMWS values are found to the west of the region between the latitudes 55°N-60°N. This corresponds to the position of the North Atlantic storm track, which is the path most frequently followed by cyclones in the area.

Figure 6.4, which has been produced in a similar fashion to Figure 6.3, shows for each site the standard deviation of DMWS values. This standard deviation

plot follows a very similar pattern to that for the mean. This suggests the standard deviation increases with the mean, which should result in a fairly constant coefficient of variation (standard deviation/mean) over locations. A plot of the coefficient of variation has been produced in Figure 6.5 and it can be seen that the coefficient of variation is indeed roughly constant across locations, with most values between 0.4 and 0.45. This plot suggests that the gamma distribution may be appropriate for modelling this data set, as one of the assumptions of the gamma distribution within the GLM framework is a constant coefficient of variation (see Section 2.2). However, note that Figure 6.5 does shows some evidence of clustering of similar sized values, for example the grouping of high values over Southern Germany.

Figure 6.6 shows the maximum DMWS value recorded over the 41 year period at each location. A similar pattern is evident to that for the means and standard deviations.

6.2.2 Seasonality

Plots of the monthly mean DMWS values by location have been produced in Figure 6.7. These provide an insight into seasonality and, as expected, show that winter months are on average windier than summer months. Further inspection of these plots suggests that seasonality is more pronounced in the north-west of the region than in the south-east. The land-sea and north-west south-east effects are again evident, along with the North Atlantic storm track.

Figure 6.8 shows the standard deviation of DMWS by location for each month of the year. A similar pattern to that for the means is evident. In Figure 6.9, plots of the monthly coefficient of variation of DMWS by location have been produced. Again, the coefficient of variation is fairly constant across locations in each of the monthly plots. However, the same clustering of high values occurs

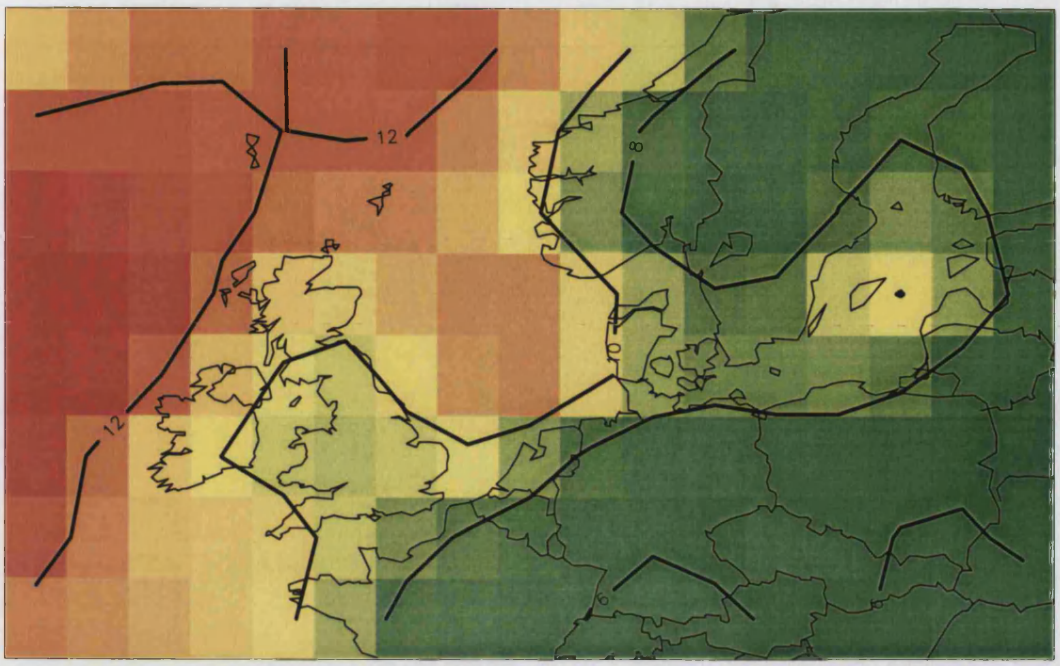

Figure 6.3: Mean DMWS values (ms^{-1}) over NCEP grid, 1958–1998. Contours are at 2 ms^{-1} intervals.

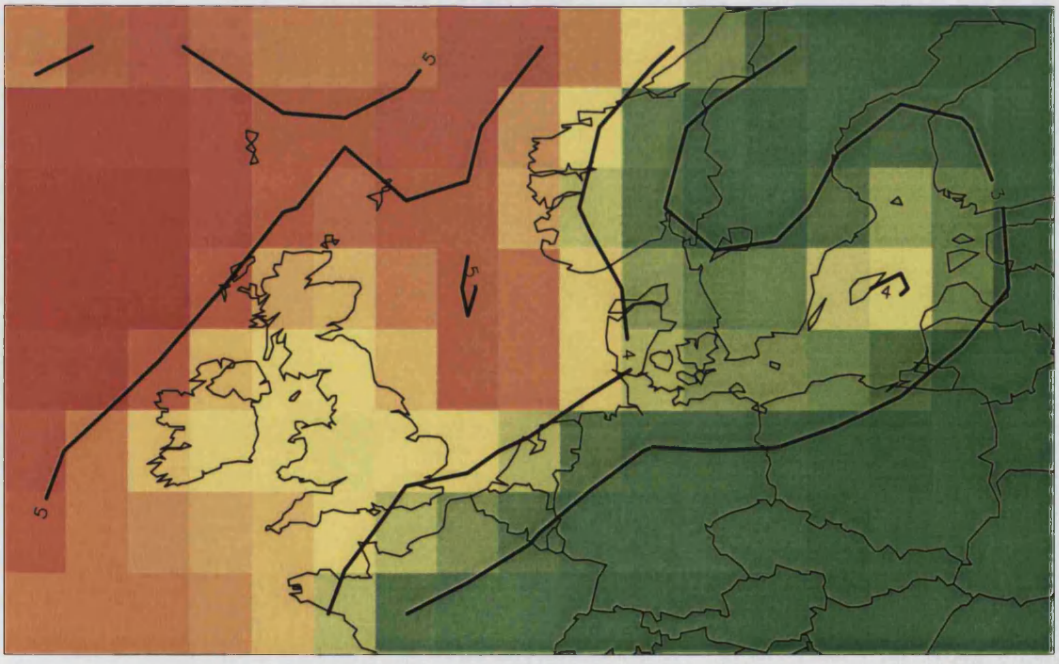

Figure 6.4: Standard deviations of DMWS values (ms^{-1}) over NCEP grid, 1958–1998. Contours are at 1 ms^{-1} intervals.

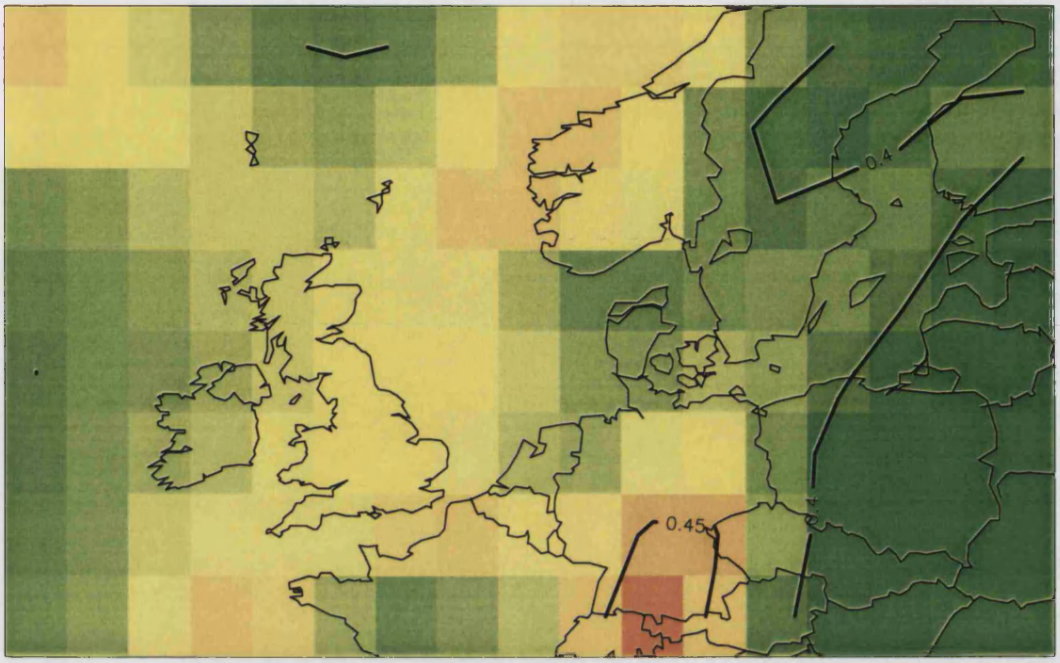

Figure 6.5: Coefficients of variation of DMWS values over NCEP grid, 1958–1998. Contours are at intervals of 0.05.

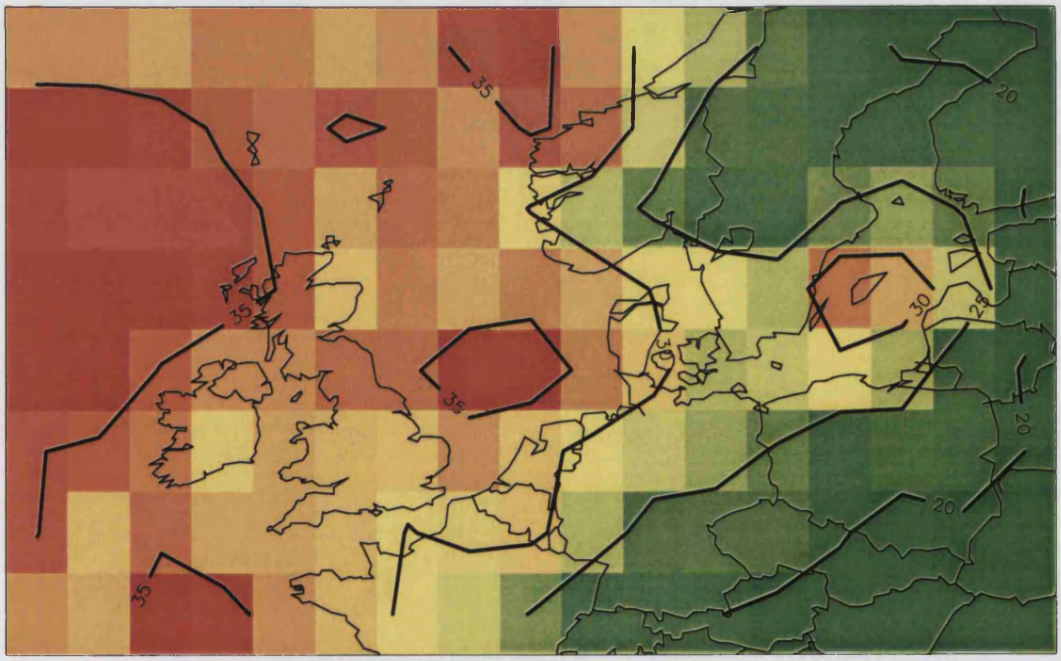

Figure 6.6: Maximum DMWS values (ms^{-1}) over NCEP grid, 1958–1998. Contours are at 5 ms^{-1} intervals.

in each of the winter months for Southern Germany. In Figure 6.10, plots of the monthly overall maximum DMWS value by location have been produced. Again, a similar pattern to that for the means and standard deviations is evident.

In Figure 6.11 means and standard deviations for each day of the year (calculated over 41 years) have been plotted for two specific locations; the most north-westerly location (65°N,12.5°W) and the most south-easterly location (47.5°N,22.5°E). Clearly, seasonality is much more evident for the north-westerly location. This figure, together with Figure 6.7, suggests the presence of an interaction between spatial location and seasonality.

6.2.3 Trend

A time series plot of the annual mean DMWS values, taken over all locations, is shown in Figure 6.12. A simple linear regression of annual mean value on year has been calculated and the fitted regression line is shown. The slope of the regression line is positive and significant at the 5% level, suggesting that DMWS values have increased over time. Admittedly, this is a rather crude piece of analysis, however, it does highlight the existence of possible long term trends in the data.

Figure 6.13 has been constructed in a similar fashion to Figure 6.12, however, this time we focus on each location separately. For each location, a linear regression of annual mean on year is performed and the gradient of the slope, multiplied by ten to obtain a decadal trend, is plotted. Naturally, the values above zero represent an increasing trend over time, whereas values below zero represent a decreasing trend. Note, however, that no test for the significance of these slopes has been undertaken. Clear regional variability is evident from this plot. For example, an increasing trend of 0.3ms^{-1} per decade has been experienced over some parts of the North Sea, while a decreasing trend of up to 0.2ms^{-1} per decade has been experienced over parts of continental Europe. These

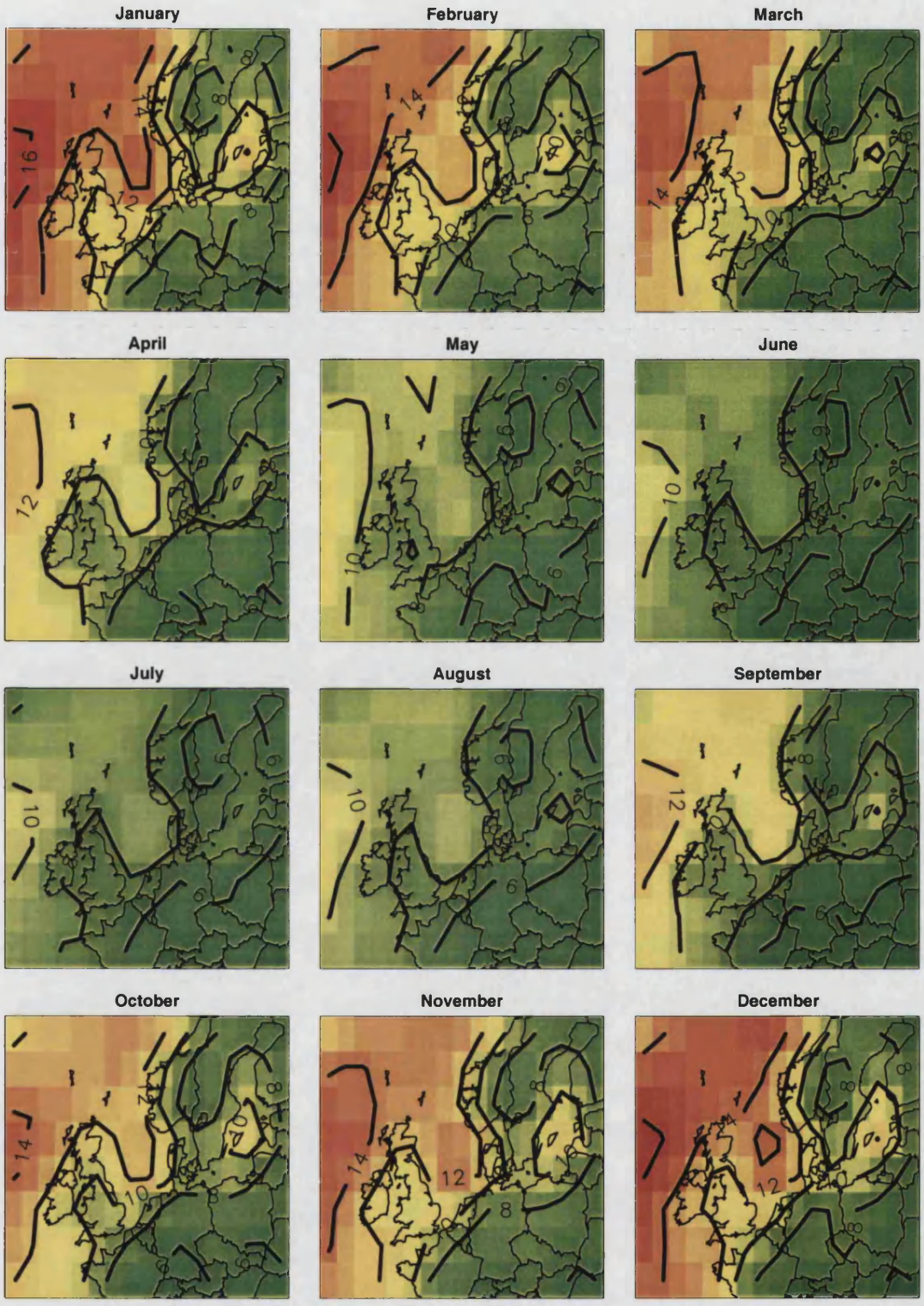

Figure 6.7: Monthly mean DMWS values (ms^{-1}) over NCEP grid, 1958–1998. Contours are at $2ms^{-1}$ intervals.

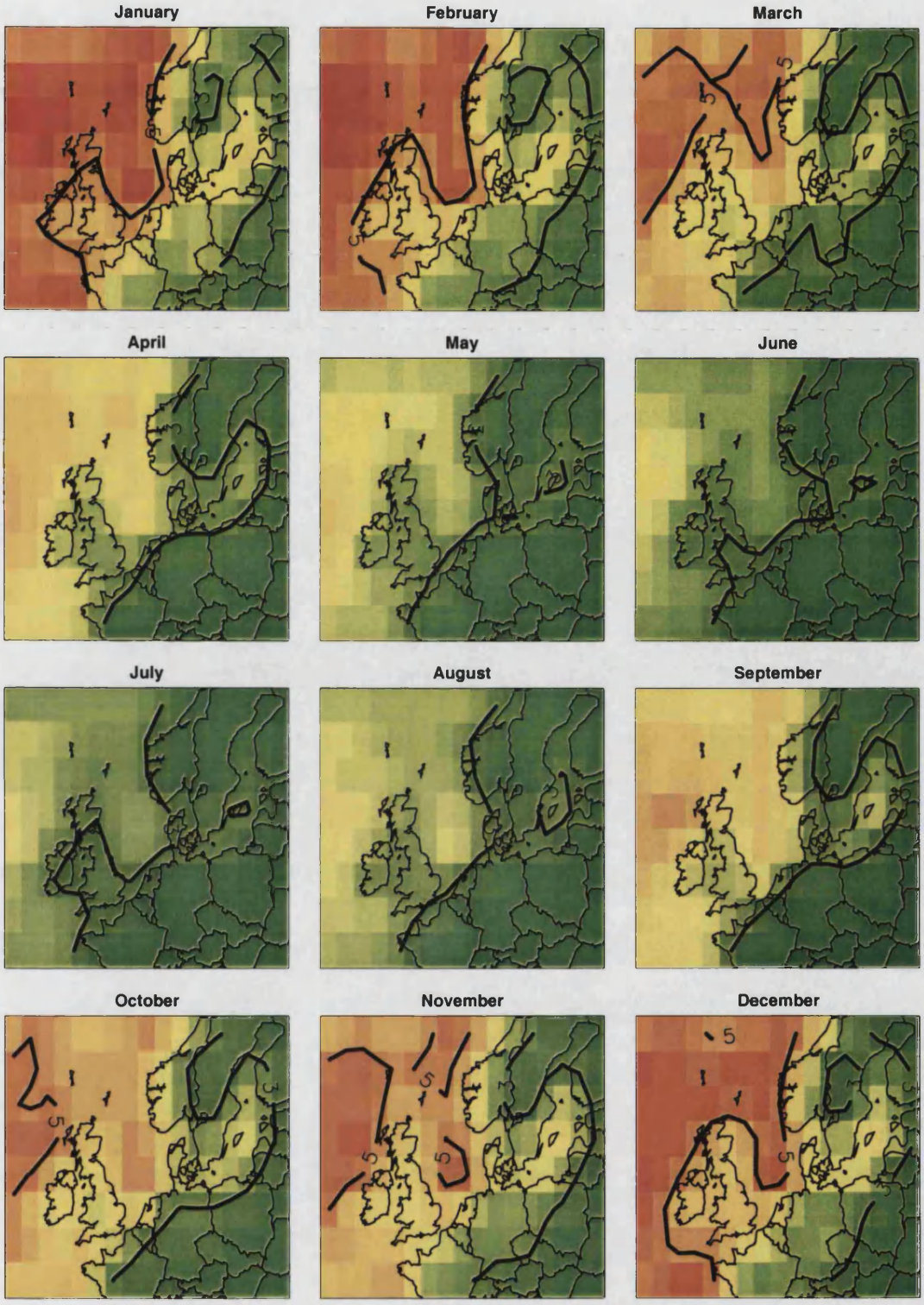

Figure 6.8: Monthly standard deviations of DMWS values (ms^{-1}) over NCEP grid, 1958–1998. Contours are at $1ms^{-1}$ intervals.

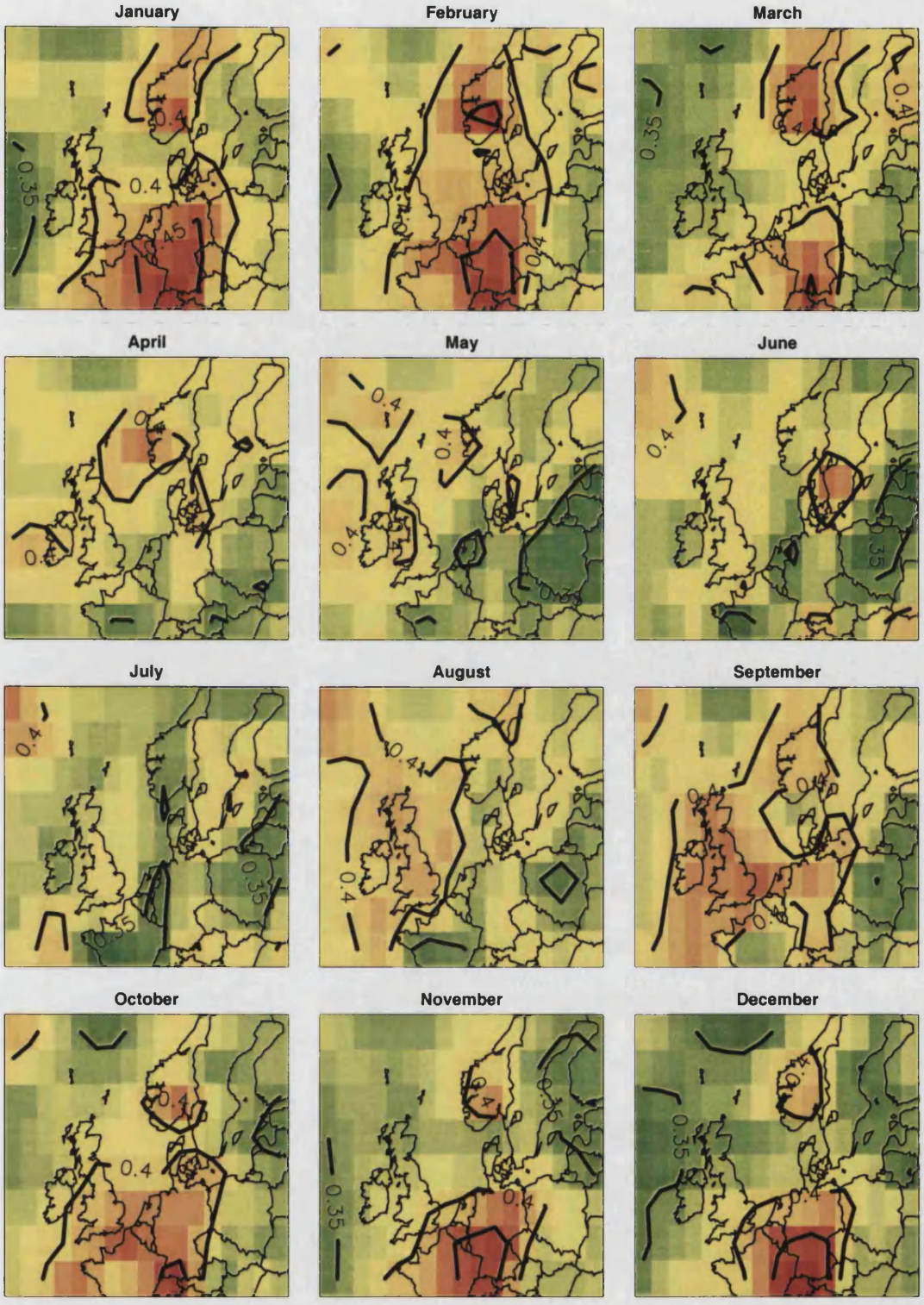

Figure 6.9: Monthly coefficients of variation of DMWS values over NCEP grid, 1958–1998. Contours are at intervals of 0.05.

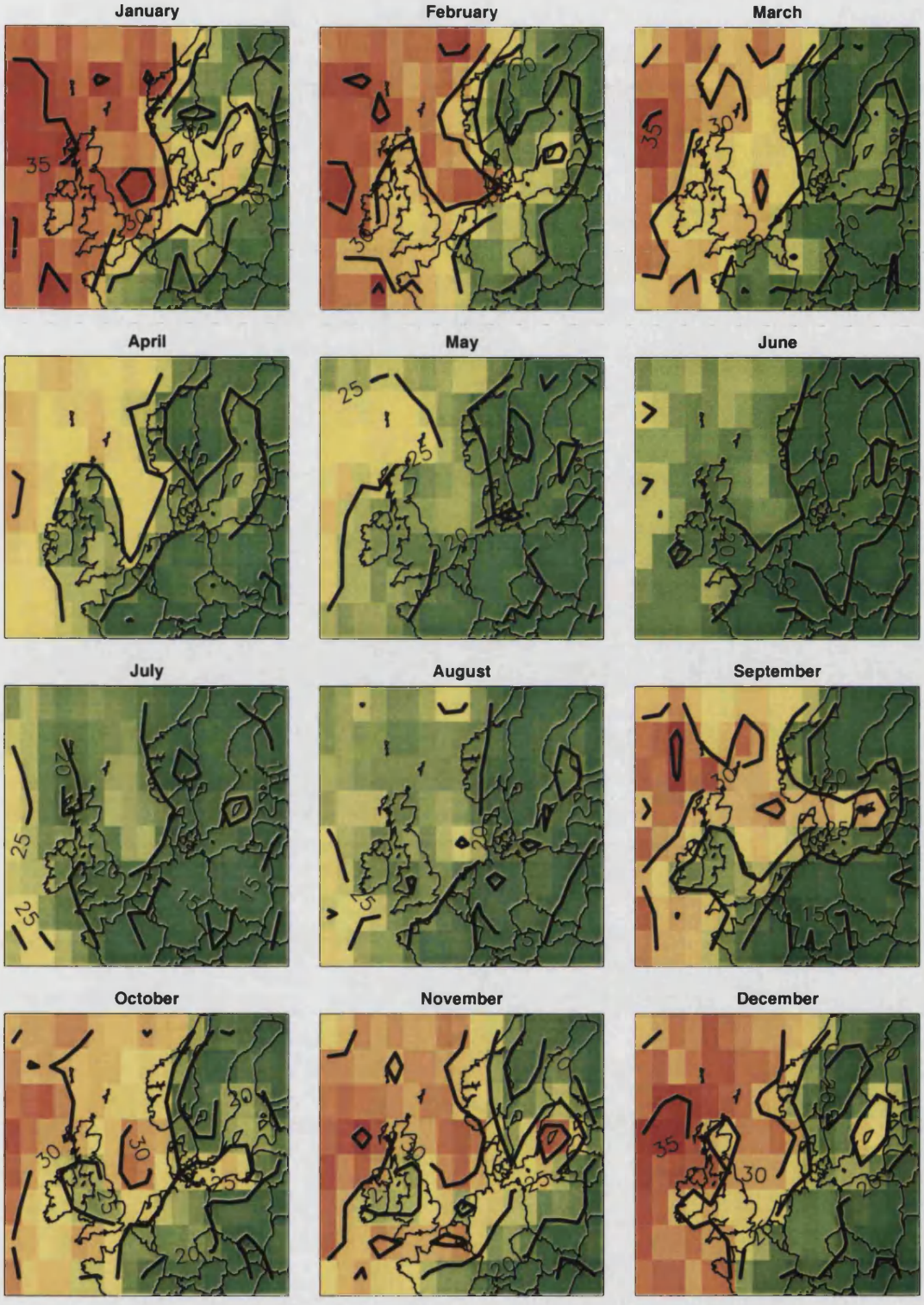

Figure 6.10: Monthly maximum DMWS values (ms^{-1}) over NCEP grid, 1958–1998. Contours are at $5ms^{-1}$ intervals

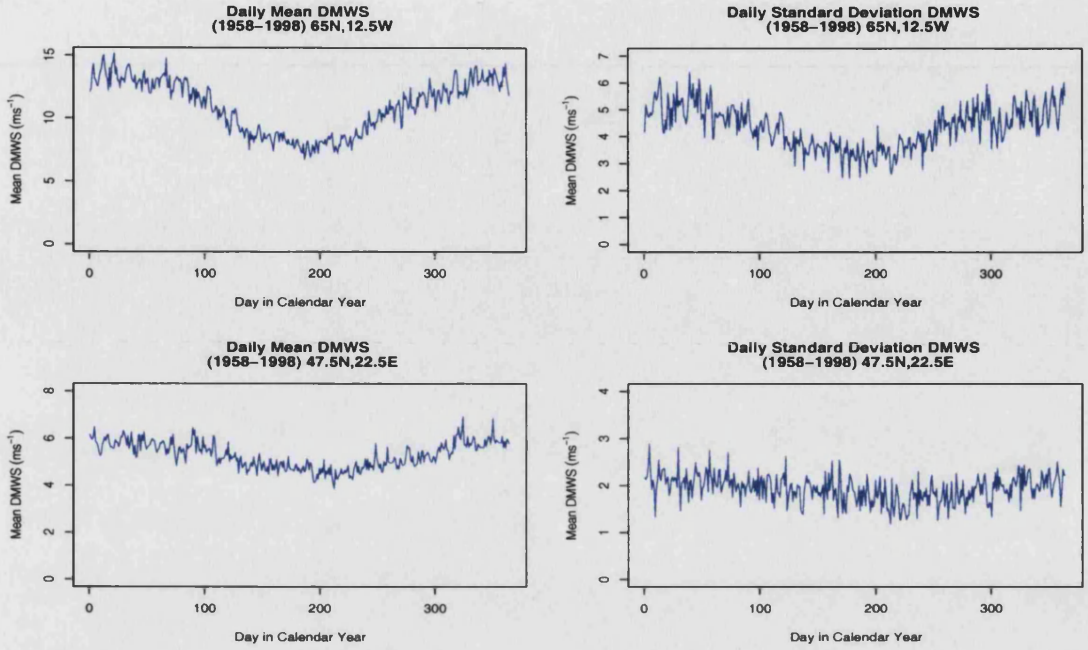

Figure 6.11: Daily mean and standard deviation DMWS for most north-westerly location (65°N,12.5°W) and most south-easterly location (47.5°N,22.5°E).

extreme trends, equate to approximately a 10% change in mean wind speeds over the 41-year study period.

In summary, the preliminary analysis undertaken above has highlighted some important features of the data, which will help to guide the model fitting process. For example, we have identified affects due to seasonality, land-sea, geographical location and various interactions. There also appears to be long-term trends in the data; these vary regionally in both their magnitude and direction.

6.3 Generalized linear modelling approach

We now model the data set using a univariate generalized linear model. Within this framework, temporal dependence is allowed for by including autoregressive

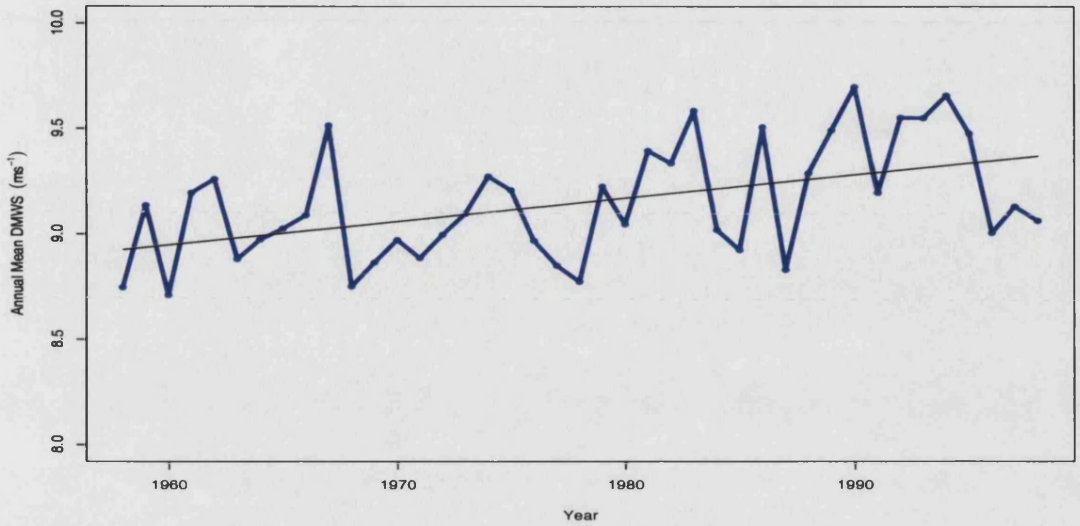

Figure 6.12: Time series of annual mean DMWS values (ms^{-1}) over all locations, 1958-1998.

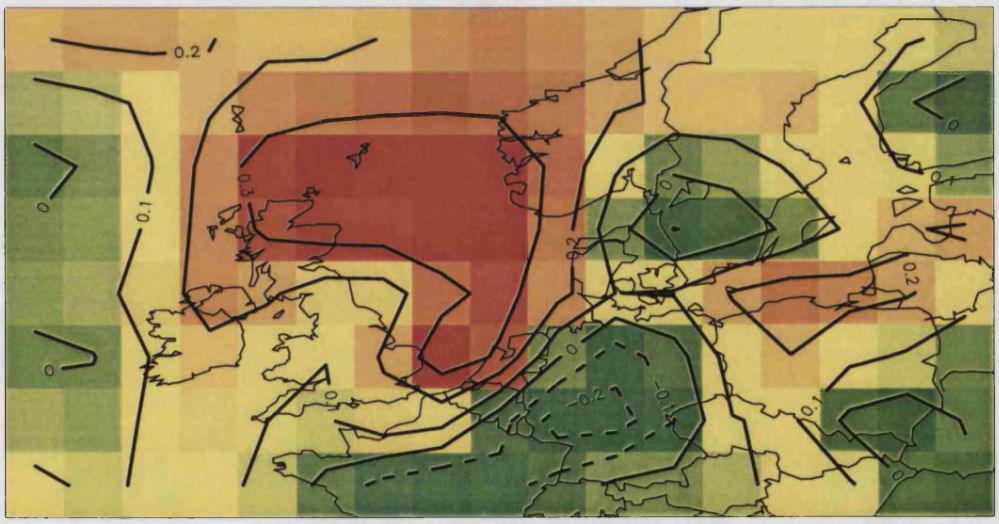

Figure 6.13: Decadal trends in annual mean DMWS values at each NCEP grid point, 1958–1998. Units are average increases in DMWS (ms^{-1}) per decade. Contours are at intervals of 0.1.

based covariates in the model (see Section 5.3). Spatial dependence, however, is not fully accounted for and therefore inference needs to be adjusted to reflect this fact. Therefore, this approach corresponds to the use of independence estimating equations (IEEs, see Section 3.1.2), where time points correspond to clusters and the within cluster dependence is a function of spatial location.

A distribution for the response variable DMWS must be chosen. Figure 6.2 suggests that the distribution chosen should be continuous, non-negative and positively skewed, and probably the most natural choice of distribution within the exponential family is the gamma. Figures 6.5 and 6.9 also suggest a fairly constant coefficient of variation, corresponding to a constant dispersion parameter for the gamma model within a GLM context. Historically however, the Weibull distribution appears to be the most widely used distribution for wind speed analysis (Conradsen et al., 1984; Tuller and Brett, 1984). Unfortunately, the Weibull does not fit naturally into the exponential family of distributions. However, the GLM algorithm may be extended to accommodate its implementation, as detailed in Section 2.7.1. A consequence of this extension, is that the Weibull model is computationally more demanding than the gamma model to fit. Within this section both the gamma and Weibull models are considered.

6.3.1 Gamma model

a) Modelling strategy

The gamma pdf is given by

$$f(y;\mu,\nu) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1}\exp\left(-\frac{\nu y}{\mu}\right), \quad 0 \leq y < \infty, \quad \mu,\nu > 0. \tag{6.1}$$

Within a GLM context, the observation y_{ts}, taken at time point t and spatial location s, is assumed to be a realization from a gamma distribution with mean μ_{ts} and constant shape parameter ν. A log link between the means μ_{ts} ($t = 1, \ldots, T$,

$s = 1, \ldots, S$) and their corresponding linear predictors $\eta_{ts} = \mathbf{x}_{ts}^T \boldsymbol{\beta}$ is assumed, to ensure that all fitted values are positive. Thus,

$$\log(\mu_{ts}) = \mathbf{x}_{ts}^T \boldsymbol{\beta}, \tag{6.2}$$

where $\mathbf{x}_{ts}$ are the values of p covariates corresponding to y_{ts} and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients to be estimated. Hence, β_j ($j = 1, \ldots, p$) measures the average multiplicative effect of the jth covariate upon the response variable DMWS.

Each of the candidate covariates under consideration can be assigned to one of four broad categories of effects, which are outlined below.

- **Geographical effects**: representing systematic regional variability. The preliminary analysis undertaken suggested that the following predictors should be included in the model: a land-sea indicator (a variable taking the value 1 for land and 0 for sea), functions of altitude, functions of latitude and functions of longitude. Altitude, latitude and longitude will be represented using Legendre polynomials (Abramowitz and Stegun, 1965). One advantage of using a Legendre polynomial representation is that it produces predictors that are approximately uncorrelated, thus benefiting model selection via Wald tests (Chandler, 1998).

- **Seasonal effects**: represented by sine and cosine waves with periods of 1 year and 6 months. Constant adjustments for individual months are also considered.

- **Autocorrelation effects**: represented by including previous days DMWS values as predictors, in the form $\log(1 + y_{t-k,s})$. Our experience is that this form is more efficient than the more natural $\log y_{t-k,s}$, perhaps because there are some values near zero in the data set, which may bias the results from a standard log-transform.

- **External effects**: indices representing different aspects of global climate. Amongst others, indices for hemispheric average annual temperatures, regional sea surface temperatures and teleconnection patterns such as the North Atlantic Oscillation (NAO) are considered. Table 6.1 provides a full list of the external effects considered, along with their abbreviations and brief descriptions. The data for most indices were obtained from the US NAOO Network Information Center (NNIC). The exceptions are AO, taken from Thompson and Wallace (1998); and SOI, SHT, NHT, NAT, and SAT which are from the Climatic Research Unit, University of East Anglia. Here we consider annual indices only.

In addition to the main effects outlined above, interactions between covariates will also be considered. Some of the external effects, for example, are likely to vary with spatial location and season, and therefore the inclusion of interactions will be necessary to represent this feature of the data.

With such a large complex data set, and vast array of potential covariates, a systematic approach to model selection is required. The approach taken here is largely drawn from Yan et al. (2002). This approach begins by introducing the most obvious covariates into the model first. These 'obvious' covariates can be identified from the preliminary analysis undertaken in Section 6.2, and include geographical, seasonal and autocorrelation effects. Covariates are added into the model on a one-by-one basis and inference is performed, at least in part, using robust Wald tests. Any terms added and subsequently found to be non-significant are discarded. Once all of the obvious main effects have been included, plausible interactions are then added into the model and tested. For example, if it is believed that temporal autocorrelation varies with season and landmark (land-sea indicator) then a three-way interaction involving these effects would be considered, to capture this aspect of the data. Only interactions of up to three factors are considered, since higher order interactions tend to be trivial and difficult to

Index name and abbreviation	Interpretation of positive value
Arctic Oscillation (AO)	Deeper polar vortex
Asian Summer Pattern (AS)	Positive pressure anomalies in summer over subtropical Asia and Africa
East Atlantic Pattern (EA)	Similar to NAO but with pressure anomaly centers shifted southward, except May–August
East Atlantic Jet Pattern (EAJ)	Enhanced westerlies over the NE Atlantic and Europe, April–August
East Atlantic / West Russia Pattern (EAWR)	Positive pressure anomaly center in the NE Atlantic and negative in W Russia, except June–August
East Pacific Pattern (EP)	Pronounced NE extension of Pacific Jet stream towards NW America, except August–September
North Atlantic Oscillation (NAO)	Sharper pressure gradient between Greenland/Iceland Low and Subtropical High in the North Atlantic
North Atlantic Temperature (NAT)	Warmer sea surface within 5°–20°N, 30°–60°W
Northern Hemisphere Temperature (NHT)	Warmer Northern Hemisphere
North Pacific Pattern (NP)	Southward shift and intensification of Pacific Jet, March–July
Pacific / North America Pattern (PNA)	Wavy pressure pattern: positive anomalies in subtropical Pacific and negative in the Aleutian, except June–July
Polar / Eurasia Pattern (POL)	Positive pressure anomalies in the polar and negative in Europe and northeastern China, winter
Pacific Transition Pattern (PT)	Wavy pressure anomalies from the Gulf of Alaska eastward to the Labrador Sea, with a prominent positive center over the western US, May–August.
South Atlantic Temperature (SAT)	Warmer sea surface within 30°W–10°E, 0°–20°S
Scandinavia Pattern (SCA)	Positive pressure anomaly sometimes due to blocking anticyclones over Scandinavia, except June–July
Southern Hemisphere Temperature (SHT)	Warmer Southern Hemisphere
Southern Oscillation Index (SOI)	Sharper pressure gradient along tropical Pacific corresponding to La Niña pattern
Tropics / N. Hemisphere Pattern (TNH)	Mainly positive pressure anomaly centers in western and southern North America and negative in northeastern North America, November–February
West Pacific Pattern (WP)	Enhanced East Asian-Western Pacific Jet, mainly winter

Table 6.1: Summary of external effects considered. Units for temperatures are °C. All other indices are dimensionless anomalies.

interpret. Once a model has been developed which incorporates all of the obvious structure in the DMWS field, we then consider adding the external effects. Those external effects that, on physical grounds, are believed to have the largest effect upon the study region are added first. External effects are time dependent, but their effects upon regional wind speeds may be site and season dependent. Therefore main external effects, along with interactions involving spatial location and season are considered. Throughout the model building process, the analysis of residual plots plays a vital role in dictating the inclusion of additional covariates. For example, when trying to account for spatial variability, mean Pearson residual plots by spatial location are extremely informative in identifying potential improvements to the model.

b) Fitted models

Initial model

This model was developed using the modelling strategy outlined above, as described by Yan et al. (2002). The model contains 38 main effects, which includes six autoregressive terms, a land-sea indicator, three Legendre polynomials for altitude, four Legendre polynomials each for latitude and longitude, annual and half-yearly seasonal cycles, a constant adjustment for August, and various external effects. In addition, significant two-way and three-way interactions are also included. In total, the model consists of 110 predictors and explains 51.5% of the variability in the data. The remaining variance is principally due to daily weather fluctuations, and is regarded as random in the GLM framework. The autoregressive terms are easily the most important influence upon DMWS, based on the percentage of variance explained. The fitted gamma distributions have an estimated shape parameter, ν, of 8.37.

To visualize the structure of the fitted model, covariate effects may be plotted.

Figure 6.14 depicts the primary dependence of DMWS upon altitude, latitude, longitude and season, according to the model. The effects plotted are multiplicative adjustments to an overall mean wind speed. Here we consider main effects only and exclude interactions. The results can be summarized as follows:

- Altitude effect: There is a clear land-sea effect, with the multiplicative factor reaching its maximum over the sea (0m altitude). Over the land, at low altitudes, say less than 100m, the wind remains stronger than average, mainly reflecting proximity to the coast. For higher altitudes, the effect is nearly constant, with a slight increase with altitude, reflecting a topographic effect.

- Latitude effect: Wind speeds reach a maximum between 55°N and 60°N, corresponding to the position of the North Atlantic storm track.

- Longitude effect: The multiplicative factor decreases significantly from west to east.

- Seasonal effect: Not surprisingly, the winter is windier than the summer.

Overall, these results are consistent with the preliminary analysis undertaken.

Fifteen of the external effects considered, were found to have a statistically significant impact upon DMWS. Of these, the effects of AO, NAO, NHT, EA, NAT and SAT varied with site and season; the effects of SHT, EAJ, PNA, SOI, EAWR and EP varied with site only; and the effects of SCA, TNH and NP were constant. The North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) were identified as being the two most important external effects, in terms of their impact upon DMWS. This is not surprising, since they are known to be the large-scale circulation patterns most directly affecting the region under study. A strong AO/NAO year corresponds to strong DMWS almost everywhere in the region, except in the most southern and eastern part of the region in summer.

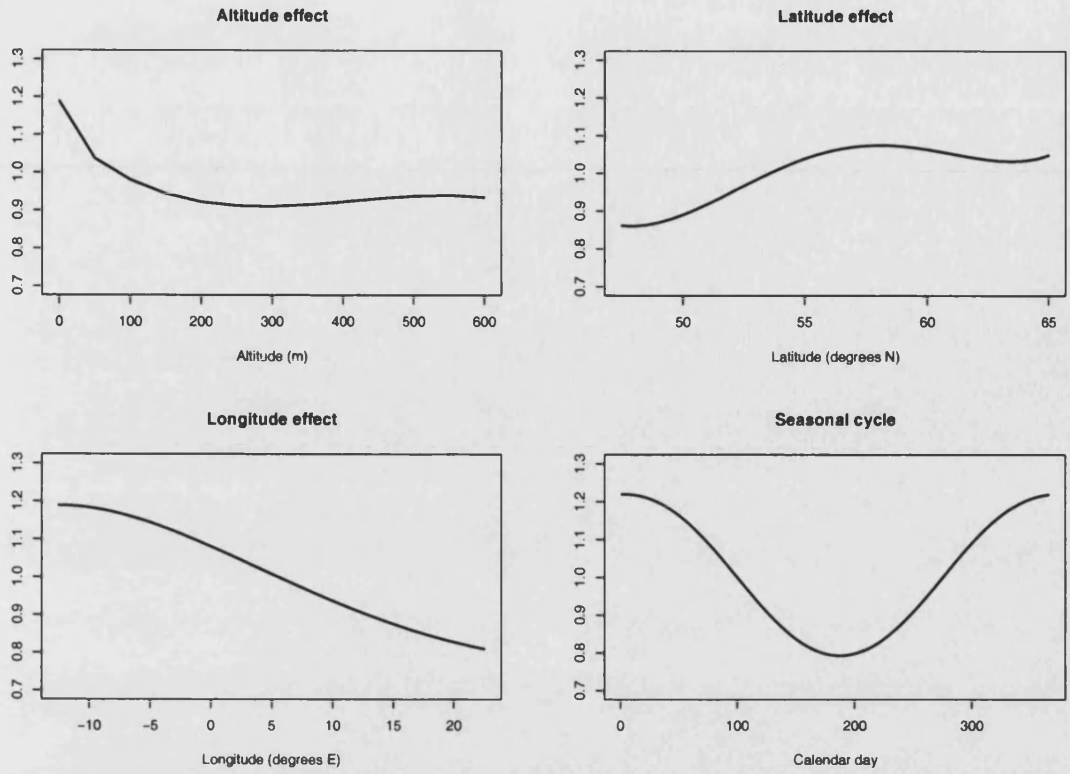

Figure 6.14: Average seasonal and regional variation in DMWS, according to the gamma GLM.

The long-term enhancement of DMWS over the ocean and most of the British Isles and Scandinavia is closely related to enhanced AO and NAO during the last few decades. Global warming also appears to have impacted upon DMWS. In particular, Southern Hemisphere temperature (SHT) exhibits a significant effect on the distinct trends in DMWS shown in Figure 6.13. A possible explanation is that the steady warming in the Southern Hemisphere during the last few decades may have forces the North Atlantic storm track to shift in such a way that storms are enhanced towards the northwestern oceanic area, but weakened throughout most of continental Europe. In summary, by forming interactions between the external effects, geographical effects and seasonal effects we have been able to capture the regional trends in DMWS identified in the preliminary analysis.

Further details regarding the initial gamma GLM can be found in Yan et al.

(2002).

Modified model

The initial model contains six autocorrelation main effects of the form

$$\log(1 + \text{DMWS value } t \text{ day's ago at same site}) \tag{6.3}$$

where $t = 1, 2, \ldots, 6$. However, it seems likely that not only will DMWS values depend upon previous values at the same site but also on previous values at a neighbourhood of sites. Therefore we propose replacing the six predictors defined in (6.3) by six autoregressive terms which are weighted averages of previous values at a neighbourhood of sites.

The chosen neighbourhood structure can be seen in Figure 6.15. As well as considering the same site (SS) itself on previous days, the 8 surrounding neighbours (N1-N8) on previous days are also considered. A weighting system based on distance is adopted, where respectively w_{SS} and w_i $(i = 1, \ldots, 8)$ denote the weights allocated to the same site and neighbouring site Ni, subject to $w_{SS} + \sum_{i=1}^{8} w_i = 1$. Under this scheme, the weight $1 - w_{SS}$ is divided amongst the eight neighbours based on their distance from SS. Thus $w_i \propto d_i^{-1} (i = 1, \ldots, 8)$, where d_i denotes the distance from SS. Under this criterion, and assuming the sites form a regular grid, we have the weighting system given in Table 6.2. In practice, the sites do not form a regular grid as they are located on a globe, however, it is felt that the above approximation was sufficient. Also note that an appropriate weighting correction needs to be applied to sites which are located on the boundary of the region and therefore do not have eight surrounding neighbours.

To implement the weighting scheme outlined above, a value of w_{SS} must be chosen. To achieve this a grid search was undertaken on the value of w_{SS}. A series of models were fitted in which the value of w_{SS} was varied. Table 6.3 shows

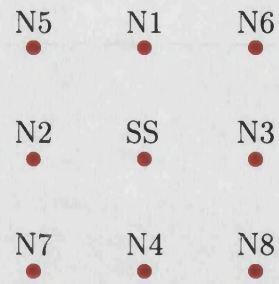

Figure 6.15: Neighbourhood structure for autoregressive covariates. The same site on previous days is labelled SS, and the eight surrounding neighbours are labelled N1-N8

Site	Weight
Same site	w_{SS}
Nearest neighbours (sites N1,N2,N3,N4)	$\frac{1-w_{SS}}{4+\sqrt{8}}$
Furthest neighbours (sites N5,N6,N7,N8)	$\frac{1-w_{SS}}{4(\sqrt{2}+1)}$

Table 6.2: Neighbourhood weighting scheme adopted for autoregressive covariates.

the performance of the various values of w_{SS} in terms of the independence log-likelihood and R^2. It can be seen that the best weighting system corresponds to $w_{SS} = 0.3$, as this maximises both the independence log-likelihood and R^2. Also note that this weighting scheme produces a far superior fit to that obtained from the original autoregressive representation defined in (6.3), where $w_{SS} = 1$.

Having assigned the same site a weighting of 0.3, this results in a weighting for the nearest neighbours of 0.1025 and a weighting for the furthest neighbours of 0.0725. We therefore take the initial model and replace the six autoregressive

w_{SS}	Independence log-likelihood	R^2
0.10	-664,373	52.85
0.15	-662,419	52.94
0.20	-661,162	53.00
0.25	-660,521	53.03
0.30	-660,423	53.04
0.35	-660,800	53.01
0.50	-664,236	52.84
0.75	-675,145	52.27
1.00	-689,721	51.51

Table 6.3: Grid search for the weight w_{SS} in the autoregressive neighbourhood.

predictors defined in (6.3) with predictors

$$\begin{aligned}\log[1 \; &+ \; 0.3(\text{DMWS value t days ago at same site}) \\ &+ \; 0.1025(\sum_{i=1}^{4} \text{DMWS value t days ago at site Ni}) \\ &+ \; 0.0725(\sum_{i=5}^{8} \text{DMWS value t days ago at site Ni})], \end{aligned} \tag{6.4}$$

for $t = 1, \ldots, 6$. This model is labelled the modified model. The results from fitting the modified gamma GLM can be seen in Appendix A.

For the remainder of this chapter, whenever we refer to the 'gamma GLM' we mean the modified gamma GLM of this section.

c) Model checking

The fit of the gamma GLM is now investigated through the use of residual plots. Pearson residuals (see Section 2.6) are initially used, which for the gamma dis-

tribution are defined as

$$r_{ts}^{(P)} = \frac{y_{ts} - \mu_{ts}}{\mu_{ts}}. \tag{6.5}$$

Typically, when producing residual plots, individual residuals are plotted, for example against the fitted values, to check for systematic structure in the residuals. With 1.8 million observations this approach is impractical, and therefore as an alternative we group together observations and analyse mean Pearson residuals.

In Figure 6.16 mean monthly Pearson residuals have been plotted, to check for unexplained seasonal structure. Also shown are approximate 95% confidence intervals, which allow for spatial dependence and have been calculated as detailed in Section 4.1 of Wheater et al. (2000). It can be seen that one of the residuals is marginally significant. Overall, however, this plot looks fine as no systematic structure is evident. In Figure 6.17, mean annual Pearson residuals have been plotted, to check for unexplained trends. Again, no systematic structure is evident. Other Pearson residual plots were produced which are not presented here, such as mean residual by spatial location, and all results suggested an adequate fit.

Finally, we now check our distributional assumptions and investigate whether the DMWS values are well represented by gamma densities. To achieve this we introduce Anscombe residuals, which for the gamma distribution are defined by

$$r_{ts}^{(A)} = \left(\frac{y_{ts}}{\mu_{ts}}\right)^{1/3}. \tag{6.6}$$

Theoretically, the distribution of these residuals should be approximately standard normal, if the studied data are gamma distributed (McCullagh and Nelder, 1989, Section 2.4.2). Therefore, a check on our gamma distribution assumption can be performed by producing a normal quantile-quantile plot of the Anscombe residuals from the model. This plot is shown in Figure 6.18, and since most points fall on the straight line, this indicates a good fit, despite there being a slight departure from the line in the upper tail.

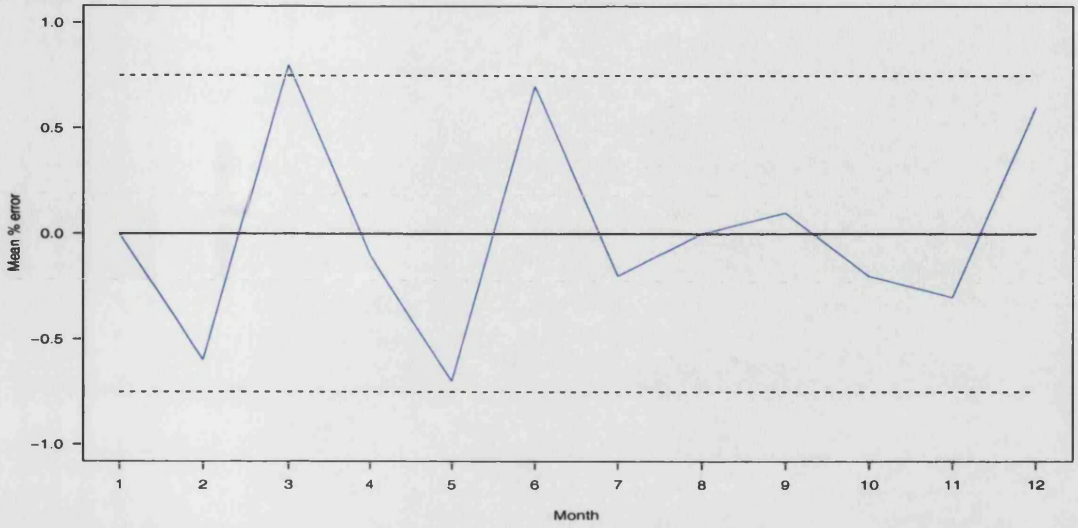

Figure 6.16: Plot of monthly Pearson residuals for gamma GLM. Dotted lines are approximate 95% confidence intervals.

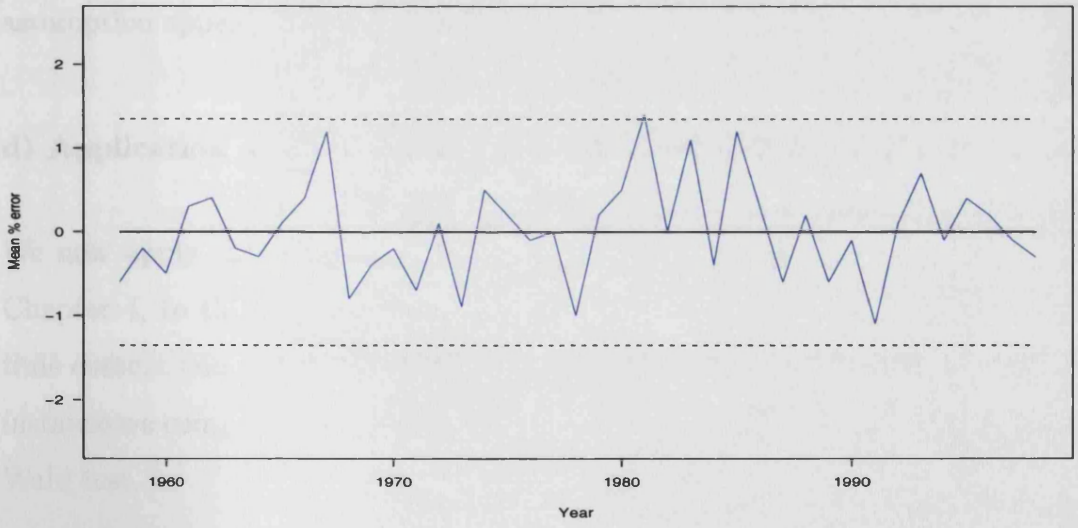

Figure 6.17: Plot of annual Pearson residuals for gamma GLM. Dotted lines are approximate 95% confidence intervals.

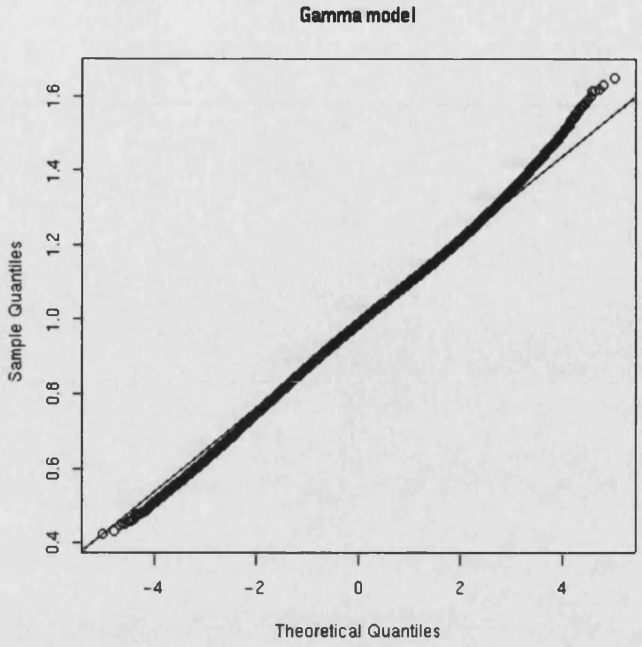

Figure 6.18: Normal quantile plot of residuals from gamma GLM.

Overall, the residual analysis presented in this section suggests that the gamma GLM provides a good fit to the DMWS data. No systematic structure in the Pearson residuals has been identified and the gamma distributional assumption appears reasonable.

d) Application of the new hypothesis testing technique

We now apply the new adjusted independence likelihood ratio test, derived in Chapter 4, to this data set. This enables us to apply the new test in a space-time context and to consider its performance relative to other techniques. In this instance we compare it with the independence likelihood ratio test and the robust Wald test.

One of the external factors included in the gamma GLM is the Northern Hemisphere Temperature (NHT). In fact, including interactions this effect is present in seven of the predictors in the model. For illustration, we now consider testing

Independence log-likelihood ratio test statistic	New adjusted log-likelihood ratio test statistic	Robust Wald test statistic
236.6	30.0	27.6

Table 6.4: Calculated test statistics for testing the NHT effects.

the statistical significance of these seven predictors using the methods outlined above. Thus, we test the null hypothesis $H_0 : \beta_N = 0$, where β_N corresponds to the seven effects involving NHT. The three calculated test statistics are given in Table 6.4. Comparing these test statistics to a χ^2_7 distribution, we conclude that all tests provide strong evidence against H_0. The independence log-likelihood test, however, is calculated on the incorrect assumption of independent data. The new method corrects this test statistic for the dependence, and we can see that this correction is considerable. Finally, in this instance, the new method provides very similar results to that of the robust Wald test, however, in general, we prefer to base inference on the new method since it is better able to allow for correlated predictors (see Section 4.4).

6.3.2 Weibull model

Within a generalized linear modelling framework, the gamma distribution is the natural choice for modelling continuous, non-negative and positively skewed data, such as wind speeds. The most prominent distribution within the existing literature on wind speeds, however, is the Weibull. The motivation for this appears to be the observation that if the u and v components of wind velocity can be modelled by a Gaussian process, then the wind speed $(= \sqrt{u^2 + v^2})$ will follow a Rayleigh distribution, which is a special case of a Weibull distribution. Of course, if the above Gaussian assumption for the wind velocity components is misplaced, so to is the assumption that wind speeds follow a Weibull distribution. Never-

theless, the Weibull is sufficiently established within the wind speed literature to warrant our attention. Therefore within this section, as an alternative to the gamma GLM of Section 6.3.1, we assume that all DMWS values are drawn from Weibull distributions.

The Weibull pdf is given by

$$f(y;\lambda,\alpha) = \frac{\alpha}{\lambda} y^{\alpha-1} \exp\left(-\frac{y^{\alpha}}{\lambda}\right), \quad 0 \leq y < \infty, \quad \lambda, \alpha > 0, \tag{6.7}$$

where

$$\mathrm{E}(Y) = \lambda^{1/\alpha}\Gamma\left(1+\frac{1}{\alpha}\right) \text{ and } \mathrm{Var}(Y) = \lambda^{2/\alpha}\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right].$$

The theory outlined in Section 2.7.1 was adopted to fit a Weibull GLM to the DMWS data. As detailed earlier, this approach involves extending the GLM algorithm, by fitting a series of exponential models for fixed shape parameter α. One of the major drawbacks of this method, in comparison to the gamma GLM, is that it is considerably more expensive to implement computationally, due to the extra level of iteration associated with the estimation of α. Due to this, it would have been extremely time consuming to build a Weibull model up from scratch, by carrying out a model selection process similar to that undertaken for the gamma model. The approach we adopted therefore was to fit a single Weibull model, containing the same 110 predictors as the modified gamma GLM of Section 6.3.1. This approach has the obvious benefit that since both the final Weibull and gamma models contain the same predictors, a direct comparison of the two models is relatively straightforward.

While the gamma GLM took slightly less than one hour to fit on a modern 3GHz pc, the equivalent Weibull GLM took more than three times as long. The Weibull model explains 52.4% of the variance in the data, which compares with the value of 53.0% for the equivalent gamma model. This suggests that the gamma model provides a slightly better fit to the data, however, it could be

argued that since the covariate selection process was based on the gamma model, then this provides the gamma model with an unfair advantage. A comparison of individual predictors in the two models was also undertaken and it was found that the estimated effect of each predictor was very similar for both models. Therefore the conclusions drawn from the gamma model of Section 6.3.1, regarding the relationships between covariates and wind speeds, also apply to the Weibull model. Hence, the conclusions drawn, in general, appear to be insensitive to the choice of distribution. Pearson residual analysis was also undertaken for the Weibull model to check for any systematic structure remaining in the residuals. The results obtained were very similar to those presented in Section 6.3.1 for the gamma model and are therefore not presented here.

We now attempt to compare the shape of the fitted gamma and Weibull distributions. The problem we have, however, is that a different distribution is fitted to each observation under each distribution, making a direct comparison difficult. By fixing the mean, however, we are able to gain an insight into the differences in the shapes of the two distributions. Figure 6.19 illustrates the shapes of the fitted gamma and Weibull distributions with mean 1, where naturally each of the shape parameters have been fixed at their constant fitted value. The plot on the left hand side compares the two density functions and it is evident that the shapes are very similar, although the Weibull has a slightly larger spread in its central range, and the gamma has a heavier upper tail. This difference in the upper tail is also evident in the quantile-quantile plot located in the right hand window of Figure 6.19. Here, corresponding percentage points (0% point to 99.99% point) of each distribution have been plotted against each other, and the straight line corresponds to perfect agreement. The heavier upper tail of the gamma distribution is clearly evident; the largest percentile plotted for the Weibull distribution is approximately 2.2, whereas for the gamma distribution it is nearly 3. Naturally, if these two models were used to simulate extremes, the results obtained would be quite different, even though their mean structures are similar.

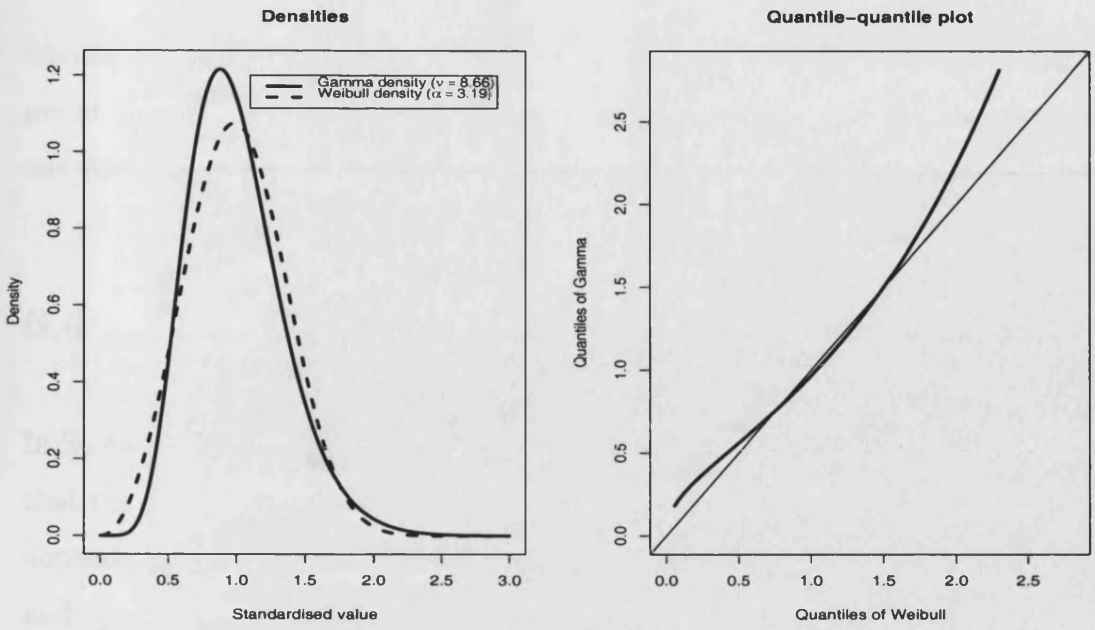

Figure 6.19: Comparison of the shape of the fitted gamma and Weibull distributions.

For the Weibull fit, the final estimate of the shape parameter α was 3.19, which is considerably larger than the value of 2 consistent with a Rayleigh distribution. This therefore casts doubt upon the validity of the historic perspective that wind speeds follow a Rayleigh distribution. A normal quantile-quantile plot of Anscombe residuals was also produced for the Weibull model, equivalent in form to the gamma model plot shown in Figure 6.18. Whereas the gamma model quantile-quantile plot generally looked fine, the Weibull model plot showed more deviation about the straight line. This suggests that a gamma distributional assumption is more appropriate for the DMWS values than a Weibull assumption. Note that since the gamma quantile-quantile plot of Anscombe residuals looked fine and Figure 6.19 shows that the quantiles of the fitted Weibull and gamma distributions are quite different, it's not surprising that the Anscombe residual quantile-quantile plot for the Weibull model suggests the Weibull is inferior to the gamma.

Overall, gamma distributions appear to provide a better fit to the DMWS values than Weibull distributions. In addition, the gamma model has a sub-

stantial computational advantage over the Weibull model. For these reasons, we prefer the gamma model, and therefore we do not consider the Weibull model any further.

6.4 Generalized estimating equations approach

In Section 6.3 the DMWS data were modelled using a GLM, effectively assuming that the responses are independent given the covariates in the model. Temporal dependence was accounted for by including autoregressive effects in the model, and spatial dependence through geographical effects such as longitude, latitude and altitude. Including a neighbourhood structure in the autoregressive terms also helps to account for spatial dependence. Inevitably though, additional spatial dependence is still present, as a result of spatial locations close together being subjected to the same weather systems at the same time. To account for this additional spatial dependence in the response we apply the GEE methodology outlined in Chapter 5. Each day represents a cluster and within each cluster there are 120 potentially correlated responses, corresponding to the 120 spatial locations.

Throughout this section we aim to build a GEE using an identical set of covariates to that of the modified gamma GLM of Section 6.3. This approach is adopted to enable a comparison of the respective fits to be undertaken.

Before a GEE is fitted to the data set we must first decide upon a relationship between the marginal mean and variance. From the analysis already undertaken within this chapter, it seems plausible to assume that the marginal distribution of DMWS values, found at each site, follows a gamma distribution. The following standard link between the marginal mean and variance then follows

$$\text{var}(Y_{ts}) = \frac{\mu_{ts}^2}{\nu},$$

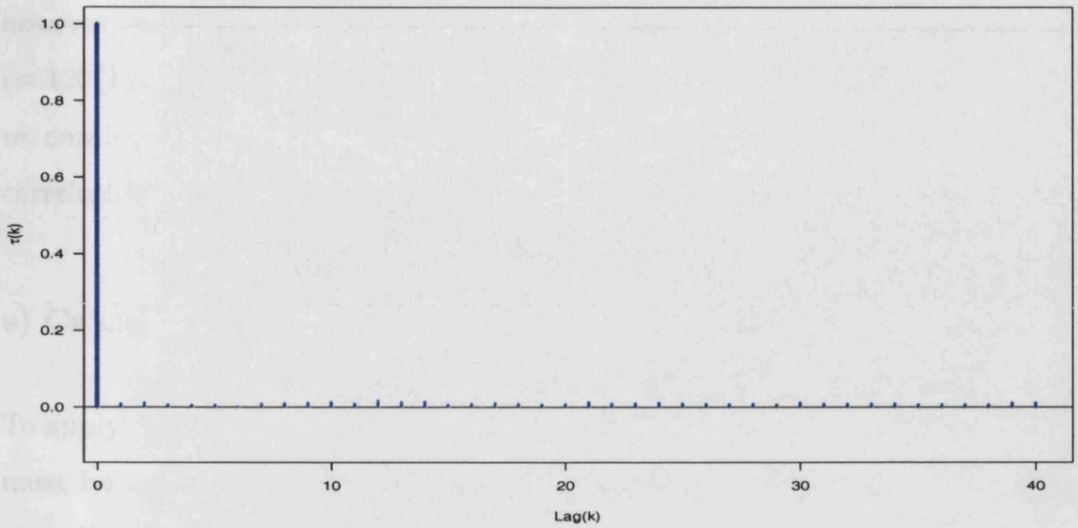

Figure 6.20: Plot of sample autocorrelation function of Pearson residuals from gamma GLM.

where ν is the common shape parameter.

6.4.1 Allowing for temporal dependence

To enable the GEE methodology to be applied we must be able to assume independent time points, conditional on the covariates. As detailed in Section 5.3, to investigate this we plot the autocorrelation function of the Pearson residuals from the gamma GLM fit, detailed in Section 6.3. This ACF is shown in Figure 6.20, and it does appear that the temporal dependence has been accounted for via the covariates. Therefore we are able to continue with the GEE approach.

6.4.2 Allowing for spatial dependence

A structure must be chosen for the working correlation matrix $\mathbf{R}(\boldsymbol{\alpha})$, and to achieve this we investigate the pairwise correlations in the Pearson residuals

obtained from the gamma GLM fit. One possible structure would be the unstructured form of $\mathbf{R}(\boldsymbol{\alpha})$ (see Section 3.1.4). The problem with this structure, however, is that since the clusters are of size 120, the parameter vector $\boldsymbol{\alpha}$ has 7140 $(= 120(120-1)/2)$ elements to be estimated, which is clearly excessive. Therefore we consider modelling the working correlation structure using an isotropic spatial correlation function as detailed in Section 5.4.1.

a) Calculating the distance between two sites

To apply this working correlation structure, the distance between all pairs of sites must be calculated. A very crude approximation to these distances is given by their Euclidean distance measured in degrees. The problem with this method, however, is that since the earth is not flat, pairs of sites which are separated by $x°$ to the north of the region are, in general, closer than two sites which are separated by $x°$ to the south of the region. We therefore decided to calculate the distance between all pairs of sites in nautical miles. From Roy and Clarke (1988), the distance (u), in nautical miles, between two sites at locations (Lat_1,$Long_1$) and (Lat_2,$Long_2$), where latitude and longitude are measured in degrees, is given by

$$\begin{aligned} u &= \frac{60 \times 180}{\pi} \times \cos^{-1}\left[\sin\left(Lat_1 \times \frac{\pi}{180}\right)\sin\left(Lat_2 \times \frac{\pi}{180}\right)\right. \\ &+ \left.\cos\left(Lat_1 \times \frac{\pi}{180}\right)\cos\left(Lat_2 \times \frac{\pi}{180}\right)\cos\left((Long_2 - Long_1) \times \frac{\pi}{180}\right)\right] \quad (6.8) \end{aligned}$$

The distance between all pairs of sites was calculated using (6.8) and then pairs of sites the same distance apart were grouped together. Correlations in Pearson residuals are then calculated using (5.4).

b) Selecting a correlation function

We now turn to the question of selecting the most appropriate spatial correlation function for this particular data set. To help answer this question we fit various correlation functions to the Pearson residuals from the gamma GLM fit. The correlation functions we consider are those introduced in Section 5.4.1, these being the 2-parameter powered exponential, the 1-parameter spherical and the 2-parameter Matérn correlation functions. These were fitted to the correlations using a non-linear regression routine. Figure 6.21 shows the fitted powered exponential function, which fits the observed correlations very well. Figure 6.22 shows the fitted spherical correlation function. Clearly, this function does not provide a good fit to the observed correlations; this is probably due to the inflexibility of this 1-parameter family (see Section 5.4.1). Finally, Figure 6.23 shows the fitted Matérn correlation functions. Within this plot there are three separate fits. This is because only the φ parameter is estimated, while the ξ parameter is fixed at one of three values $\xi = 1, 1.5, 2$, resulting in the three separate fits. From studying this plot we see that setting $\xi = 1$ or $\xi = 1.5$ appears to provide the best fit.

To decide between the powered exponential and the Matérn functions, the sum of squared errors for each fit has been calculated and these values, along with the parameter estimates, can be seen in Table 6.5. The best fitting Matérn function is obtained when $\xi = 1$. The powered exponential function, however, is marginally superior to this Matérn function, according to the sum of squared errors criterion. We therefore select the 2-parameter powered exponential model.

c) Testing for anisotropy

We now investigate for evidence of anisotropy. To achieve this we proceed in the manner detailed in Section 5.4.2. Thus, each pair of sites is allocated to one of four orientation groups. These groups are centered on 0°, 45°, 90° and

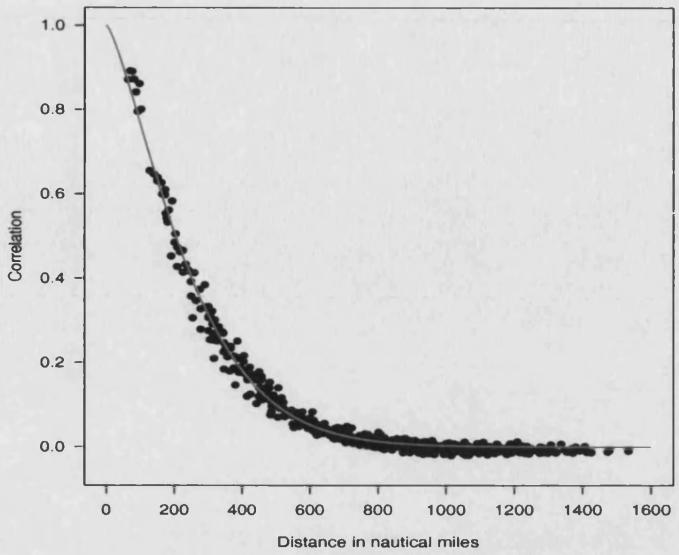

Figure 6.21: Powered exponential correlation function fit to the pairwise correlations in Pearson residuals from the gamma GLM.

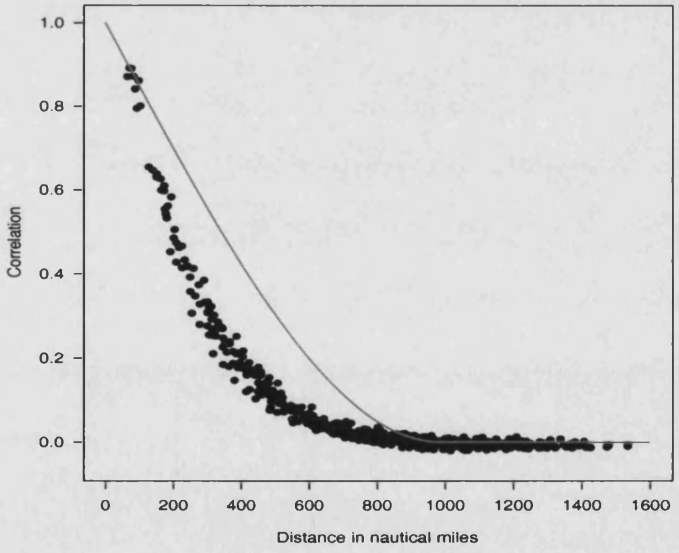

Figure 6.22: Spherical correlation function fit to the pairwise correlations in Pearson residuals from the gamma GLM.

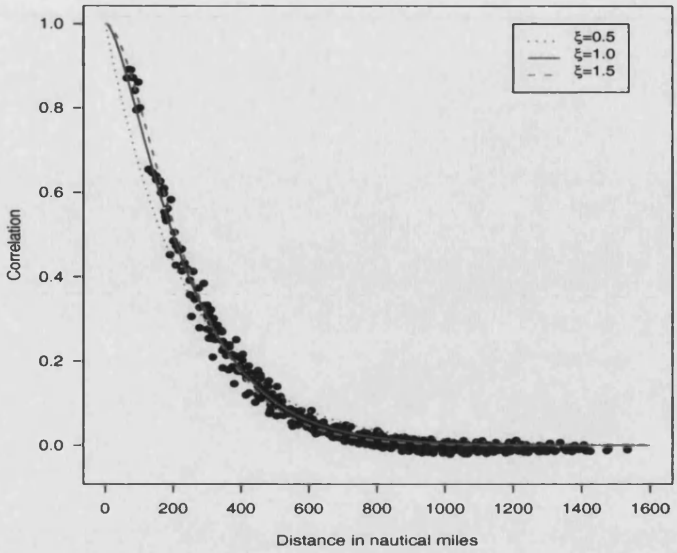

Figure 6.23: Matérn correlation function fits to the pairwise correlations in Pearson residuals from the gamma GLM.

135°, with each group covering 45° (where 0° corresponds to the north and angles are standard compass bearings). Within groups, inter-site Pearson residual correlations are calculated over the range of the inter-site distances. For each of the four groups a separate 2-parameter isotropic powered exponential correlation function is fitted to the pairwise correlations. These individual fits can be seen in Figure 6.24. The powered exponential correlation function fits the correlations well in all directions. In Figure 6.25, the four separate fitted correlation functions have been overlaid onto the same plot. This shows that the correlation decay rate varies with direction, with the angle of 45° possessing the greatest amount of correlation. This plot suggests some evidence of anisotropy, with the strongest correlation present in the south-west north-east direction. We therefore adopt the four parameter powered anisotropic exponential correlation function given by (5.10).

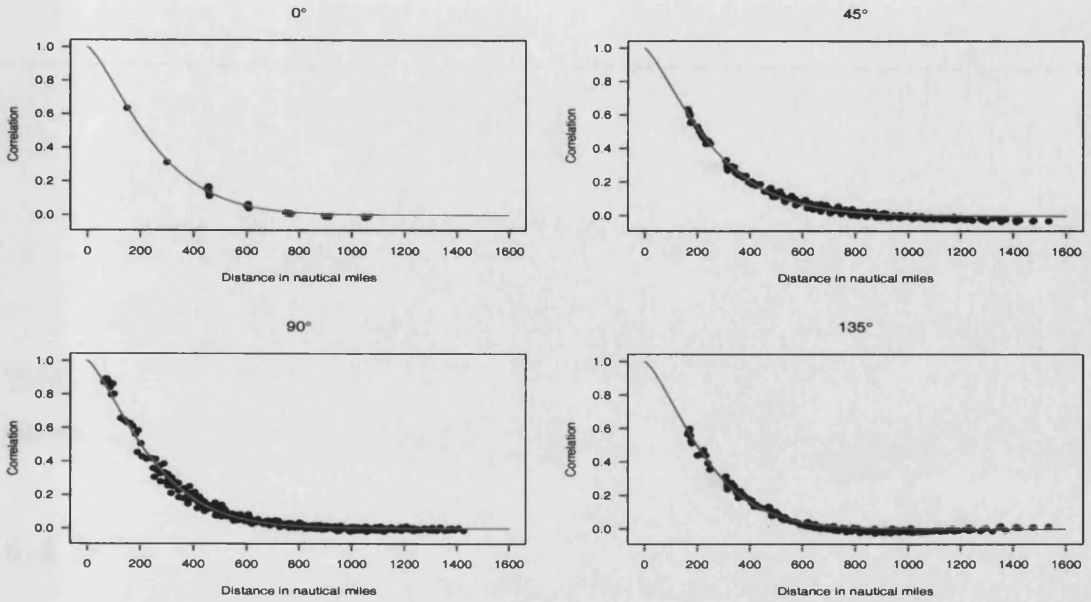

Figure 6.24: Powered exponential correlation function fits to the pairwise correlations in Pearson residuals from the gamma GLM. The four separate plots correspond to the directions 0°, 45°, 90° and 135°.

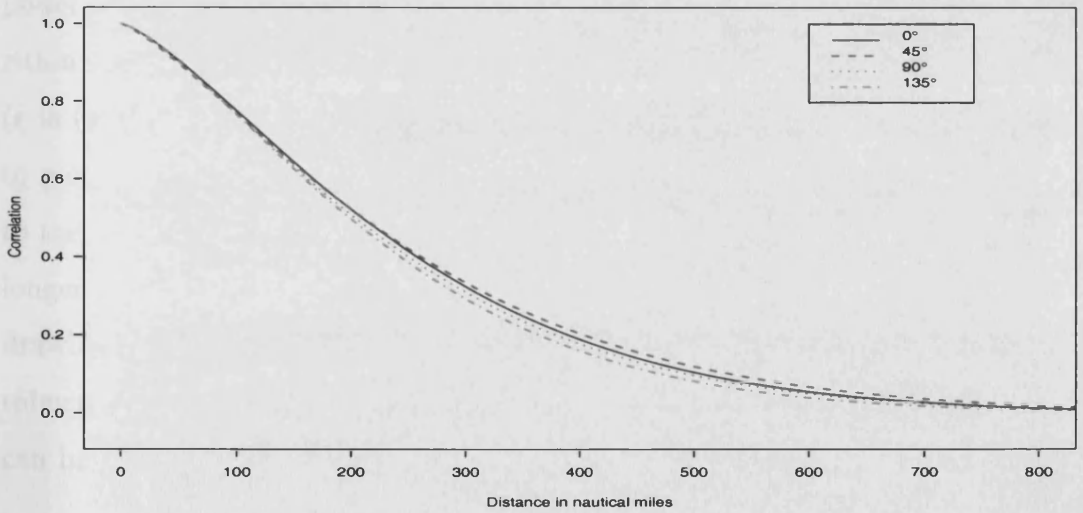

Figure 6.25: Powered exponential correlation function fits to the pairwise correlations in Pearson residuals from the gamma GLM, all four directions.

Correlation function	φ	ξ	Sum of squared errors
Exponential	268.03	1.34	0.159
Spherical	984.09	-	11.20
Matérn	240.83	0.50	0.649
Matérn	158.09	1.00	0.166
Matérn	124.96	1.50	0.180
Matérn	106.23	2.00	0.252

Table 6.5: Performance of the correlation functions in terms of sum of squared errors. Parameter estimates are also given.

6.4.3 Results

Having identified a suitable spatial working correlation structure, the modified GLM of Section 6.3 was refitted using a GEE approach. A four parameter anisotropic exponential correlation function was fitted to the Pearson residuals, using non-linear regression, each time the $\mathbf{R}$ matrix was re-estimated. In total the $\mathbf{R}$ matrix was estimated 3 times before the parameters of the anisotropic powered exponential correlation function converged, resulting in the whole algorithm converging. The final estimate obtained for the anisotropy ratio parameter (r in (5.10)) was statistically smaller than 1 and therefore supported our decision to model the spatial correlation using an anisotropic structure. When compared to the gamma GLM of Section 6.3.1, the spatial GEE took more than five times longer to converge to its final parameter estimates, highlighting the computational drawback of this method. The final estimated covariate effects and corresponding robust standard errors (see Section 3.1.2) obtained from fitting the spatial GEE can be seen in Appendix B.

In general, the overall fit obtained from the spatial GEE method was similar to the gamma GLM of Section 6.3.1. More will be said about this in Section 6.5.

Due to the similarity of the fits, we do not present any model checking results here since the overall message is similar to that of the gamma GLM.

6.4.4 The one-step estimator

As discussed earlier, one of the drawbacks of the GEE method is that it is computationally more expensive to fit when compared with the GLM, due to the extra level of iteration involved in estimating the working correlation matrix $\mathbf{R}$. Naturally, increasing the number of predictors or the number of observations only adds to the problem.

To investigate whether the one-step estimator is a viable option for this particular data set, we consider the convergence properties of the full spatial GEE algorithm implemented above. The GEE fit detailed above began with the $\mathbf{R}$ matrix set equal to the identity matrix. After this independence fit the $\mathbf{R}$ matrix needed to be reestimated 3 more times before the algorithm converged. At each of these 4 iterations of the $\mathbf{R}$ matrix we monitor the convergence of 4 separate quantities, these being, 1) the 110-element parameter vector $\boldsymbol{\beta}$, 2) the variance-covariance matrix of $\boldsymbol{\beta}$ denoted by $\mathbf{Q} = \mathbf{F^{-1}VF^{-1}}$, 3) the anisotropic powered exponential correlation function 4-parameter vector $\boldsymbol{\alpha}$ and 4) the dispersion parameter ϕ. At the end of each $\mathbf{R}$ matrix iteration we compare the estimates of these four quantities with their estimate at the previous $\mathbf{R}$ iteration. To achieve this we introduce the following 4 statistics, one for each of the 4 quantities.

1.
$$\beta_{\Delta}^{(i)} = \sqrt{\frac{\sum_{j=1}^{110}\left(\frac{\beta_j^{(i+1)}-\beta_j^{(i)}}{\beta_j^{(i)}}\right)^2}{110}},$$

where $\beta_j^{(i)}$ denotes the estimate of β_j at the end of the ith $\mathbf{R}$ iteration.

2.

$$Q_{\Delta}^{(i)} = \sqrt{\frac{\sum_{j=1}^{110} \left(\frac{Q_{jj}^{(i+1)} - Q_{jj}^{(i)}}{Q_{jj}^{(i)}} \right)^2}{110}},$$

where $Q_{jj}^{(i)}$ denotes the estimate of the jth diagonal element of the matrix $\mathbf{Q}$ at the end of the ith $\mathbf{R}$ iteration.

3.

$$\alpha_{\Delta}^{(i)} = \sqrt{\frac{\sum_{j=1}^{4} \left(\frac{\alpha_j^{(i+1)} - \alpha_j^{(i)}}{\alpha_j^{(i)}} \right)^2}{4}},$$

where $\alpha_j^{(i)}$ denotes the estimate of the jth element of the 4-parameter vector which determines the working correlation structure at the end of the ith $\mathbf{R}$ iteration.

4.

$$\phi_{\Delta}^{(i)} = \frac{\text{abs}\left(\phi^{(i+1)} - \phi^{(i)}\right)}{\phi^{(i)}},$$

where $\phi^{(i)}$ denotes the estimate of ϕ at the end of the ith $\mathbf{R}$ iteration.

The results from calculating these statistics can be seen in Table 6.6. Note that column $i = 1$ corresponds to the comparison of the gamma GLM estimates with the one-step estimates. Some of these values are difficult to interpret since one or two of the elements explode due to estimates near zero. The most important feature we can take from these values therefore is the sharp decrease in magnitude of the values across iterations. This suggests that the bulk of the convergence effort takes place early on in the iteration process.

Another way to monitor convergence is to calculate the Mahalanobis distance between the estimate of $\boldsymbol{\beta}$ at each $\mathbf{R}$ iteration and the final GEE $\boldsymbol{\beta}$ estimate, under the final $\boldsymbol{\beta}$ variance-covariance structure. The Mahalanobis distance at the ith $\mathbf{R}$ iteration is given by

$$M^{(i)} = (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(F)})^T (\boldsymbol{\Sigma}^{(F)})^{-1} (\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(F)}),$$

	i=1	i=2	i=3
$\beta_{\Delta}^{(i)}$	368.7	65.9	3.8
$Q_{\Delta}^{(i)}$	54.3	2.0	0.2
$\alpha_{\Delta}^{(i)}$	18.0	1.7	0.0
$\phi_{\Delta}^{(i)}$	1.6	0.2	0.0

Table 6.6: Convergence properties of the full spatial GEE algorithm. Values have been multiplied by 100 to represent average percentage change per vector element.

i	$M^{(i)}$
1	40,543.01
2	105.97
3	0.64

Table 6.7: Mahalanobis distance for the spatial GEE.

where $\boldsymbol{\beta}^{(i)}$ is the estimate of $\boldsymbol{\beta}$ at the ith **R** iteration, $\boldsymbol{\beta}^{(F)}$ is the final estimate of $\boldsymbol{\beta}$ under the spatial GEE and $\boldsymbol{\Sigma}^{(F)}$ is the final estimate of the variance-covariance matrix.

The results from calculating these Mahalanobis distances can be seen in Table 6.7. Again, note the sharp decrease in the values early on in the iteration process, suggesting that the GEE one-step estimator should provide us with similar results to that of the full GEE algorithm, for considerably less computational cost.

Finally, we graphically compare the coefficient estimates and corresponding robust t-values obtained from the full GEE algorithm and the one step algorithm.

In Figure 6.26 the two sets of coefficient estimates are plotted against each other. Since all points lie close to the line of perfect agreement, this suggests that the one-step estimator produces very similar results to the full algorithm. In Figure 6.27 the corresponding robust t-values are plotted against each other and again the results are very similar, suggesting that inference carried out on the one-step estimator would produce very similar results to that of the full GEE algorithm. Overall, it is felt that the GEE one-step estimator offers a realistic alternative to the full GEE algorithm in this instance.

6.5 Comparison of approaches

Within this section we compare the gamma GLM and spatial GEE fits discussed in Sections 6.3 and 6.4. Table 6.8 lists various models along with their R^2 values. There are two different models in terms of predictors: the initial model which contains the 110 predictors detailed in Yan et al. (2002) and the modified model which replaces the 6 autoregressive terms in the initial model with the neighbourhood autoregressive terms detailed in Section 6.3. For each of these two different sets of predictors the gamma GLM and spatial GEE are compared. Focusing on the initial model, the best performing model, in terms of R^2, is the gamma GLM, with the spatial GEE performing substantially worse. Once we introduce the autoregressive neighbourhood structure we see that both the gamma GLM and spatial GEE gain in terms of R^2, however, the spatial GEE gains most. The spatial GEE has benefited significantly from the additional modelling of the spatial dependence within the predictors. This may be related to the findings of Pepe and Anderson (1994), who found that biased estimates can be obtained when responses within clusters depend on the covariates of other responses within the same cluster.

An alternative method of comparing the gamma GLM and spatial GEE over-

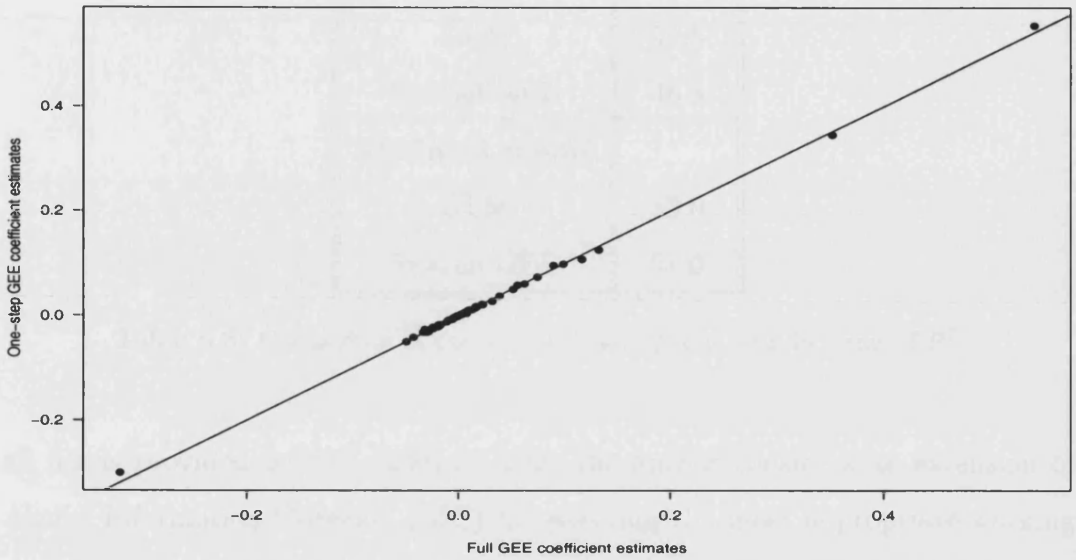

Figure 6.26: Comparison of coefficient estimates obtained from the full GEE algorithm and the one step GEE.

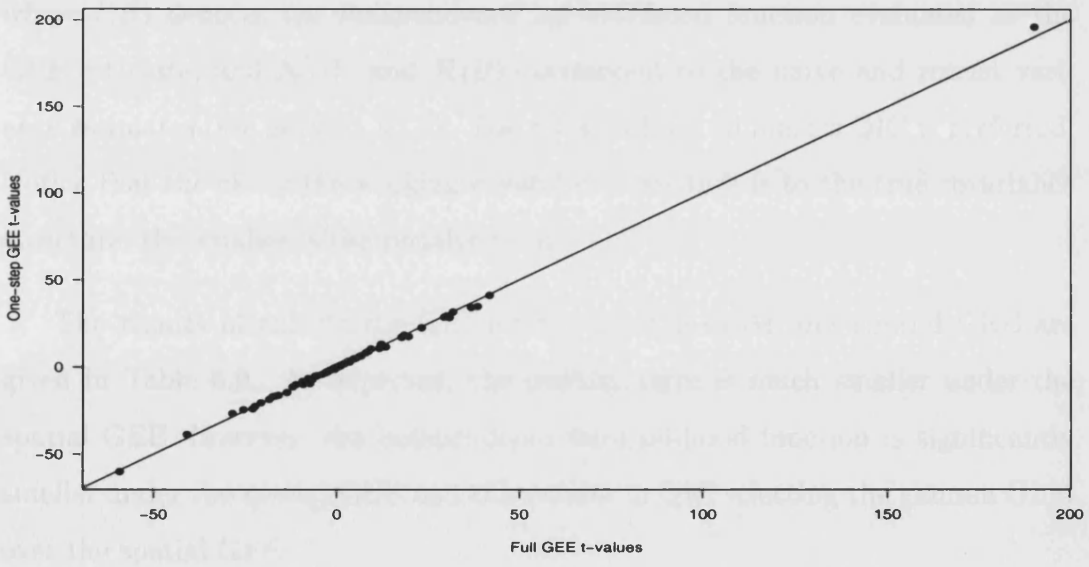

Figure 6.27: Comparison of t-values obtained from the full GEE algorithm and the one step GEE.

Model	R^2(%)
Initial model	
GLM	51.5
Spatial GEE	46.3
Modified model	
GLM	53.0
Spatial GEE	51.0

Table 6.8: Comparison of gamma GLM and spatial GEE in terms of R^2.

all fits is provided by Pan (2001a). Here the author considers an extension of Akaike Information Criterion (AIC) for selecting the most appropriate working correlation structure within a GEE context. Now since the gamma GLM is equivalent to independence estimating equations in this context, we are able to apply this theory. Pan's selection criterion is called QIC and takes the form

$$QIC = -2\ell(\hat{\beta}) + 2\text{trace}\{\mathcal{N}(\hat{\beta})^{-1}\mathcal{R}(\hat{\beta})\}, \tag{6.9}$$

where $\ell(\hat{\beta})$ denotes the independence log-likelihood function evaluated at the GEE estimate, and $\mathcal{N}(\hat{\beta})$ and $\mathcal{R}(\hat{\beta})$ correspond to the naive and robust variance estimates (see Section 3.1.2). The model which minimizes QIC is preferred. Notice that the closer the working covariance structure is to the true covariance structure, the smaller is the penalty term.

The results of calculating QIC for the gamma GLM and spatial GEE are given in Table 6.9. As expected, the penalty term is much smaller under the spatial GEE. However, the independence log-likelihood function is significantly smaller under the spatial GEE and this results in QIC selecting the gamma GLM over the spatial GEE.

In Figure 6.28 the gamma GLM coefficient estimates have been plotted against the spatial GEE coefficient estimates. There does appear to be a fair amount of

Model	Independence log-likelihood	Penalty term	QIC
Gamma GLM	-660,423	1856	1,322,702
Spatial GEE	-690,359	211	1,380,929

Table 6.9: Comparison of gamma GLM and spatial GEE in terms of QIC.

deviation from the line of perfect agreement, suggesting that the estimates obtained by the two methods are reasonably different. Figure 6.29 compares the corresponding robust t-values. Once again there does appear to be a fair amount of scatter, which suggests that inference based on the different methods could lead to different conclusions.

The boxplots in Figure 6.30 provide us with an insight into the differences between the fitted values under the different fitting methods. For each observation in the data set, the difference between the gamma GLM and spatial GEE fitted values has been calculated. These differences have then been grouped according to their gamma GLM fitted values. For each of these groups a boxplot of the differences has been produced. The median difference is negative for the two left-most boxplots, and is positive for the other 4 boxplots. Therefore the lowest gamma GLM fitted values are, on average, smaller than the lowest spatial GEE fitted values, and the larger gamma GLM fitted values exceed those for the spatial GEE. As the GLM appears to generate more fitted values in both tails of the distribution this will have a substantial impact on its ability to simulate extremes.

6.6 Summary

Within this chapter a generalized linear modelling approach has been adopted to model daily maximum wind speeds over northwestern Europe. The GLM framework has enabled us to explain various patterns in the data and identify

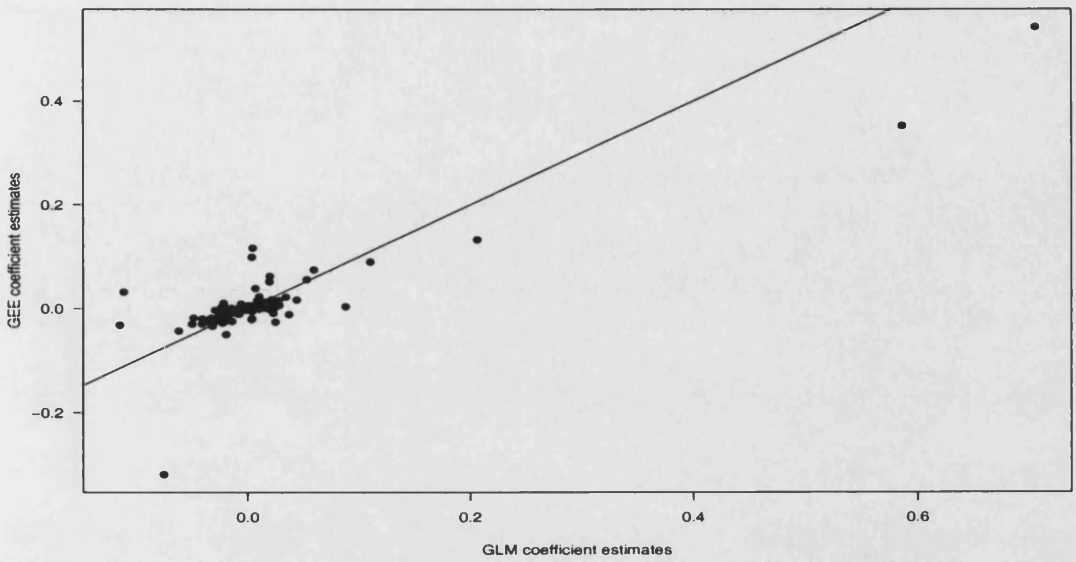

Figure 6.28: Comparison of coefficient estimates obtained from the gamma GLM and the spatial GEE.

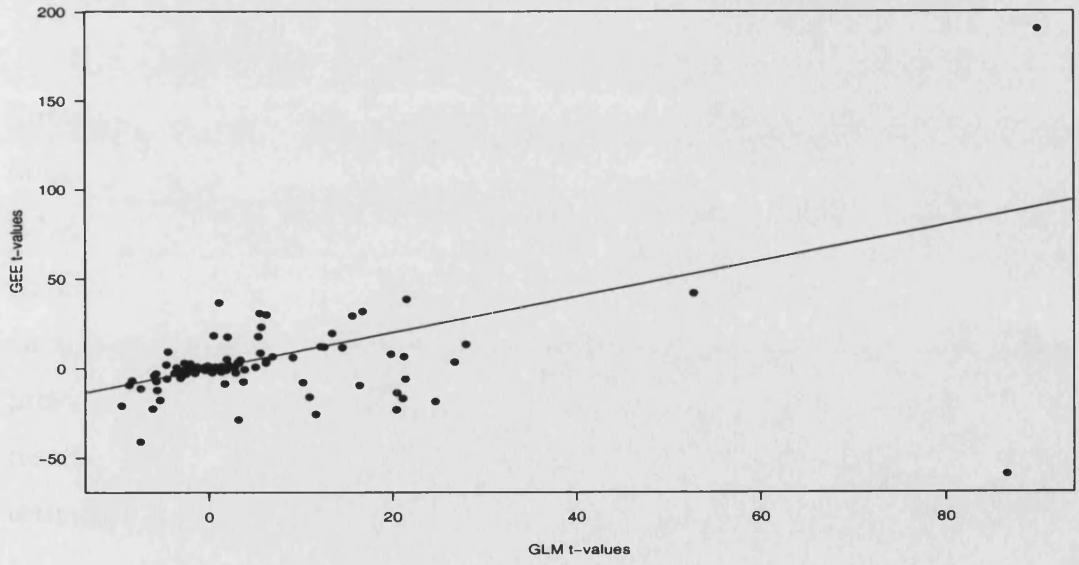

Figure 6.29: Comparison of t-values obtained from gamma GLM and spatial GEE.

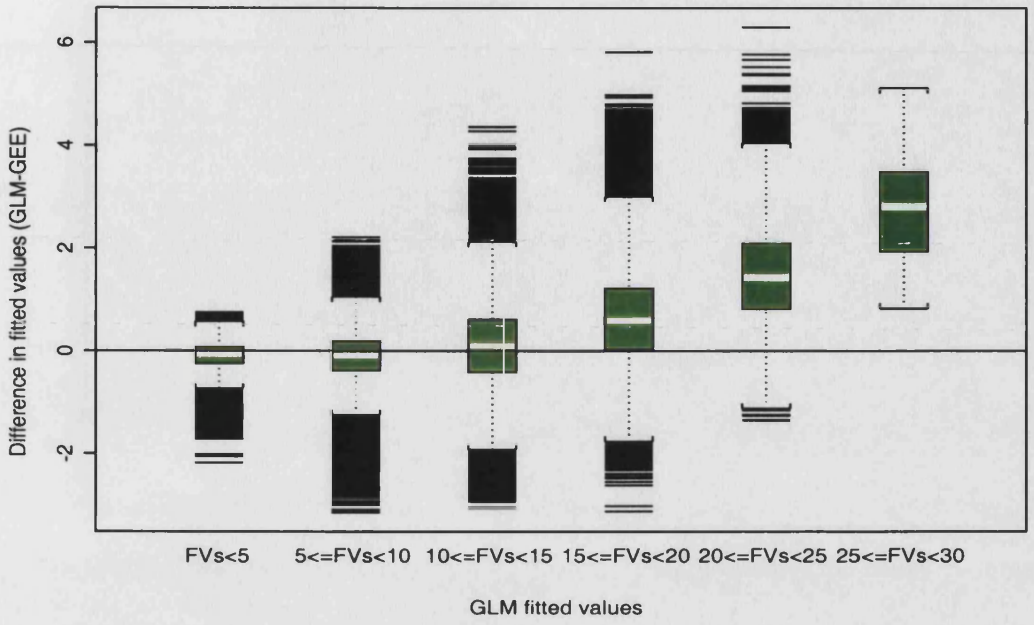

Figure 6.30: Plot of difference in fitted values (gamma GLM - spatial GEE) against GLM fitted values.

factors which effect wind speed within the region under study.

This case study has also enabled us to apply the new techniques developed throughout the thesis, to a space-time data set. A comparison of the results obtained from employing different estimation techniques has also been undertaken. When comparing the gamma GLM and spatial GEE it was discovered that difference did exist in parameter estimates and robust t-values, which could impact on any conclusions drawn. Also, it was found that the GEE one-step estimator provided very similar results to the full algorithm in this case. Based on these results there would be a strong argument in favour of employing the one-step estimator over the full algorithm for large data sets.

Chapter 7

Conclusions and further work

This thesis has focused on the application of generalized linear models to dependent response data. Having covered mostly standard GLM theory in Chapter 2, we then proceeded to consider extensions of the standard theory for cluster correlated data in Chapter 3. Chapter 4 then built on the ideas discussed in the previous chapters to propose a new hypothesis testing technique, appropriate for the application of GLMs to cluster correlated data. This method essentially adjusts the independence log-likelihood ratio test statistic to allow for the within cluster dependence. Using simulations the performance of the new method was compared with established techniques, and it was found that in all cases considered, the new method did at least as well as the established techniques considered. In some instances, the performance of the new test was superior in terms of power. For example, the new method outperformed Rotnitzky and Jewell's likelihood ratio test, when testing more than one parameter. Also, robust Wald tests were outperformed when correlated predictors were present, and only a subset of these were being tested.

In Chapter 5 we then proposed applying generalized estimating equations to space-time data. Under this approach, the temporal dependence was accounted

for via autoregressive covariates and the spatial dependence was modelled using working correlation structures of a spatial nature. This method has many appealing properties, for example, it can be used to model geometrically anisotropic and non-stationary processes. The method can also be applied to non-lattice data, and is computational efficient to implement, relative to other existing space-time approaches. Also, within Chapter 5 the application of the GEE one-step estimator was proposed within a large data context, to ease computational concerns.

Chapter 6 then considered a climate case study, involving wind speeds over northwestern Europe. Here the GLM methodology was applied to explore and identify important factors which impact upon wind speeds. This case study also enabled us to apply many of the techniques proposed earlier. In particular, we were able to apply, the new hypothesis testing technique, the space-time GEE approach, the GEE one-step estimator, in addition to Weibull and gamma GLMs. Comparisons of the various approaches were also undertaken. While the full GEE algorithm and the one-step algorithm appear to produce very similar results, the parameter estimates and standard errors obtained from the GEE and GLM approaches appeared to differ.

With regards to further work, there are several areas in which the work undertaken could be developed further. The first of these relates to the new hypothesis testing technique introduced in Chapter 4. Within this chapter, we considered the geometry of the test through specific examples and investigated the performance of the test using simulations. However, to gain a greater understanding of the test, more work needs be undertaken. This further work could take the form of additional simulations, formulated within other space-time settings. Varying the simulation parameters in relation to factors such as the design of the clusters, the extent and nature of the within cluster correlation and the covariate design would provide further insight into the performance of the test.

A second possible extension involves carrying out further analysis on esti-

mator comparison, for the various estimating techniques studied. In Chapter 6, the estimates obtained from the spatial GEE and gamma GLM differed significantly. This was somewhat surprising, since due to the sheer size of the data set, it was expected that asymptotic theory would take effect and result in much closer estimates. A possible reason for this discrepancy is that the asymptotic theory for the GEE approach is breaking down and biased estimates are being obtained. This claim could be investigated through theoretical work and simulations. In addition to the above, a full investigation into the performance of the GEE one-step algorithm, in comparison with the full GEE algorithm, could be undertaken. Since we are operating within a large data context, significant interest lies in trying to identify methods which provide comparable results for less computational effort. This work could also involve simulations.

Bibliography

Abramowitz, M. and Stegun, I. (1965). *Handbook of mathematical functions: with formulas, graphs and mathematical tables.* Dover, New York.

Aitkin, M. and Clayton, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics*, 29:156–63.

Albert, P. S. and McShane, L. M. (1995). A generalized estimating equations approach for spatially correlated binary data: Applications to the analysis of neuroimaging data. *Biometrics*, 51, No.2:627–638.

Bahadur, R. (1961). A representation of the joint distribution of responses to n dichotomous items. In Solomon, H., editor, *Studies in Item Analysis and Prediction, Volume VI, Stanford Mathematical Studies in the Social Sciences*, pages 158–168. Stanford University Press, Stanford, California.

Bowman, A. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations.* Oxford, Clarendon.

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25.

Brix, A. and Diggle, P. (2001). Spatiotemporal prediction for log-Guassian Cox processes. *Journal of the Royal Statistical Society, Series B*, 63, No.4:823–841.

Brown, P., Karesen, K., Roberts, G., and Tonellato, S. (2000). Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society, Series B*, 62, No.4:847–860.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier test: An expository note. *American Statistician*, 36:153–157.

Casella, G. and Berger, R. (2002). *Statistical Inference (second edition)*. Duxbury, Cambridge.

Chandler, R. (1998). Orthogonality. In Armitage, P. and Colton, T., editors, *Encyclopedia of Biostatistics*. Wiley, Chichester.

Chandler, R. and Wheater, H. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, 38, No.10:doi:10.1029/2001WR000906.

Chatfield, C. (2003). *The analysis of time series: an introduction (sixth edition)*. Chapman & Hall, London.

Coe, R. and Stern, R. (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology*, 21:1024–1031.

Conradsen, K., Nielsen, L., and Prahm, L. (1984). Review of Weibull statistics for estimation of wind speed distributions. *Journal of Climate and Applied Meteorology*, 23:1173–83.

Cox, D. (1972). The analysis of multivariate binary data. *Applied Statistics*, 21, No.2:113–120.

Cox, D. and Hinkley, D. (1974). *Theoretical statistics*. Chapman & Hall, London.

Cressie, N. (1991). *Statistics for spatial data*. Wiley, New York.

Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika*, 82:407–410.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002). *Analysis of longitudinal data (second edition).* Oxford University Press, Oxford.

Dobson, A. (2002). *An introduction to generalized linear models (second edition).* Chapman and Hall, London.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models (second edition).* Springer-Verlag, New York.

Fitzmaurice, G. (1995). A caveat concerning independence estimating equations with multivaraite binary data. *Biometrics*, 51:309–317.

Goodall, C. and Mardia, K. (1994). Challenges in multivariate spatio-temporal modeling. XVIIth International Biometric Conference, Ontario Canada.

Hardin, J. and Hilbe, J. (2002). *Generalized Estimating Equations.* Chapman & Hall/CRC, Boca Raton.

Haslett, J. and Raftery, A. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics*, 38:1–50.

Hughes, J., Guttorp, P., and Charles, S. (1999). A nonhomogeneous hidden Markov model for percipitation. *Applied Statistics*, 48:15–30.

Johnson, N. and Kotz, S. (1970). *Continuous univariate distributions - 2.* John Wiley & Sons, New York.

Journel, A. and Huijbregts, C. (1978). *Mining geostatistics.* Academic Press, New York.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.*, 77:437–471.

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.

Liang, K. and Zeger, S. (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science*, 10, No. 2:158–173.

Liang, K., Zeger, S., and Qaqish, B. (1992). Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society, Series B*, 54:3–40.

Lipsitz, S., Fitzmaurice, G., Orav, E., and Laird, N. (1994). Performance of generalized estimating equations in practical situations. *Biometrics*, 50:270–278.

McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics*, 11:59–67.

McCullagh, P. and Nelder, J. (1989). *Generalized linear models (second edition).* Chapman and Hall, London.

McDonald, B. (1993). Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society, Series B*, 55:391–397.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135:370–384.

Pan, W. (2001a). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57:120–125.

Pan, W. (2001b). On the robust variance estimator in generalized estimating equations. *Biometrika*, 88, No.3:901–906.

Pepe, M. and Anderson, G. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communication in Statistics*, 23:939–951.

Pickles, A. (1998). Generalized estimating equations. In Armitage, P. and Colton, T., editors, *Encyclopedia of Biostatistics.* Wiley, Chichester.

Pierce, D. and Schafer, D. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*, 81, No. 396:977–986.

Prentice, R. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44:1033–1048.

Priestley, M. (1981). *Spectral analysis and time series.* Academic Press, London.

Rotnitzky, A. and Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77:485–497.

Roy, A. and Clarke, D. (1988). *Astronomy: principles and practice (third edition).* Adam Hilger, Bristol.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78:719–727.

Stern, R. and Coe, R. (1984). A model fitting analysis of daily rainfall data (with discussion). *Journal of the Royal Statistical Society, Series A*, 147:1–34.

Stroud, J. R., Muller, P., and Sanso, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, Series B*, 63, No.4:673–689.

Sutradhar, B. and Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika*, 86:459–465.

Thall, P. and Vail, S. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46:657–671.

Thompson, D. and Wallace, J. (1998). The Arctic Oscillation signature in the winter-time geopotential height and temperature fields. *Geophys. Res. Lett.*, 25(9):1297–1300.

Tuller, S. and Brett, A. (1984). The characteristics of wind velocity that favour the fitting of a Weibull distribution in wind speed analysis. *Journal of Climate and Applied Meteorology*, 23:124–134.

Venables, W. N. and Ripley, B. (1994). *Modern applied statistics with S-Plus.* Springer-Verlag, New York.

Wackerley, D., Mendenhall, W., and Scheaffer, R. (2002). *Mathematical statistics with applications (sixth edition).* Duxbury, California.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–447.

Wheater, H., Isham, V., Onof, C., Chandler, R., Northrop, P., P.Guiblin, Bate, S., Cox, D., and Koutsoyiannis, D. (2000). Generation of spatially consistent rainfall data. Report to the Ministry of Agriculture, Fisheries and Food (2 volumes). Also available as Research Report No. 204, Department of Statistical Science, University College London (http://www.ucl.ac.uk/Stats/research/abstracts.html).

Wikle, C., Berliner, M., and Cressie, N. (1999). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5:117–154.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a psuedo-likelihood approach. *Journal of Statistical Computing and Simulation*, 48:233–243.

Yan, Z., Bate, S., Chandler, R., Isham, V., and Wheater, H. (2002). An analysis of daily maximum wind speed in northwestern Europe with generalized linear modelling. *Journal of Climate*, 15:2073–2088.

Zeger, S. and Karim, M. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, 86:79–86.

Zeger, S. and Liang, K. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130.

Zeger, S., Liang, K., and Albert, P. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44:1049–1060.

Appendix A

Gamma GLM coefficient estimates

Main effects:	Coefficient Estimate	Robust Std Err
Constant	0.586264	0.011115
Landmark (0 sea - land 1)	-0.061695	0.006498
Legendre polynomial 1 for Eastings	-0.074910	0.010064
Legendre polynomial 1 for Northings	0.110178	0.009079
Legendre polynomial 1 for Elevation	0.003913	0.001225
Legendre polynomial 2 for Eastings	0.053014	0.009530
Legendre polynomial 2 for Northings	0.003789	0.007619
Legendre polynomial 2 for Elevation	0.016201	0.000753
Legendre polynomial 3 for Eastings	-0.028675	0.001405
Legendre polynomial 3 for Northings	-0.005976	0.001339
Legendre polynomial 3 for Elevation	-0.048346	0.000558
Legendre polynomial 4 for Eastings	0.003826	0.001027
Legendre polynomial 4 for Northings	0.024583	0.000919
NHT	-0.014466	0.016882
SHT	0.023048	0.015456
SOI	0.002793	0.002741
NAO	0.012242	0.007050
EA	-0.000654	0.004071
EAJ	0.005394	0.002541
EP	-0.001446	0.003294
NP	0.001452	0.002212

PNA	0.005418	0.003990
EAWR	-0.006141	0.003957
SCA	-0.007616	0.003813
TNH	-0.004682	0.002419
NAT	0.001732	0.007863
SAT	-0.000664	0.005797
AO	0.016114	0.008032
August effect	-0.016203	0.004312
ln(1+Previous days weighted neighbourhood)	0.705660	0.003713
ln(1+Weighted neighbourhood 2 days before)	-0.111034	0.003969
ln(1+Weighted neighbourhood 3 days before)	0.059496	0.003557
ln(1+Weighted neighbourhood 4 days before)	0.019506	0.003444
ln(1+Weighted neighbourhood 5 days before)	0.006885	0.003462
ln(1+Weighted neighbourhood 6 days before)	0.019825	0.003173
Daily seasonal effect, cosine component	0.003618	0.010956
Daily seasonal effect, sine component	-0.006251	0.001881
Daily half-year cycle, cosine component	-0.008917	0.001558
Daily half-year cycle, sine component	-0.040660	0.007059

2-way interactions:	Coefficient Estimate	Robust Std Err
Landmark (0 sea - land 1)	0.010360	0.001886
with Legendre polynomial 1 for Northings		
Landmark (0 sea - land 1)	-0.023959	0.002062
with Legendre polynomial 2 for Eastings		
Landmark (0 sea - land 1)	0.011014	0.001209
with Legendre polynomial 2 for Northings		
Landmark (0 sea - land 1)	0.206264	0.013210
with Legendre polynomial 3 for Eastings		
Landmark (0 sea - land 1)	0.021178	0.001582
with Legendre polynomial 3 for Northings		
Landmark (0 sea - land 1)	-0.114491	0.011208
with Legendre polynomial 4 for Eastings		
Landmark (0 sea - land 1)	-0.011253	0.003653
with Previous days weighted neighbourhood		
Landmark (0 sea - land 1)	0.021303	0.003993
with Weighted neighbourhood 2 days before		
Legendre polynomial 1 for Eastings	-0.020186	0.002717
with Legendre polynomial 1 for Northings		
Legendre polynomial 1 for Eastings	-0.049758	0.002433
with Legendre polynomial 2 for Northings		
Legendre polynomial 1 for Eastings	0.027072	0.001367
with Legendre polynomial 4 for Northings		
Legendre polynomial 1 for Eastings	0.004595	0.004184

with Previous days weighted neighbourhood		
Legendre polynomial 1 for Eastings	-0.013997	0.002486
with Daily seasonal effect, cosine component		
Legendre polynomial 1 for Northings	0.027805	0.011454
with Legendre polynomial 2 for Eastings		
Legendre polynomial 1 for Northings	0.028008	0.001929
with Legendre polynomial 3 for Eastings		
Legendre polynomial 1 for Northings	0.028315	0.001335
with Legendre polynomial 4 for Eastings		
Legendre polynomial 1 for Northings	-0.033078	0.003806
with Previous days weighted neighbourhood		
Legendre polynomial 1 for Northings	0.012653	0.002042
with Daily seasonal effect, cosine component		
Legendre polynomial 1 for Northings	-0.010872	0.001822
with Daily seasonal effect, sine component		
Legendre polynomial 2 for Eastings	-0.022434	0.002054
with Legendre polynomial 2 for Northings		
Legendre polynomial 2 for Eastings	-0.035402	0.001435
with Legendre polynomial 3 for Northings		
Legendre polynomial 2 for Eastings	-0.032583	0.003886
with Previous days weighted neighbourhood		
Legendre polynomial 2 for Northings	0.022229	0.001776
with Legendre polynomial 3 for Eastings		
Legendre polynomial 2 for Northings	-0.023148	0.001412
with Legendre polynomial 4 for Eastings		
Legendre polynomial 2 for Northings	-0.019318	0.003168
with Weighted neighbourhood 3 days before		
Legendre polynomial 3 for Eastings	-0.022599	0.001058
with Legendre polynomial 3 for Northings		
Legendre polynomial 3 for Eastings	-0.027058	0.001283
with Legendre polynomial 4 for Northings		
Daily seasonal effect, cosine component	-0.011420	0.004463
with Previous days weighted neighbourhood		
Daily seasonal effect, cosine component	0.034282	0.004980
with Weighted neighbourhood 2 days before		
Daily seasonal effect, cosine component	0.008431	0.004412
with Weighted neighbourhood 3 days before		
Daily seasonal effect, cosine component	0.013963	0.009439
with NHT		
Daily seasonal effect, cosine component	0.007260	0.006885
with NAO		
Daily seasonal effect, cosine component	0.008169	0.003880
with EA		
Daily seasonal effect, cosine component	-0.010709	0.003393
with EAJ		

Daily seasonal effect, cosine component with AO	0.005555	0.005839
Daily seasonal effect, sine component with NHT	0.044024	0.015792
Daily seasonal effect, sine component with NAO	-0.012888	0.007268
Daily seasonal effect, sine component with EA	-0.008412	0.004290
Daily seasonal effect, sine component with NAT	-0.020932	0.009694
Daily seasonal effect, sine component with SAT	-0.014821	0.005618
Daily seasonal effect, sine component with AO	0.006050	0.007409
Daily half-year cycle, sine component with Previous days weighted neighbourhood	0.018181	0.002949

3-way interactions:	Coefficient Estimate	Robust Std Err
Landmark (0 sea - land 1) with Legendre polynomial 3 for Eastings and Weighted neighbourhood 2 days before	-0.031494	0.005927
Landmark (0 sea - land 1) with Legendre polynomial 4 for Eastings and Weighted neighbourhood 2 days before	0.019281	0.005012
Legendre polynomial 1 for Northings with Legendre polynomial 2 for Eastings and Previous days weighted neighbourhood	-0.022374	0.004885
Landmark (0 sea - land 1) with Legendre polynomial 1 for Eastings and NHT	0.024908	0.014622
Landmark (0 sea - land 1) with Legendre polynomial 3 for Eastings and NHT	-0.021595	0.010608
Legendre polynomial 1 for Northings with Legendre polynomial 2 for Eastings and NHT	-0.003946	0.008985
Legendre polynomial 1 for Northings with Legendre polynomial 3 for Eastings and NHT	0.024142	0.007430
Landmark (0 sea - land 1) with Legendre polynomial 1 for Eastings and SHT	-0.029194	0.018102
Landmark (0 sea - land 1)	0.087907	0.017472

with Legendre polynomial 1 for Northings		
and SHT		
Landmark (0 sea - land 1)	0.037098	0.012878
with Legendre polynomial 3 for Eastings		
and SHT		
Legendre polynomial 1 for Eastings	-0.040853	0.014987
with Legendre polynomial 1 for Northings		
and SHT		
Legendre polynomial 1 for Northings	0.008120	0.002627
with Legendre polynomial 2 for Eastings		
and SOI		
Landmark (0 sea - land 1)	-0.019122	0.005412
with Legendre polynomial 1 for Eastings		
and NAO		
Landmark (0 sea - land 1)	0.002699	0.008440
with Legendre polynomial 1 for Northings		
and NAO		
Legendre polynomial 1 for Eastings	-0.016731	0.008403
with Legendre polynomial 1 for Northings		
and NAO		
Landmark (0 sea - land 1)	-0.023574	0.005078
with Legendre polynomial 1 for Northings		
and EA		
Legendre polynomial 1 for Eastings	0.014995	0.005430
with Legendre polynomial 1 for Northings		
and EA		
Legendre polynomial 1 for Northings	-0.008935	0.003480
with Legendre polynomial 3 for Eastings		
and EA		
Legendre polynomial 1 for Northings	-0.008548	0.003395
with Legendre polynomial 2 for Eastings		
and EP		
Legendre polynomial 1 for Eastings	-0.007460	0.004899
with Legendre polynomial 1 for Northings		
and PNA		
Landmark (0 sea - land 1)	0.010079	0.005450
with Legendre polynomial 1 for Eastings		
and EAWR		
Landmark (0 sea - land 1)	-0.021317	0.007942
with Legendre polynomial 1 for Northings		
and NAT		
Legendre polynomial 1 for Northings	0.008129	0.006040
with Legendre polynomial 2 for Eastings		
and NAT		
Landmark (0 sea - land 1)	-0.008100	0.006307

with Legendre polynomial 1 for Northings		
and SAT		
Landmark (0 sea - land 1)	0.024026	0.007106
with Legendre polynomial 1 for Northings		
and A0		
Landmark (0 sea - land 1)	0.004663	0.003194
with Legendre polynomial 3 for Eastings		
and A0		
Legendre polynomial 1 for Eastings	-0.010640	0.007507
with Legendre polynomial 1 for Northings		
and A0		
Legendre polynomial 1 for Eastings	0.003943	0.003035
with Legendre polynomial 2 for Northings		
and A0		
Legendre polynomial 1 for Northings	0.006212	0.003266
with Legendre polynomial 3 for Eastings		
and A0		

Appendix B

Spatial GEE coefficient estimates

Main effects:	Coefficient Estimate	Robust Std Err
Constant	0.352266	0.008424
Landmark (0 sea - land 1)	-0.042684	0.002057
Legendre polynomial 1 for Eastings	-0.319110	0.007814
Legendre polynomial 1 for Northings	0.089923	0.007363
Legendre polynomial 1 for Elevation	-0.019290	0.000676
Legendre polynomial 2 for Eastings	0.056178	0.006502
Legendre polynomial 2 for Northings	0.099198	0.005389
Legendre polynomial 2 for Elevation	0.013548	0.000351
Legendre polynomial 3 for Eastings	-0.023658	0.001029
Legendre polynomial 3 for Northings	0.009422	0.001007
Legendre polynomial 3 for Elevation	-0.017845	0.000301
Legendre polynomial 4 for Eastings	-0.005536	0.000755
Legendre polynomial 4 for Northings	0.002767	0.000789
NHT	0.001267	0.015348
SHT	-0.008627	0.013652
SOI	0.000966	0.002567
NAO	0.005964	0.006407
EA	0.005904	0.003690
EAJ	0.004822	0.002374
EP	-0.002828	0.003042
NP	0.002100	0.002083
PNA	0.004954	0.003749
EAWR	0.000118	0.003529
SCA	-0.003949	0.003603

TNH	-0.000866	0.002246
NAT	0.006432	0.007186
SAT	0.005091	0.005315
AO	0.008972	0.007386
August effect	-0.010430	0.004040
ln(1+Previous days weighted neighbourhood)	0.542549	0.002851
ln(1+Weighted neighbourhood 2 days before)	0.032149	0.002407
ln(1+Weighted neighbourhood 3 days before)	0.074840	0.002352
ln(1+Weighted neighbourhood 4 days before)	0.051984	0.002243
ln(1+Weighted neighbourhood 5 days before)	0.039047	0.002210
ln(1+Weighted neighbourhood 6 days before)	0.062578	0.002084
Daily seasonal effect, cosine component	-0.020568	0.009183
Daily seasonal effect, sine component	-0.001288	0.001788
Daily half-year cycle, cosine component	-0.010212	0.001445
Daily half-year cycle, sine component	-0.019200	0.006788

2-way interactions:	Coefficient Estimate	Robust Std Err
Landmark (0 sea - land 1)	0.023285	0.000755
with Legendre polynomial 1 for Northing		
Landmark (0 sea - land 1)	-0.025033	0.000980
with Legendre polynomial 2 for Eastings		
Landmark (0 sea - land 1)	0.003573	0.000577
with Legendre polynomial 2 for Northing		
Landmark (0 sea - land 1)	0.132796	0.004524
with Legendre polynomial 3 for Eastings		
Landmark (0 sea - land 1)	0.012552	0.000642
with Legendre polynomial 3 for Northing		
Landmark (0 sea - land 1)	-0.031702	0.004072
with Legendre polynomial 4 for Eastings		
Landmark (0 sea - land 1)	-0.004832	0.000915
with Previous days weighted neighbourhood		
Landmark (0 sea - land 1)	0.017646	0.000989
with Weighted neighbourhood 2 days before		
Legendre polynomial 1 for Eastings	-0.024140	0.002154
with Legendre polynomial 1 for Northing		
Legendre polynomial 1 for Eastings	-0.029528	0.002166
with Legendre polynomial 2 for Northing		
Legendre polynomial 1 for Eastings	0.009783	0.001223
with Legendre polynomial 4 for Northing		
Legendre polynomial 1 for Eastings	0.116774	0.003175
with Previous days weighted neighbourhood		
Legendre polynomial 1 for Eastings	-0.024328	0.002018
with Daily seasonal effect, cosine		

Legendre polynomial 1 for Northings	0.007842	0.007576
with Legendre polynomial 2 for Eastings		
Legendre polynomial 1 for Northings	0.016271	0.001402
with Legendre polynomial 3 for Eastings		
Legendre polynomial 1 for Northings	0.006531	0.000977
with Legendre polynomial 4 for Eastings		
Legendre polynomial 1 for Northings	-0.028315	0.003081
with Previous days weighted neighbourhood		
Legendre polynomial 1 for Northings	0.008318	0.001801
with Daily seasonal effect, cosine		
Legendre polynomial 1 for Northings	-0.006540	0.001649
with Daily seasonal effect, sine		
Legendre polynomial 2 for Eastings	-0.027429	0.001721
with Legendre polynomial 2 for Northing		
Legendre polynomial 2 for Eastings	-0.020946	0.001134
with Legendre polynomial 3 for Northing		
Legendre polynomial 2 for Eastings	-0.017691	0.002633
with Previous days weighted neighbourhood		
Legendre polynomial 2 for Northings	0.016557	0.001377
with Legendre polynomial 3 for Eastings		
Legendre polynomial 2 for Northings	-0.009424	0.001006
with Legendre polynomial 4 for Eastings		
Legendre polynomial 2 for Northings	-0.049307	0.002204
with Weighted neighbourhood 3 days before		
Legendre polynomial 3 for Eastings	-0.004871	0.000828
with Legendre polynomial 3 for Northing		
Legendre polynomial 3 for Eastings	-0.016797	0.001003
with Legendre polynomial 4 for Northing		
Daily seasonal effect, cosine component	-0.000996	0.003785
with Previous days weighted neighbourhood		
Daily seasonal effect, cosine component	0.022530	0.003384
with Weighted neighbourhood 2 days before		
Daily seasonal effect, cosine component	0.015611	0.002997
with Weighted neighbourhood 3 days before		
Daily seasonal effect, cosine component	0.007764	0.008856
with NHT		
Daily seasonal effect, cosine component	0.002937	0.006511
with NAO		
Daily seasonal effect, cosine component	0.006138	0.003626
with EA		
Daily seasonal effect, cosine component	-0.005781	0.003176
with EAJ		
Daily seasonal effect, cosine component	0.001921	0.005468
with AO		
Daily seasonal effect, sine component	0.017149	0.014477

with NHT		
Daily seasonal effect, sine component	-0.007836	0.006729
with NAO		
Daily seasonal effect, sine component	-0.005188	0.003987
with EA		
Daily seasonal effect, sine component	-0.010458	0.008967
with NAT		
Daily seasonal effect, sine component	-0.006463	0.005176
with SAT		
Daily seasonal effect, sine component	0.005989	0.006844
with AO		
Daily half-year cycle, sine component	0.008957	0.002821
with Previous days weighted neighbourhood		

3-way interactions:	Coefficient Estimate	Robust Std Err
Landmark (0 sea - land 1)	-0.033404	0.0019
with Legendre polynomial 3 for Eastings		
and Weighted neighbourhood 2 days before		
Landmark (0 sea - land 1)	-0.000936	0.0017
with Legendre polynomial 4 for Eastings		
and Weighted neighbourhood 2 days before		
Legendre polynomial 1 for Northings	-0.018031	0.0032
with Legendre polynomial 2 for Eastings		
and Previous days weighted neighbourhood		
Landmark (0 sea - land 1)	-0.025574	0.0030
with Legendre polynomial 1 for Eastings		
and NHT		
Landmark (0 sea - land 1)	0.010893	0.0039
with Legendre polynomial 3 for Eastings		
and NHT		
Legendre polynomial 1 for Northings	0.005963	0.0064
with Legendre polynomial 2 for Eastings		
and NHT		
Legendre polynomial 1 for Northings	0.016046	0.0052
with Legendre polynomial 3 for Eastings		
and NHT		
Landmark (0 sea - land 1)	-0.002843	0.0040
with Legendre polynomial 1 for Eastings		
and SHT		
Landmark (0 sea - land 1)	0.004278	0.0051
with Legendre polynomial 1 for Northings		
and SHT		
Landmark (0 sea - land 1)	-0.010937	0.0046

with Legendre polynomial 3 for Eastings		
and SHT		
Legendre polynomial 1 for Eastings	-0.029324	0.0109
with Legendre polynomial 1 for Northings		
and SHT		
Legendre polynomial 1 for Northings	0.006361	0.0019
with Legendre polynomial 2 for Eastings		
and SOI		
Landmark (0 sea - land 1)	0.000795	0.0013
with Legendre polynomial 1 for Eastings		
and NAO		
Landmark (0 sea - land 1)	-0.000496	0.0024
with Legendre polynomial 1 for Northings		
and NAO		
Legendre polynomial 1 for Eastings	-0.006973	0.0060
with Legendre polynomial 1 for Northings		
and NAO		
Landmark (0 sea - land 1)	0.002687	0.0013
with Legendre polynomial 1 for Northings		
and EA		
Legendre polynomial 1 for Eastings	0.001232	0.0042
with Legendre polynomial 1 for Northings		
and EA		
Legendre polynomial 1 for Northings	-0.005235	0.0024
with Legendre polynomial 3 for Eastings		
and EA		
Legendre polynomial 1 for Northings	-0.007658	0.0025
with Legendre polynomial 2 for Eastings		
and EP		
Legendre polynomial 1 for Eastings	-0.010362	0.0040
with Legendre polynomial 1 for Northings		
and PNA		
Landmark (0 sea - land 1)	-0.000939	0.0010
with Legendre polynomial 1 for Eastings		
and EAWR		
Landmark (0 sea - land 1)	0.007263	0.0024
with Legendre polynomial 1 for Northings		
and NAT		
Legendre polynomial 1 for Northings	0.001295	0.0043
with Legendre polynomial 2 for Eastings		
and NAT		
Landmark (0 sea - land 1)	0.001873	0.0017
with Legendre polynomial 1 for Northings		
and SAT		
Landmark (0 sea - land 1)	0.008868	0.0018

with Legendre polynomial 1 for Northings		
and A0		
Landmark (0 sea - land 1)	0.000076	0.0011
with Legendre polynomial 3 for Eastings		
and A0		
Legendre polynomial 1 for Eastings	-0.006343	0.0055
with Legendre polynomial 1 for Northings		
and A0		
Legendre polynomial 1 for Eastings	-0.004342	0.0024
with Legendre polynomial 2 for Northings		
and A0		
Legendre polynomial 1 for Northings	-0.000882	0.0023
with Legendre polynomial 3 for Eastings		
and A0		