Universidade Federal de Pernambuco
Centro de Ciências Exatas e Naturais
Departamento de Estatística

Programa de Pós-Graduação em Estatística

INFERENCE AND DIAGNOSTICS IN SPATIAL MODELS

Fernanda De Bastiani

Supervisor: **Professor Dra. Audrey Helen Mariz de Aquino Cysneiros**

Co-supervisor: **Professor Dr. Miguel Angel Uribe Opazo**

Doctoral thesis

Recife, 2016

Inference and diagnostics in spatial models

Fernanda De Bastiani

Doctoral thesis submitted to the Graduate Program in Statistics, Department of Statistics, Universidade Federal de Pernambuco as a requirement to obtain a PhD. in Statistics.

During the development of this work the author received scholarship from FACEPE and Capes (sandwich doctorate at London Metropolitan University under the supervision of Professor Dr. Dimitrios Mikis Stasinopoulos)

Recife, 2016.

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

D278i De Bastiani, Fernanda.
Inference and diagnostics in spatial models / Fernanda De Bastiani. – 2016.
182 f.: il., fig., tab.

Orientadora: Audrey Helen Mariz de Aquino Cysneiros.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2016.
Inclui referências.

1. Processos estocásticos. 2. Estatística espacial. 3. Modelos de regressão. 4. Inferência estatística. I. Cysneiros, Audrey Helen Mariz de Aquino (orientadora). II. Título.

519.23 CDD (23. ed.) UFPE- MEI 2016-033

FERNANDA DE BASTIANI

INFERENCE AND DIAGNOSTICS IN SPATIAL MODELS

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutora em Estatística.

Aprovada em: 22 de fevereiro de 2016.

BANCA EXAMINADORA

Prof.ª Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof. Francisco José de Azevedo Cysneiros
UFPE

Prof. Klaus Leite Pinto Vasconcellos
UFPE

Prof.ª Dione Maria Valença
UFRN

Prof.ª Verônica Cadena Lima
UFBA

For my parents and my brother.

Acknowledgements

I would like to thank my brother (Felipe), my parents (Nelson and Angélica), my grandparents (Rosalina, Máximo, Azélia and Mario) and my whole family for their unceasing support throughout my life.

My gratitude to my supervisors, Professor Dra. Audrey Helen Mariz de Aquino Cysneiros from Universidade Federal de Pernambuco (UFPE) and Professor Dr. Miguel Angel Uribe Opazo from Universidade Estadual do Oeste do Paraná (UNIOESTE), for their invaluable guidance and direction of this research, encouragement, support, patience, and critical reviews of this work. I am also grateful for the great support of Professor Dr. Manuel Galea from Pontificia Universidad Católica de Chile (PUC), Professor Dr. Dimitrios Mikis Stasinopoulos and Professor Dr. Robert Anthony Rigby from London Metropolitan University. I would like to thank you their considerable contribution to the topics, conversations, patience, encouragement and support.

I would also like to thank you the Professors from UFPE for their incentive, Prof. Dr. Francisco Cribari Neto, Prof. Dr. Franciso José de Azevêdo Cysneiros, Prof. Dra. Betsabé Blas Achic, Prof. Dr. Sérgio de Carvalho Bezerra, Prof. Dra. Claudia Regina Oliveira de Paiva Lima, Prof. Dra. Maria Cristina Falcão Raposo, Prof. Dr. Getúlio José Amorim do Amaral, Prof. Dr. Raydonal Ospina Martinez, Prof. Dr. Gauss Moutinho Cordeiro and Prof. Dr. Klaus Leite Pinto Vasconcelos. And some Professors from other universities that shared some of their experiences with me. Thank you for the examiners for their contribution to improve the work presented and the suggestions about topics for future research.

My UFPE classmates. Thank you for all the staff member, and a special thank you to Valéria, she is really efficient. Friends from the spatial statistics laboratory at UNIOESTE, from London Metropolitan University, PUC and from different places. I will not mention here all the names, because I am afraid of forgetting someone. My friends know how much they are important for me.

My PhD scholarship from FACEPE, financial support for sandwich doctorate from CAPES, and financial support from FACEPE, Capes, CNPq, Fundação Araucária and UFPE to participate in scientific conferences.

There are so many persons that contributed to my professional and personal experience during the development of this thesis, that I would like to say my heartiest thanks to everyone.

Thanks God, I finished!

Resumo

Neste trabalho, apresentamos inferência e diagnósticos para modelos espaciais. Inicialmente, os modelos espaciais lineares Gaussianos são estendidos para os modelos espaciais lineares elípticos, e desenvolve-se a metodologia de influência local para avaliar a sensibilidade dos estimadores de máxima verossimilhança para pequenas perturbações nos dados e/ou nos pressupostos do modelo. Posteriormente, considera-se os modelos espaciais lineares Gaussianos com repetições. Para estes modelos obteve-se em notação matricial um fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas. É também realizada inferência para estimar o parâmetro de suavização da classe de modelos da família Matérn. Os estimadores de máxima verossimilhança são obtidos, e uma expressão explícita para a matriz de informação de Fisher é apresentada, mesmo quando o parâmetro de suavização da classe de modelos da família Matérn da estrutura de covariância é estimado. Desenvolve-se técnicas de diagnósticos de influência local e global para avaliar a influência de observações em modelos espaciais lineares Gaussianos com repetições. Os conceitos de distância de Cook e alavanca generalizada são revisados e estendidos para estes modelos. Para influência local são consideradas perturbações apropriadas na variável resposta e ponderação de casos. Finalmente, é descrita a modelagem para os componentes espaciais dos campos aleatórios Markovianos nos modelos aditivos generalizados de locação escala e forma. Isto permite modelar qualquer ou todos os parâmetros da distribuição para a variável resposta utilizando as variáveis explanatórias e efeitos espaciais. Alguns estudos de simulações são apresentados e as metodologias são ilustradas com conjuntos de dados reais.

Palavras-chave: Correções de Bartlett. Distribuições elípticas. GAMLSS. Geoestatística. Influência global. Influência local. Máxima verossimilhança. Medidas repetidas.

Abstract

In this work, we present inference and diagnostics in spatial models. Firstly, we extend the Gaussian spatial linear model for the elliptical spatial linear models, and present the local influence methodology to assess the sensitivity of the maximum likelihood estimators to small perturbations in the data and/or the spatial linear model assumptions. Secondly, we consider the Gaussian spatial linear models with repetitions. We obtain in matrix notation a Bartlett correction factor for the profiled likelihood ratio statistic. We also present inference approach to estimate the smooth parameter from the Matérn family class of models. The maximum likelihood estimators are obtained, and an explicit expression for the Fisher information matrix is also presented, even when the smooth parameter for Matérn class of covariance structure is estimated. We present local and global influence diagnostics techniques to assess the influence of observations on Gaussian spatial linear models with repetitions. We review concepts of Cook's distance and generalized leverage and extend it. For local influence we consider two different approach and for both we consider appropriated perturbation in the response variable and case weight perturbation. Finally, we describe the modeling and fitting of Markov random field spatial components within the generalized additive models for locations scale and shape framework. This allows modeling any or all of the parameters of the distribution for the response variable using explanatory variables and spatial effects. We present some simulations and real data sets illustrate the methodology.
Keywords: Bartlett correction. Elliptical distributions. GAMLSS. Geostatistical. Global influence. Local influence. Maximum likelihood. Repeated measures.

List of Figures

List of Tables

Contents

Chapter 1

Preliminary

1.1 Resumo

Este capítulo apresenta uma introdução e breve revisão bibliográfica sobre inferência e diagnósticos em modelos espaciais. Apresenta os principais objetivos do trabalho, a organização do mesmo e os recursos computacionais utilizados.

1.2 Introduction

Spatial statistics are useful in subjects as diverse as climatology, ecology, economics, environmental and earth sciences, epidemiology, image analysis, agriculture and more. This field of study is concerning statistical methods that use space and spatial relationships such as distance, area, volume, length, height, orientation, centrality and more. The most well known branches in spatial statistics are discrete spatial variation, spatial point processes and geostatistics. Gaetan and Guyon (2010) covers these types of spatial analysis.

The study of discrete spatial variation, where the variables are defined on discrete domains, such as regions, regular grids or lattices, are studied by the Markov random field theory. Extensive theoretical and practical details are provided by Rue and Held (2005). Another area in spatial statistics is the spatial point process. A spatial point process is a set of locations, irregularly distributed within a designated region and presumed to have been generated by some form of stochastic mechanism, (Diggle, 2003).

Geostatistics study a response variable (and potentially explanatory variables) that are measured at points in space. Important work by Krige (1951) and Matheron (1963) laid the foundation for the field of geostatistics where some of the first methods for modelling spatial dependence were proposed, see (Schabenberger and Gotway, 2005) for more details. The methodology developed there after is referred in the literature as “kriging”. Isaaks and Srisvastava (1989) says that geostatistics offers a way of describing the spatial continuity of natural phenomena and provides adaptations of classical regression techniques to take advantage of this continuity.

For more details about inference methods and applications of these models, see for example, (Mardia and Marshall, 1984) that described the maximum likelihood method for fitting the linear model when residuals are correlated and when the covariance among the residuals is determined by a parametric model containing unknown parameters. Isaaks and Srisvastava (1989) and Cressie (1993) contents coverage the principles of Geostatistics. Waller and Gotway (2004) provided a text that moves from a basic understanding of multiple linear regression to an application-oriented introduction to statistical methods used to analyze spatially referenced health data. Zhang (2004) showed inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Diagnostic techniques are discussed in Cerioli and Riani (1999), Militino et al. (2006), Uribe-Opazo et al. (2012) and Filzmoser et al. (2014). Among references which have taken a variety of approaches we mention Stein (1999), Journel and Huijbregts (2004), Schabenberger and Gotway (2005), Webster and Oliver (2007) and Diggle and Ribeiro Jr. (2007).

Problems of identifiability sometimes are present in geostatistical models. Cressie (1993) presents a practical illustration of non-identifiability of the deterministic and stochastic components contributing to a data set. Genton and Zhang (2012) showed some identifiability problems in skew-Gaussian and elliptical spatial linear models. Perrin and Meiring (1999) discussed identifiability issues for the class of models with gemometrical anisotropy. In the case of multivariate geostatistical models, Diggle and Ribeiro Jr. (2007) said that even very simple multivariate constructions quickly lead to models with either large numbers of parameters and consequent problems of poor identifiability, or potentially severe restrictions on the allowable form of cross-correlation structure. According to the same authors, replicated observations are needed at each sampling location in order to identify the required transformation.

There are some works that dealed with non-Gaussian and/or bayesian geostatistical models, for instance, a model-based approach to geostatistics for non-Gaussian data based on generalized linear mixed models has been advanced by Diggle et al. (1998) in a Bayesian inferential framework, see also (Banerjee et al., 2014; Diggle and Ribeiro Jr., 2007). Minozzo and Fruttini (2004) proposed an extension of the proportional covariance model following a hierarchical model-based approach. Zhu et al. (2005) proposed maximum likelihood inference using a Monte Carlo EM algorithm for a spatio-temporal framework. Reich and Fuentes (2007) considered a framework in which they used a stick-breaking prior in a semiparametric Bayesian context. Other references are Bailey and Krzanowski (2000) and Christensen and Amemiya (2002).

An alternative to kriging in geostatistics is the smoothing techniques popularized by Hastie and Tibshirani (1990), and also by the P-spline approach of Eilers and Marx (1996). The P-spline models were extended to smoothing spatial data, which requires use of tensor product and row-wise Kronecker product (Eilers (2003); Currie et al. (2006); Eilers and Marx (2010)). Thin plate regression splines are another candidate since they are invariant to rotation of the covariate space, Wood (2006).

The generalized additive models for location scale and shape (GAMLSS) introduced by Rigby and Stasinopoulos (2005) provide spatial modelling facilities and a very general and flexible system for modelling a response variable. GAMLSS are (semi) parametric univariate regression models, where all the parameters of the assumed distribution for the response can be modelled as additive functions of the explanatory variables. The addtive terms that can be incorporated are for instance, P-splines, cubic splines, simple random effects and varying coefficient.

GAMLSS has been used in a variety of fields including: actuarial science, biology, biosciences, energy economic, genomics, finance, fisheries, food consumption, growth curves estimation, marine research, medicine, meteorology, rainfalls, vaccines. Beyerlein et al. (2008) present GAMLSS to assess increase in childhood body mass index. Fenske et al. (2008) used GAMLSS in the detection of risk factors for obesity in early childhood with quantile regression methods for longitudinal data. Verschuren et al. (2010) used GAMLSS To establish reference values and reference curves for anaerobic performance and agility in ambulatory children and adolescents with cerebral palsy.

1.3 Objective and organization

The aim of this work is to present different frameworks and techniques to model spatial data. The whole thesis is written with independent chapters and each chapter has new research contribution. The Chapters are related in the sense they talk about spatial modelling and the notation is consistent inside each Chapter. Chapter 2 presents the elliptical spatial linear models for one single realization, an extension of the Gaussian spatial linear model and presents the local influence methodology. Chapters 3, 4 and 5 present inference, local influence diagnostics and global diagnostics, respectively, on Gaussian spatial linear models with repetitions. These Chapters are related to geostatistical models. Chapter 6 presents the GAMLSS in the scope of Markov random fields.

More specifically, in Chapter 2 we extend the Gaussian spatial linear model to the elliptical spatial linear models, for the case where we have no repetitions. For this case we also use the local influence methodology to assess the sensitivity of the maximum likelihood estimators to small perturbations in the data and/or the spatial linear model assumptions. In Chapter 3 we present inference techniques in Gaussian spatial linear models with repetitions. Mainly we present hypothesis test and estimation approach to estimate the smooth parameter from the Matérn family class of models. The maximum likelihood estimators are obtained, and an explicit expression for the Fisher information matrix is also presented, even when the smooth parameter for Matérn class of covariance structure is estimated. Chapter 4 and 5 present local and global diagnostics techniques, respectively, to assess the influence of observations on Gaussian spatial linear models with repetitions. We review concepts of Cook's distance based on the likelihood and Q-function, local influence for two different approaches and generalized leverage. Moreover, for local influence we consider appropriate perturbation in the response variable, in the scale matrix and case weight perturbation. Chapter 6 describes the modelling and fitting of Markov random field spatial components within the GAMLSS framework. The methodologies are illustrated with a real data set, and for some situations we present simulations. In Chapter 7 we present some conclusion remarks and future research. Each Chapter presents own

Appendices and a general Bibliography is presented in the end.

1.4 Computing plataform

We implemented the obtained results in the `R` statistical software R Core Team (2015). For the Monte Carlo simulation presented in Chapter 3 we use `Ox` matrix programming language (Cribari-Neto and Zarkos, 2003; Doornik, 2006). For graphical representation of results, we use the `R` software. For Chapter 6 we created the new package `gamlss.spatial` and the latest version of this software is freely available at `http://www.R-project.org`.

Chapter 2

Influence diagnostics in elliptical spatial linear models

2.1 Resumo

Neste capítulo, os modelos espaciais lineares Gaussianos foram estendidos para a família de distribuições elípticas, a qual foi considerada para estimar os parâmetros que definem a estrutura de dependência espacial em dados georreferenciados. Além disso, por meio da metodologia de influência local foi possível avaliar a sensibilidade dos estimadores de máxima verossimilhança a pequenas perturbações nos dados e/ou nas suposições do modelo espacial linear. A metodologia foi ilustrada com um conjunto de dados reais. Os resultados permitiram concluir que a presença de observações atípicas no conjunto de dados amostrados tem forte influência, alterando a estrutura de dependência espacial. Também foi incluído um estudo de simulação.

2.2 Introduction

Spatial statistics is a rapidly developing field which involves the quantitative analysis of spatial data and the statistical modelling of spatial variability and uncertainty. Applications of spatial statistics can be found not only for environmental disciplines such as agriculture, geology, soil science, hydrology, ecology, oceanography, forestry, meteorology and climatology, but also socio-economic disciplines such as human geography, spatial econometrics, epidemiology and spatial planning. Recent proposals have discussed the use of Gaussian spatial linear models to study the structure of dependence in spatially referenced data. For more details about estimation, inference methods and applications of these models, see, for example, (Cressie, 1993; Isaaks and Srisvastava, 1989; Mardia and Marshall, 1984; Schabenberger and Gotway, 2005; Waller and Gotway, 2004; Webster and Oliver, 2007). Diagnostic techniques are discussed in (Cerioli and Riani, 1999; Filzmoser et al., 2014; Militino et al., 2006; Uribe-Opazo et al., 2012).

Certainly the multivariate normal distribution is useful in many cases, and most of the statistical inference for continuous variables has been developed under the assumption of normality. This is particularly the case of multivariate analysis, mixed effects models and linear regression. However it is well known that the normal distribution is not always suitable for modelling multivariate continuous data (Lange and Sinsheimer, 1993; Lange et al., 1989). In this Chapter we extend the model proposed by Mardia and Marshall (1984), relaxing the assumption of normality. We consider the spatial linear model under the family of elliptical distributions. This class that contains distributions such as the normal one, $\boldsymbol{t}$, power exponential and slash, has received greater interest in the literature (Cambanis et al., 1981; Fang and Anderson, 1990; Fang et al., 1990; Fang and Zhang, 1990; Gupta and Varga, 1993; Kelker, 1970). The family of elliptical distributions offers a more flexible framework for modelling continuous spatial data. This class contains many distributions with heavier tails than the normal one, allowing us to model tails that are frequently observed in multivariate symmetric data sets.

To assess the effect of small perturbations in the model (or data) on the parameter estimates, Cook (1986) has proposed an interesting method, named local influence. This analysis does not involve recomputing the parameter estimates for each case deletion, so it is often computationally simpler. Several authors have extended the local influence method to various regression models. Paula (1993) developed local influence on linear models with restriction in the parameters in the form of linear inequalities. Lesaffre and Verbeke (1998) extended the local influence methodology to normal linear mixed models in repeated-measurement context and under the case-weight perturbation scheme. Influence diagnostics based on the likelihood displacement have been developed for multivariate elliptical linear models by Galea et al. (1997) and by Liu (2000). Galea et al. (2003) presented influence diagnostics in univariate elliptical linear regression models, and mentioned that the lack of correlation among the errors in the multivariate elliptical distribution is not equivalent to independent elliptical errors, as occurs in the normal case. Galea et al. (2005) studied diagnostics in symmetrical nonlinear models. Osorio et al. (2007) derived local influence curvatures under various perturbation schemes for elliptical linear models with longitudinal structure. Ibacache-Pulgar and Paula (2011) applied the approach of local influence in Student-t partial linear models.

So, in this Chapter we extend the Gaussian spatial model relaxing the assumption of normality of the observations, considering the family of elliptical distributions, which provides greater flexibility in modeling tails or extremes that are frequently observed in multivariate symmetric data sets. Apart from this flexibility, it preserves several well-known properties of the normal distribution allowing one to derive attractive explicit solution forms. More specifically, in this paper, we study and develop spatial linear models where the random errors follow an elliptical distribution, generalizing the previous Gaussian spatial linear models considered in the literature. On the other hand, there are only a few works in the literature about influence diagnostics in geostatistical analysis (Christensen et al., 1992a; Diamond and Armstrong, 1984; Warnes, 1986). Recently, Uribe-Opazo et al. (2012) used diagnostic techniques to assess the sensitivity of the maximum likelihood estimators, covariance functions and linear predictor to small perturbations in the data and/or in the Gaussian spatial linear model assumptions and Borssoi et al. (2011) applied local influence of explanatory variables in the same model.

We also discuss maximum likelihood estimation and some diagnostic tools such as local influence and generalized leverage. Moreover, we consider the perturbation in the response variable proposed by Zhu et al. (2007). The Chapter unfolds as follows. Section 2.3 presents the Elliptical spatial linear model. In Section 2.4, the maximum likelihood estimators are obtained, and an explicit expression for the Fisher information matrix is also presented. The likelihood ratio test is briefly discussed. Section 2.5 reviews concepts of local influence and generalized leverage. Furthermore, we discuss the selection of an appropriate perturbation scheme by using the methodology proposed by Zhu et al. (2007). Section 2.6 contains an application, with real data, to illustrate the methodology developed in this paper. A simulation study is also included. Finally, Section 2.7 contains some concluding remarks. Calculations are presented in the appendices.

2.3 The Elliptical Spatial Linear Model

We say that the random vector $\mathbf{Y}$, $n\times 1$ dimensional has an elliptical distribution with location parameter $\boldsymbol{\mu}$ an $n \times 1$ vector and an $n \times n$ scale matrix $\boldsymbol{\Sigma}$, if its density is given by

$$f(\mathbf{Y};\boldsymbol{\mu},\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} g\{(\mathbf{Y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\}, \quad \mathbf{Y} \in \mathbb{R}^n, \tag{2.1}$$

where the function $g : \mathbb{R} \to [0,\infty)$ is such that $\int_0^\infty u^{n-1}g(u^2)du < \infty$. The function g is known as density generator. For a vector $\mathbf{Y}$ distributed according to the density (2.1), we use the notation $\mathbf{Y} \sim El_n(\boldsymbol{\mu},\boldsymbol{\Sigma};g)$ or simply $El_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Kelker (1970) and Cambanis et al. (1981) have discussed many properties of the elliptical distributions. The characteristic function is $exp(i\boldsymbol{t}^\top\boldsymbol{\mu})\varphi(\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t})$ for some function φ, where $i = \sqrt{-1}$. If it exists, $E(\mathbf{Y}) = \boldsymbol{\mu}$ and $\text{Var}(\mathbf{Y}) = c_g\boldsymbol{\Sigma}$, where $c_g = -2\varphi^{(1)}(0)$ is a positive constant. The random vector $\mathbf{Y}$ has the representation $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}U$, where R is a positive random variable, U has the uniform distribution on $\boldsymbol{u}^\top\boldsymbol{u} = 1$, that is $U \stackrel{d}{=} \mathbf{P}U$, for any orthogonal matrix $\mathbf{P}$; R and U are independent and $\mathbf{A}$ is a nonsingular matrix such that $\mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$. The moments of R are related to the characteristic function. For example, $E(R^2) = -2n\varphi^{(1)}(0)$, $E(R^4) = 4n(n+2)\varphi^{(2)}(0)$, $E(R^6) = -8n(n+2)(n+4)\varphi^{(3)}(0)$ and $E(R^8) = 16n(n+2)(n+4)(n+6)\varphi^{(4)}(0)$. If $\mathbf{Y}$ has finite fourth moments each component of $\mathbf{Y}$, z_i, has zero skewness and the same kurtosis, $3\{(\varphi^{(2)}(0)/[\varphi^{(1)}(0)]^2)-1\} = 3\{(E[(z_i-\mu_i)^4]/3[\text{var}(z_i)]^2)-1\} = 3\kappa_g$, for $i = 1,\ldots,n$, where $3\kappa_g$ is called the kurtosis parameter of the corresponding elliptical distribution with density generator g.

In the case where $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$ (identity matrix of dimension n), we obtain the spherical family of densities.

To model a data set with spatial correlation structure (Mardia and Marshall, 1984), we consider an Elliptical stochastic process $\{\mathbf{Y}(\mathbf{s}), \mathbf{s} \in D\}$, where D is a subset of $\mathbb{R}^h$, the $h-$dimensional Euclidean space, and $\mathbf{Y} \sim El_n(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$. It is supposed that data $Y(\mathbf{s}_1),\ldots,Y(\mathbf{s}_n)$ of this process are observed at known sites (locations) $\mathbf{s}_i$, for $i = 1,\ldots,n$, where $\mathbf{s}_i$ is an $h-$dimensional vector of spatial site coordinates, and generated from the model,

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i),$$

where both the deterministic term $\mu(\mathbf{s}_i)$ and the stochastic term $\epsilon(\mathbf{s}_i)$ may depend on the spatial location in which $Y(\mathbf{s}_i)$ is observed. We assume that the stochastic errors have zero mean, $E\{\epsilon(\mathbf{s}_i)\}=0$, and that the variation between spatial points is determined by a covariance function $C(\mathbf{s}_i, \mathbf{s}_j)=\text{cov}\{\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)\}$.

Suppose that for some known functions of $\mathbf{s}_i$, $x_1(\mathbf{s}_i), \ldots, x_p(\mathbf{s}_i)$, the mean of the stochastic process is

$$\mu(\mathbf{s}_i) = \sum_{j=1}^{p} x_j(\mathbf{s}_i)\beta_j,$$

where $\beta_1, \ldots, \beta_p$ are unknown parameters to be estimated. In addition, each family of covariance functions $C(\mathbf{s}_i, \mathbf{s}_j)$, is fully specified by a $q-$dimensional parameter vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$. In our case $q = 3$. We use the following notations: $Y(\mathbf{s}_i) = y_i$, $\mathbf{Y} = (y_1, \ldots, y_n)^\top$, $x_{ij} = x_j(\mathbf{s}_i)$, $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$, $\mathbf{X}$ as the $n \times p$ matrix with ith row $\mathbf{x}_i^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\epsilon_i = \epsilon(s_i)$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$, with $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Thus, $\mu(\mathbf{s}_i) = \mathbf{x}_i^\top\boldsymbol{\beta}$ and then $y_i = \mathbf{x}_i^\top\boldsymbol{\beta} + \epsilon_i$, $i = 1, \ldots, n$. Equivalently, in matrix notation, we have the spatial linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.2}$$

Then, $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and the scale matrix of $\boldsymbol{\epsilon}$ is $\boldsymbol{\Sigma} = [(\sigma_{ij})]$, where σ_{ij} is proportional to $C(\mathbf{s}_i, \mathbf{s}_j)$. We assume that $\boldsymbol{\Sigma}$ is nonsingular and that $\mathbf{X}$ has a full rank. We concentrate on a particular parametric form for the scale matrix given by

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\phi}) = \phi_1\mathbf{I}_n + \phi_2\mathbf{R}, \tag{2.3}$$

where ϕ_1 can be viewed as a measurement error variance or a nugget effect (the magnitude of the apparent discontinuity at the origin), ϕ_2 is defined as sill (the value for distances beyond the range value), $\mathbf{R} = \mathbf{R}(\phi_3) = [(r_{ij})]$, is an $n \times n$ symmetric matrix with diagonal elements $r_{ii} = 1$, $r_{ij} = \phi_2^{-1} C(\mathbf{s}_i, \mathbf{s}_j)$ if $\phi_2 \neq 0$, $r_{ij} = 0$ if $\phi_2 = 0$ for $i \neq j = 1, \ldots, n$ and ϕ_3 is a function of the range of the model. This parametric form occurs for several isotropic processes, where $C(\mathbf{s}_i, \mathbf{s}_j)$ is defined via the function $C(d_{ij}) = \phi_2 r_{ij}$, with $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ being the Euclidean distance between the points $\mathbf{s}_i$ and $\mathbf{s}_j$. For example, the Matérn is a covariance function particulary attractive given by

$$c_g C(d_{ij}) = \begin{cases} \dfrac{c_g \phi_2}{2^{\kappa-1}\Gamma(\kappa)} \left(d_{ij}/\phi_3\right)^{\kappa} K_{\kappa}\left(d_{ij}/\phi_3\right), & d_{ij} > 0, \\ c_g(\phi_1 + \phi_2), & d_{ij} = 0, \end{cases}$$

where once again the parameters are assumed to be non-negative, i.e., $\phi_1 \geq 0$, $\phi_2 \geq 0$ and $\phi_3 \geq 0$; $K_\kappa(u) = \frac{1}{2}\int_0^\infty x^{\kappa-1} e^{-\frac{1}{2}u(x+x^{-1})} dx$ is the modified Bessel function of the third kind of order κ (Gradshteyn and Ryzhik, 2000), where $\kappa > 0$ is fixed, and c_g is a constant that depends on g. The Gaussian covariance function is a special case when $\kappa \to \infty$ and it is given by

$$c_g C(d_{ij}) = \begin{cases} c_g \phi_2 \exp\left[-\left(d_{ij}/\phi_3\right)^2\right], & d_{ij} > 0, \\ c_g(\phi_1 + \phi_2), & d_{ij} = 0. \end{cases}$$

The exponential covariance is also a special case of Matérn family, it corresponds to $\kappa = \frac{1}{2}$ and can be written more simply as

$$c_g C(d_{ij}) = \begin{cases} c_g \phi_2 \exp\left[-\left(d_{ij}/\phi_3\right)\right], & d_{ij} > 0, \\ c_g(\phi_1 + \phi_2), & d_{ij} = 0. \end{cases}$$

Under this specification the variance-covariance matrix of $\mathbf{Y}$ is given by $\mathrm{Var}(\mathbf{Y}) = c_g\{\phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}\}$. For the normal case, $c_g = 1$. Other examples of covariance function that also yield the form (2.3) can be viewed at Diggle and Ribeiro Jr. (2007). Although we work in terms of covariances functions, the variogram, defined by $\gamma(d_{ij}) = c_g\{C(0) - C(d_{ij})\}$, can be used.

2.4 Maximum likelihood estimation (MLE)

Under the hypothesis that in (2.2) the errors have an elliptical distribution, then $\mathbf{Y} \sim El_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}, g)$, has density given by

$$f(\mathbf{Y}, \boldsymbol{\theta}) = |\boldsymbol{\Sigma}|^{-1/2} g\{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\}, \quad \mathbf{Y} \in \mathbb{R}^n, \tag{2.4}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)^\top$. The unknown model parameters, $\boldsymbol{\theta}$, may be estimated by maximizing the corresponding log-likelihood function given by

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| + \log g(\delta), \tag{2.5}$$

where $\delta = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, is known as Mahalanobis distance and $\boldsymbol{\Sigma}$ is defined in (2.3).

Assuming that g is continuous and differentiable, we can define the functions

$$W_g(\delta) = \frac{\partial \log g(\delta)}{\partial \delta} = \frac{g'(\delta)}{g(\delta)} \quad \text{and} \quad W_g'(\delta) = \frac{\partial W_g(\delta)}{\partial \delta},$$

that depend on the distribution of the elliptical contour family assumed. Table 2.1 shows functions $g(\delta)$ and $W_g(\delta)$ for some elliptical distributions, where c is a normalizing constant.

Table 2.1: Functions $g(\delta)$ and $W_g(\delta)$ for some elliptical distributions.

Distribution	$g(\delta)$	$W_g(\delta)$	Parameters
Power Exponential $(\lambda)^*$	$c \exp\left(-\delta^\lambda/2\right)$	$-\frac{1}{2}\lambda\delta^{\lambda-1}$	$\lambda \neq 1/2$
Normal	$c \exp\left(-\delta/2\right)$	$-\frac{1}{2}$	
Generalized $\boldsymbol{t}$ $(\nu, \gamma)^*$	$\gamma^{-n/2}(1 + \delta/\gamma)^{-(n+\nu)/2}$	$-\frac{1}{2}(\nu + n)/(\gamma + \delta)$	$\nu, \gamma > 0$
$\boldsymbol{t}$ $(\nu)^*$	$c(1 + \delta/\nu)^{-(n+\nu)/2}$	$-\frac{1}{2}\left(\frac{\nu+n}{\nu+\delta}\right)$	$\nu > 0$

*λ, ν and γ are parameters of the respective distributions.

Score functions for the spatial linear models with distribution of the elliptical contour family are given by

$$U(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -2W_g(\delta)\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad \text{and} \quad U(\boldsymbol{\phi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}},$$

where $\dfrac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \phi_j} = -\dfrac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\dfrac{\partial \boldsymbol{\Sigma}}{\partial \phi_j}\right) - W_g(\delta)(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}\dfrac{\partial \boldsymbol{\Sigma}}{\partial \phi_j}\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ is the jth element of $U(\boldsymbol{\phi})$, for $j = 1, 2, 3$.

Unfortunately, the equation $U(\boldsymbol{\phi}) = \mathbf{0}$ does not lead to an explicit solution for $\boldsymbol{\phi}$. A common practice is to maximize the concentrated log likelihood obtained as follows. Given $\boldsymbol{\Sigma}$, for any density generating function g, the log-likelihood function (2.5) is maximized at

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Y}. \tag{2.6}$$

By substituting the expression (2.6) into the log-likelihood function, we obtain a concentrated log-likelihood

$$\mathcal{L}_c(\boldsymbol{\phi}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| + \log g(\hat{\delta}),$$

where $\hat{\delta} = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\beta}}$ is given in 2.6. $\mathcal{L}_c(\boldsymbol{\phi})$ must be maximized numerically with respect to ϕ_1, ϕ_2 and ϕ_3.

Given $\hat{\boldsymbol{\phi}}$, the MLE of $\boldsymbol{\phi}$, the MLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}, \tag{2.7}$$

where $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\phi}})$. Asymptotic standard errors can be calculated by inverting either the observed information matrix or the expected information matrix. For the elliptical distribution, the expected information matrix is given by (Lange et al., 1989; Mitchell, 1989)

$$K(\boldsymbol{\theta}) = \begin{pmatrix} K(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & K(\boldsymbol{\phi}) \end{pmatrix},$$

where

$$K(\boldsymbol{\beta}) = \frac{4d_g}{n}\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X}$$

and $K(\boldsymbol{\phi}) = [(k_{ij}(\phi))]$, with

$$k_{ij}(\phi) = \frac{b_{ij}}{4}\left(\frac{4f_g}{n(n+2)} - 1\right) + \frac{2f_g}{n(n+2)}\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_j}\right),$$

$d_g = \text{E}[W_g(U)^2U]$, $f_g = \text{E}[W_g(U)^2U^2]$, where $U = ||\mathbf{Y}||^2$, $\mathbf{Y} \sim El_n(\mathbf{0}, \mathbf{I}_n, g)$ and $b_{ij} = \text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_i}\right)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_j}\right)$, for $i, j = 1, 2, 3$.

For some distributions of the elliptical contour family it is possible to obtain closed expressions for the expected values d_g and f_g. For $\boldsymbol{t}$ distribution, $d_g = \frac{n}{4}\frac{\nu+n}{\nu+n+2}$ and $f_g = \frac{n(n+2)}{4}\frac{\nu+n}{\nu+n+2}$, and for normal distribution $d_g = \frac{n}{4}$ and $f_g = \frac{n(n+2)}{4}$.

The covariance matrix of $\hat{\boldsymbol{\beta}}$, [see (2.7)], can be estimated by,

$$\mathbf{V}_\beta = [(v_{ij})] = (n/4d_g)(\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}.$$

Then, a $100(1-\alpha)\%$ confidence interval for β_j is given by

$$\hat{\beta}_j \pm z_{(1-\alpha/2)}\sqrt{v_{jj}},$$

where $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ percentage point from a standard normal distribution. Similarly we can construct confidence intervals for ϕ_j, for $j = 1, 2, 3$.

Also, for the Elliptical spatial linear model, we can consider linear hypothesis of the form $H : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{a}$, where $\mathbf{C}$ is a full rank matrix of contrasts, of known order $q \times p$, range $q \leq p$ while $\boldsymbol{a}$ is a known vector of order $q \times 1$. To test the hypothesis H we can use the likelihood ratio test, given by

$$\tau_E = \log\left\{\frac{|\tilde{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}|}\right\} + 2\log\left\{\frac{g(\hat{\delta})}{g(\tilde{\delta})}\right\}, \tag{2.8}$$

where, $\hat{\delta} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ and $\tilde{\delta} = (\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$, with $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\Sigma}}$ being the MLE of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, under H. The asymptotic distribution of τ_E is $\chi^2(q)$ under H. For the $\boldsymbol{t}$ spatial linear model the statistics (2.8) takes the form

$$\tau_t = \log\left\{\frac{|\tilde{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}|}\right\} + (n+\nu)\log\left\{\frac{\nu+\tilde{\delta}}{\nu+\hat{\delta}}\right\},$$

and for the normal spatial linear model, $\tau_N = \log(|\tilde{\mathbf{\Sigma}}|/|\hat{\mathbf{\Sigma}}|) + (\tilde{\delta} - \hat{\delta})$.

Finally, an important goal in studies that generate spatial data, is a future measurement prediction at a new location inside the same spatial region known as *kriging*. Let $Z_0 = Z(s_0)$ be a future observation at location $s_0 \in D$. The mean of Z_0 is $x_0^\top \boldsymbol{\beta}$, where $x_0^\top = (x_{01}, \ldots, x_{0p})$ and $x_{0j} = x_j(s_0)$, for $j = 1, \ldots, p$. The minimum mean square error predictor [best linear unbiased (Schabenberger and Gotway, 2005)] is

$$p(s_0, \boldsymbol{\theta}) = x_0^\top \boldsymbol{\beta} + \mathbf{C}_0^\top \mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),$$

where, $\mathbf{C}_0^\top = (C(d_{10}), \ldots, C(d_{n0}))$, with $d_{i0} = \|s_i - s_0\|$, for $i = 1, \ldots, n$. So, a point estimator of Z_0 is

$$\hat{Z}_0 = p(s_0, \hat{\boldsymbol{\theta}}) = x_0^\top \hat{\boldsymbol{\beta}} + \hat{\mathbf{C}}_0^\top \hat{\mathbf{\Sigma}}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

An alternative expression for $\hat{Z}_0$ using robust estimators of $\mathbf{C}_0$, $\boldsymbol{\beta}$ and $\mathbf{\Sigma}$ is given in Atkinson and Riani (2004) and Barnett (2004).

2.5 Influence diagnostics

An important step in spatial data analysis is the examination of possible deviations of the assumptions of the statistical model, as well as the detection of atypical observations which have a disproportionate influence on the results of statistical analysis. There are various techniques to assess the influence of perturbations in a data set and in the model assumptions.

There are some papers in the literature on diagnostic influences on spatial linear models. Diamond and Armstrong (1984) and Warnes (1986) observed the sensitivity of predictions to perturbations in the covariance function. Christensen et al. (1992a) discussed case deletion diagnostics for detecting observations that are influential for prediction based on universal kriging, while Christensen et al. (1993) considered diagnostic with the deletion of points to estimate the parameters of the covariance function by the method of restricted maximum likelihood. More recently, Cerioli and Riani (1999) and Militino et al.

(2006) showed that case deletion diagnostics do suffer from masking and suggest robust procedures based on subsets of data free from outliers. In this direction, Filzmoser et al. (2014) proposed the Mahalanobis distance to identify multivariate outliers. Borssoi et al. (2011) and Uribe-Opazo et al. (2012) discussed diagnostic techniques, using local influence methodology, to evaluate the sensitivity of MLE, the covariance functions and the linear predictor under small perturbations in the data and/or spatial linear model with normal distribution.

2.5.1 Local influence

The local influence method suggested by Cook (1986) evaluates the simultaneous effect of observations on the ML estimator without removing it from the data set.

Let $\boldsymbol{\omega}$ be a vector of perturbation $r \times 1 \in \boldsymbol{\Omega}$ subset of $\mathbb{R}^r$, and the perturbed statistical model $\mathcal{M} = \{f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$, where $f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega})$ is the density function of $\mathbf{Y}$ (2.4) perturbed by $\boldsymbol{\omega}$ and $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega}) = \log f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega})$ the correspondent log-likelihood function. Denoting the vector of no perturbation (the null vector) by $\boldsymbol{\omega}_0$, we suppose that $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega}_0) = \mathcal{L}(\boldsymbol{\theta})$ given in (2.5).

The influence of the perturbation $\boldsymbol{\omega}$ on the ML estimator can be evaluated by the likelihood displacement given by $LD(\boldsymbol{\omega}) = 2\{\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_\omega)\}$, where $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ in the postulated model, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)^\top$ and $\hat{\boldsymbol{\theta}}_\omega$ is the ML estimator of $\boldsymbol{\theta}$ in the perturbed model by $\boldsymbol{\omega}$. Cook (1986) proposed to study the local behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$ and shows that the normal curvature C_l of $LD(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in direction of some unit vector $\boldsymbol{l}$, is given by $C_l = C_l(\boldsymbol{\theta}) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|$, with $||\boldsymbol{l}|| = 1$, where $\mathbf{L}$ is the observed information matrix (given in Appendix A), evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Delta}$ is a $(p+3) \times r$ matrix given by $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\phi^\top)^\top$, where $\boldsymbol{\Delta}_\beta = \partial^2 \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top$ and $\boldsymbol{\Delta}_\phi = \partial^2 \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\omega}^\top$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

The plot of the elements $|\boldsymbol{l}_{max}|$ versus i (order of data) can reveal what type of perturbation has more influence on $LD(\boldsymbol{\omega})$, in the neighborhood of $\boldsymbol{\omega}_0$, (Cook, 1986). Even considering $C_i = 2|f_{ii}|$, where f_{ii} form the main diagonal of the matrix $\mathbf{F} = \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta}$, can be used the index plot of C_i to evaluate the presence of influential observations.

Since C_l is not invariant under uniform change of scale, Poon and Poon (1999) proposed the conformal normal curvature $B_l = C_l/tr(2\mathbf{F})$, (Zhu and Lee, 2001). An interesting property of conformal curvature is that for any direction unit $\boldsymbol{l}$, it follows that $0 \leq B_l \leq 1$. This allows, for example, comparison of curvatures among different elliptical models. We denote by $B_i = 2|f_{ii}|/tr(2\mathbf{F})$ the conformal curvature in the unit direction with $i-$th entry 1 and all other entries 0. According to Zhu and Lee (2001), the ith observation is potentially influential if $B_i > \bar{B} + 2\text{sd}(B)$, where $\bar{B} = \sum_{i=1}^{n} B_i/n$ and $\text{sd}(B)$ is the standard deviation of $B_1, \ldots, B_n$.

In this paper we consider as perturbation scheme the model shift in mean, i.e. $\mathbf{Y} = \boldsymbol{\mu}(\boldsymbol{\omega}) + \boldsymbol{\epsilon}$, with $\boldsymbol{\mu}(\boldsymbol{\omega}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\omega}$ where $\mathbf{A}$, is an $n \times n$ matrix that does not depend on $\boldsymbol{\beta}$ or on $\boldsymbol{\omega}$. In this case $\boldsymbol{\omega}_0 = \mathbf{0}$. Equivalently we can write $\mathbf{Y}_\omega = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\mathbf{Y}_\omega = \mathbf{Y} + (-1)\mathbf{A}\boldsymbol{\omega}$, that corresponds to a perturbation scheme of the response vector. The $\boldsymbol{\Delta}_\beta$ and $\boldsymbol{\Delta}_\phi$ matrices for this perturbation scheme are presented in the Appendix A.2. In Appendix A.3, using the proposal of Zhu et al. (2007), we show that $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$ produces an adequate perturbation scheme.

2.5.2 Generalized Leverage

The general concept of a generalized leverage is related to a certain value observed y_i, over the corresponding adjusted value $\hat{y}_i$, see for example (Hoaglin and Welsh, 1978; Ross, 1987; St. Laurent and Cook, 1992).

Based on the generalized leverage by Wei et al. (1998), which is defined as $GL(\hat{\boldsymbol{\theta}}) = \partial\hat{\mathbf{Y}}/\partial\mathbf{Y}^\top = [(\partial\hat{y}_i/\partial y_j)]$ where $\hat{\mathbf{Y}} = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}$. The element (i, j) of $GL(\hat{\boldsymbol{\theta}})$, is the instantaneous rate of change in ith predicted values with respect to the jth value of the response.

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$, assuming it exists and it is unique, and assuming that the log-likelihood function has continuous second derivatives with respect to $\boldsymbol{\theta}$ and $\mathbf{Y}$, and using results from Wei et al. (1998) it can be shown that the generalized leverage may be expressed as $GL(\boldsymbol{\theta}) = GL_1 + GL_2$, where

$$
\begin{aligned}
GL_1 &= \mathbf{X}(L_{\beta\beta} - L_{\beta\phi}L_{\phi\phi}^{-1}L_{\phi\beta})^{-1}(-L_{\beta Y}) \quad \text{and} \\
GL_2 &= \mathbf{X}(L_{\beta\beta} - L_{\beta\phi}L_{\phi\phi}^{-1}L_{\phi\beta})^{-1}(L_{\beta\phi}L_{\phi\phi}^{-1}L_{\phi Y}),
\end{aligned}
$$

with $L_{\beta Y} = -2\mathbf{X}^\top\mathbf{\Sigma}^{-1}\{W_g(\delta)\mathbf{\Sigma} + 2W_g'(\delta)\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\}\mathbf{\Sigma}^{-1}$ and $L_{\phi Y} = \dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\phi}\partial\mathbf{Y}^\top}$, with elements $\dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\phi_j\partial\mathbf{Y}^\top} = -2\boldsymbol{\epsilon}^\top\mathbf{\Sigma}^{-1}\dfrac{\partial\mathbf{\Sigma}}{\partial\phi_j}\mathbf{\Sigma}^{-1}\{W_g'(\delta)\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top + W_g(\delta)\mathbf{\Sigma}\}\mathbf{\Sigma}^{-1}$ for $j = 1, 2, 3$.

Note that if $L_{\beta\phi} \approx \mathbf{0}$ then $GL(\boldsymbol{\theta}) \approx \mathbf{X}L_{\beta\beta}^{-1}(-L_{\beta Y})$. Furthermore, if we use the Fisher information matrix $K(\boldsymbol{\theta})$ instead of $-\mathbf{L}$, we have $GL(\boldsymbol{\theta}) \approx (n/4d_g)\mathbf{X}(\mathbf{X}^\top\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}L_{\beta Y}$. For the normal case, $d_g = n/4$, $L_{\beta Y} = \mathbf{X}^\top\mathbf{\Sigma}^{-1}$ and then $GL(\boldsymbol{\theta}) \approx \mathbf{X}(\mathbf{X}^\top\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Sigma}^{-1}$, that coincides with the generalized leverage matrix proposed by Martin (1992).

The diagonal elements of the matrix $GL(\hat{\boldsymbol{\theta}})$, i.e., the elements GL_{ii} for $i = 1, \ldots, n$, are used as diagnostics tool of the influence in the vector $\hat{\mathbf{Y}}$. The ith response is potentially influential if $GL_{ii} > \bar{GL} + 2\text{sd}(GL)$, where $\bar{GL} = \sum_{i=1}^n GL_{ii}/n$ and $\text{sd}(GL)$ is the standard deviation of $GL_{11}, \ldots, GL_{nn}$.

2.6 Application

In this Section the methodology developed in this paper is illustrated using observations from 93 wells in a single aquifer near Saratoga Valley, Wyoming. The water heights, y, are in meters above mean sea level, and the size of the aquifer is about 1300 square kilometers. The x_1 and x_2 coordinates are in kilometers. The first goal is to obtain a predicted surface map for the response. The data set appears in Jones (1989). As in Christensen et al. (1992a), for the normal spatial linear models, we use the Gaussian covariance function which is given by

$$C(d_{ij}) = \begin{cases} \phi_2 \exp\left[-\left(d_{ij}/\phi_3\right)^2\right], & d_{ij} > 0, \\ \phi_1 + \phi_2, & d_{ij} = 0, \end{cases}$$

with ϕ_1, ϕ_2 and ϕ_3 non-negative. We use the model $y = \mu + \epsilon$, where $\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

To illustrate the methodology, we use the $\boldsymbol{t}$ distribution. This basically for two reasons. First, within the class of elliptical distributions undoubtedly the $\boldsymbol{t}$ distribution is one of the most used, with applications in literature of several areas, see for instance (Lange and Sinsheimer, 1993; Lange et al., 1989). Second, in this case the degrees of freedom for the $\boldsymbol{t}$ spatial linear model cannot be selected using the likelihood function, as proposed by

Lange et al. (1989), and we think that is interesting to discuss this point. In effect, as noted by Zellner (1976), for the case of the usual $\boldsymbol{t}-$linear regression model, (see Appendix A.5), the likelihood function of the $\boldsymbol{t}$ model is an increasing function of ν, and then can not be used to select the degree of freedom of the $\boldsymbol{t}$ distribution. In this paper we used, initially, cross-validation to select the degree of freedom of the $\boldsymbol{t}$ spatial linear model. In effect, the cross-validation is defined by

$$CV = CV(\nu) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{y(\mathbf{s}_i) - \hat{y}_{(i)}(\mathbf{s}_i)}{1-h_{ii}}\Big\}^2,$$

where $\hat{y}_{(i)}(\mathbf{s}_i) = \mathbf{x}_i^\top\hat{\boldsymbol{\beta}}_{(i)}$, with $\mathbf{x}_i^\top$ the ith row of the matrix $\mathbf{X}$, is the prediction in the location $\mathbf{s}_i$ without considering this observation, $(y_i, \mathbf{x}_i^\top)$, $\hat{\boldsymbol{\beta}}_{(i)}$ is the ML estimator of $\boldsymbol{\beta}$ without considering the ith observation, and h_{ii} is the ith diagonal element of the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}$, for $i = 1,\ldots,n$. To simplify the calculations we can use the approximation $\hat{\boldsymbol{\beta}}_{(i)} \approx \hat{\boldsymbol{\beta}} + K^{-1}(\hat{\boldsymbol{\beta}})U_{(i)}(\hat{\boldsymbol{\beta}})$, where

$$K(\hat{\boldsymbol{\beta}}) = \Big(\frac{\nu+n}{\nu+n+2}\Big)\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}$$

and

$$U_{(i)}(\hat{\boldsymbol{\beta}}) = \Big(\frac{\nu+n-1}{\nu+\delta_{(i)}}\Big)\mathbf{X}_{(i)}^\top\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}), \tag{2.9}$$

where $\delta_{(i)} = (\mathbf{Y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}})^\top\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}})$, with $\mathbf{X}_{(i)}$, $(n-1)\times p$, the matrix $\mathbf{X}$ without the ith row, $\mathbf{x}_i^\top$, and $\mathbf{Y}_{(i)}$, $(n-1)\times 1$ denotes the vector $\mathbf{Y}$ without the response y_i, for $i = 1,\ldots,n$. Note that in (2.9) the matrix $\hat{\boldsymbol{\Sigma}}^{-1}$ is of order $(n-1)\times(n-1)$.

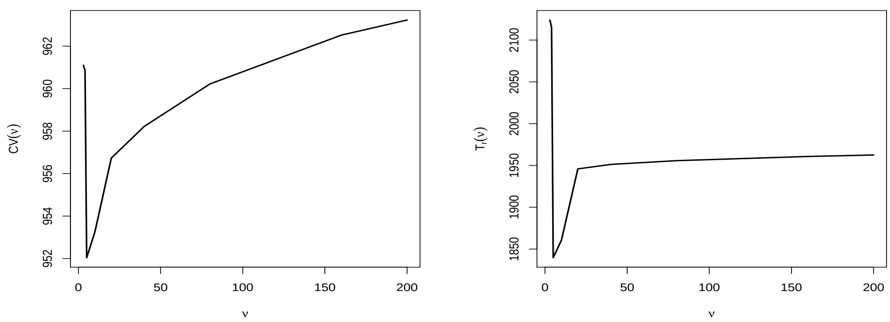
Alternatively, Kano et al. (1993) within the class of elliptical distributions, proposed to use the trace of the asymptotic covariance matrix of an estimated mean as a criterion in selecting a better model. Let $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. From the results of Section 3, it follows that the trace of the asymptotic covariance matrix of $\hat{\boldsymbol{\mu}}$ is given by,

$$T_r = T_r(\nu) = \Big\{\frac{\nu+n+2}{\nu+n}\Big\}tr\{(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\}.$$

Table 2.2: Criteria CV and T_r for some values of ν for Jones' data.

ν	CV	T_r
3	961.102	2123.977
4	960.857	2114.881
5	**952.036**	**1839.644**
10	953.252	1860.548
20	956.735	1945.884
40	958.222	1951.239
80	960.224	1955.717
160	962.515	1960.754
200	963.234	1962.506
∞	1220.443	1975.651

Figure 2.1: Plots of $CV(\nu)$ (a) and $T_r(\nu)$ (b) versus ν for the Jones' data.



(a) (b)

Source: From the author.

Table 2.2 shows the CV and T_r criteria for some values of ν and Fig. 2.1 shows the graphs corresponding for the Jones' data. We can see that both criteria suggest that a suitable value for ν is 5. According to these criteria the $\boldsymbol{t}_5$ spatial linear model presents a better fit. Thus, we use the model to illustrate the methodology developed in this paper. For comparative purposes we also fit the normal spatial linear model. Table 2.3 presents the MLEs for the parameters under the normal and $\boldsymbol{t}_5$ spatial linear models. The standard errors were estimated using the expected information matrix.

Table 2.3 shows the marked differences between the values of the estimators, especially $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\phi}_3$ between the two models. Also the standard errors under the $\boldsymbol{t}_5$ model are smaller when compared to the ones obtained under the normal model, except for $\hat{\phi}_1$ and $\hat{\phi}_2$. One hypothesis of interest in this case is $H : \beta_1 = \beta_2 = 0$. Using the results of Section 3,

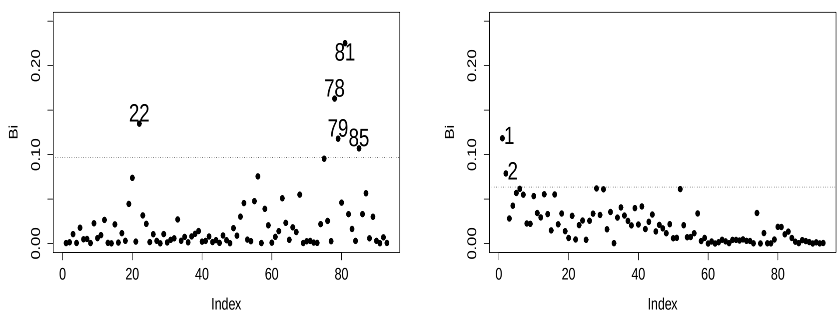
Table 2.3: MLEs and asymptotic standard errors (in parentheses) estimates under normal and $\boldsymbol{t}_5$ distributions for the Jones' data.

Distribution	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
normal	2248.2280	-0.6881	-3.0336	25.7923	950.4528	7.7435
	(30.9242)	(1.2133)	(0.4974)	(5.3874)	(320.2882)	(0.7143)
$\boldsymbol{t}_5$	2222.5280	-0.3403	-2.9237	7.4101	651.1403	0.0020
	(15.0299)	(0.6758)	(0.2186)	(7.9408)	(432.2509)	(0.0002)

we have the values of the likelihood ratio test (p-value) are given by $\tau_N = 19.068$ (< 0.001) and $\tau_t = 130.568$ (< 0.0001), for the normal and $\boldsymbol{t}_5$ spatial models, respectively. In both models, the coordinates (x_1, x_2) are significant.

Fig. 2.2 shows two index plots of B_i for the normal and $\boldsymbol{t}_5$ models. As expected the MLEs are more sensitive under the normal model. We can see that, although we highlight some points, possible outliers are not detected, at least clearly, under the $\boldsymbol{t}_5$ spatial linear model.

Figure 2.2: Index plots of B_i for the normal (a) and $\boldsymbol{t}_5$ (b) for perturbation scheme the model shift in mean for Jones' data.

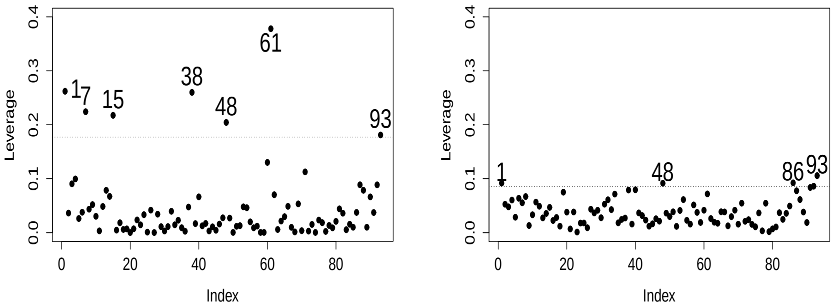


(a) Normal (b) $\boldsymbol{t}_5$

Source: From the author.

Fig. 2.3 shows the leverage plots. We note that the predicted values $\hat{y}_i$ are more stable to changes in y_i, for $i = 1, \ldots, n$, under the $\boldsymbol{t}_5$ spatial linear model. Table 2.4 presents the MLEs for the parameters under the normal and $\boldsymbol{t}_5$ spatial linear models for the Jones' data without observation #1. Interestingly, this observation only produces some changes in $\hat{\phi}_2$ and in the respective standard error, in both spatial models, with less effect under the $\boldsymbol{t}_5$ model. The analysis without #1 incrases the value of $\hat{\phi}_2$, which leads to a model that consider even more the spatial correlation between the observations.

Figure 2.3: Generalized leverage plots for the normal (a) and $\boldsymbol{t}_5$ (b) models for Jones' data.



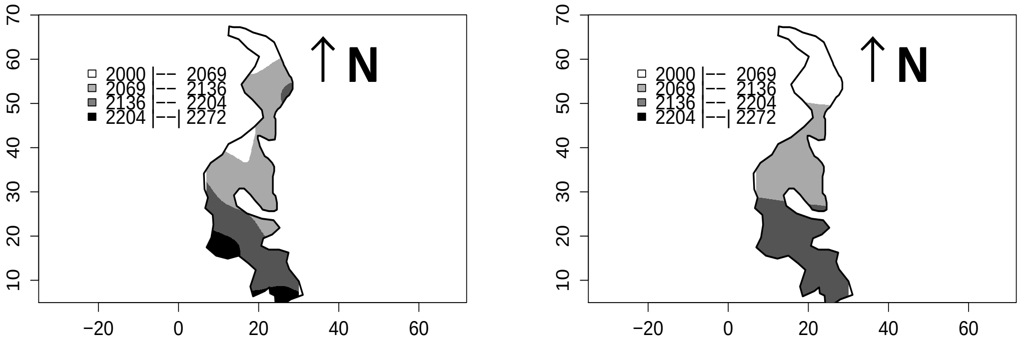
(a) Normal (b) $\boldsymbol{t}_5$

Source: From the author.

Table 2.4: Case deletion: MLEs and standard errors (in parentheses) estimates under normal and $\boldsymbol{t}_5$ spatial linear models, fitted to Jones' data without observation #1.

Distribution	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
normal	2247.1563	-0.2185	-3.3351	24.9241	1053.0808	8.0731
	(32.7975)	(1.2980)	(0.5488)	(5.1528)	(365.2411)	(0.7401)
$\boldsymbol{t}_5$	2222.6580	-0.3574	-2.9144	7.9756	710.1933	0.0020
	(15.6539)	(0.7039)	(0.2276)	(8.6383)	(471.3820)	(0.0002)

Figure 2.4: Thematic maps for the normal (a) and $\boldsymbol{t}_5$ (b) models for Jones' data.



(a) Normal (b) $\boldsymbol{t}_5$

Source: From the author.

In the Fig. 2.4 thematic maps of the spatial variability built by universal kriging are presented, for normal and $\boldsymbol{t}_5$ spatial linear models. Both maps show an increasing trend of the response variable on the north-south direction, which is consistent with the hypothesis test, where the coordinates (x, y) are significant.

Finally we conducted a small simulation study. To analyze the performance of the criterion CV and T_r, based on the above example, we generate 500 data sets considering $n = 93$ for each case, of the $\boldsymbol{t}_5$ spatial linear models, with Gaussian covariance function. The parameters were $\beta_0 = 2126$, $\beta_1 = -0.3400$, $\beta_2 = -2.9200$, $\phi_1 = 7$, $\phi_2 = 650$, $\phi_3 = 0.002$. For each data set we estimate the parameters and calculate CV and T_r for $\nu = 5$ and for $\nu = \infty$, normal model. For the 500 simulated data set we count how many times were chosen the $\boldsymbol{t}_5$ and normal models. The results are presented in Table 2.5, where we can see the effectiveness of CV in this case. The T_r criterion presents, in this situation, a 67% of effectiveness.

Table 2.5: N^0 of times was chosen the normal and $\boldsymbol{t}_5$ models using the criteria CV and T_r in 500 simulated data set. In parentheses percentage.

Model	CV	T_r
normal spatial	1 (0.2)	163 (32.6)
$\boldsymbol{t}_5$ spatial	499 (99.8)	337 (67.4)

Of course it is necessary to make a deeper simulation study, considering different scenarios, to evaluate the performance of the criteria CV and T_r in the selection of ν in a $\boldsymbol{t}$ spatial linear model.

2.7 Conclusion

It is well known that the normal distribution is not always suitable for modelling multivariate continuous data. In this paper we extended the Gaussian spatial linear model relaxing the assumption of normality of observations. We considered the spatial linear model under the family of elliptical distributions, which offers a more flexible framework for modeling tails or extremes that are frequently observed in multivariate symmetric data sets. Moreover, it preserves several well-known properties of the normal distribution allowing one to derive attractive explicit solution forms.

In the Elliptical spatial linear model we discussed maximum likelihood estimation and some diagnostic tools such as local influence and generalized leverage. Explicit expressions for the Fisher information matrix and for the Delta matrix were presented. The likelihood ratio test for linear hypothesis, was also discussed.

To illustrate the methodology developed in the paper, we used the $\boldsymbol{t}$ spatial linear model. Because the likelihood function of the $\boldsymbol{t}$ spatial linear model is an increasing function of the degree of freedom, we used cross-validation to select the degree of freedom. The trace of the asymptotic covariance matrix of the ML estimator of the vector mean also was used to fixed the degree of freedom. A small simulation study, showed that cross-validation performed better than the trace of the asymptotic covariance matrix. However, it is necessary to make a deeper simulation study, considering different scenarios, to evaluate the performance of these criterion in the selection of the degree of freedom in a $\boldsymbol{t}$ spatial linear model. This topic will be addressed in a subsequent study.

Appendix A

A.1 The observed information matrix for elliptical spatial linear models

The log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| + \log g(\delta),$$

where $\delta = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

The second derivatives matrix, is given by

$$\mathbf{L}(\boldsymbol{\theta}) = \mathbf{L} = \begin{pmatrix} L_{\beta\beta} & L_{\beta\phi} \\ L_{\phi\beta} & L_{\phi\phi} \end{pmatrix},$$

where $L_{\beta\beta} = \dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = 2\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\{W_g(\delta)\boldsymbol{\Sigma} + 2W_g'(\delta)\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\}\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $L_{\beta\phi} = \dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\phi}^\top}$, with $\dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\phi_j} = 2\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\Big\{W_g'(\delta)\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\boldsymbol{\Sigma}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\phi_j} + W_g(\delta)\dfrac{\partial\boldsymbol{\Sigma}}{\partial\phi_j}\Big\}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}$, as its jth column, for $j = 1, 2, 3$; $L_{\phi\beta} = L_{\beta\phi}^\top$ and $L_{\phi\phi} = \dfrac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top}$, with elements

$$\begin{aligned}
\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\phi_i\partial\phi_j} &= \frac{1}{2}\,\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_j} - \frac{\partial^2\boldsymbol{\Sigma}}{\partial\phi_i\partial\phi_j}\right)\right\} \\
&+ \boldsymbol{\epsilon}^\top\boldsymbol{\Sigma}^{-1}\Big\{W_g'(\delta)\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_j}\right) \\
&+ W_g(\delta)\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_j} - \frac{\partial^2\boldsymbol{\Sigma}}{\partial\phi_i\partial\phi_j} + \frac{\partial\boldsymbol{\Sigma}}{\partial\phi_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\phi_i}\right)\Big\}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon},
\end{aligned}$$

for $i, j = 1, 2, 3$. The derivatives of first and second-order of the scale matrix $\mathbf{\Sigma}$, (2.3), with respect to ϕ_j, for $j = 1, 2, 3$; for some covariance functions are presented in Uribe-Opazo et al. (2012).

A.2 $\mathbf{\Delta}$ matrix for perturbation scheme of the mean for elliptical spatial linear models

In this case we have that $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega})$ is given by

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega}) = -\frac{1}{2}\log|\mathbf{\Sigma}| + \log g(\delta_\omega), \tag{A.1}$$

where $\delta_\omega = \{\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})\}^\top \mathbf{\Sigma}^{-1}\{\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})\} = \boldsymbol{\epsilon}_\omega^\top \mathbf{\Sigma}^{-1}\boldsymbol{\epsilon}_\omega$, $\boldsymbol{\epsilon}_\omega = \mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})$ and $\boldsymbol{\mu}(\boldsymbol{\omega}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\omega}$. Then

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} = -2W_g(\delta_\omega)\{\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})\}^\top \mathbf{\Sigma}^{-1}\mathbf{A}. \tag{A.2}$$

Differentiating (A.2) with respect to $\boldsymbol{\beta}$, see (Nel, 1980),

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\omega}^\top} = 2\mathbf{X}^\top \mathbf{\Sigma}^{-1}\{W_g(\delta_\omega)\mathbf{\Sigma} + 2W_g^{'}(\delta_\omega)\boldsymbol{\epsilon}_\omega \boldsymbol{\epsilon}_\omega^\top\}\mathbf{\Sigma}^{-1}\mathbf{A}. \tag{A.3}$$

The derivative with respect to ϕ_j is given by,

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \phi_j \partial \boldsymbol{\omega}^\top} = -2\boldsymbol{\epsilon}_\omega^\top \big\{W_g(\delta_\omega)(\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{A}}{\partial \phi_j} - \mathbf{D}_j\mathbf{A}) - W_g^{'}(\delta_\omega)\boldsymbol{\epsilon}_\omega^\top \mathbf{D}_j \boldsymbol{\epsilon}_\omega \mathbf{\Sigma}^{-1}\mathbf{A}\big\}, \tag{A.4}$$

for $j = 1, 2, 3$ and $\mathbf{D}_j = \mathbf{\Sigma}^{-1}\partial\mathbf{\Sigma}/\partial\phi_j \mathbf{\Sigma}^{-1}$. Evaluating (A.3) and (A.4) at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ we obtain the $\mathbf{\Delta} = (\mathbf{\Delta}_\beta^\top, \mathbf{\Delta}_\phi^\top)^\top$ matrix.

A.3 The Fisher information matrix $\mathbf{G}(\boldsymbol{\omega})$ for elliptical spatial linear models

To select an adequate matrix $\mathbf{A}$ we can use the methodology proposed by Zhu et al. (2007). In effect, the score function for $\boldsymbol{\omega}$ in the perturbed log-likelihood function (A.1) is given by

$$U(\boldsymbol{\omega}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = -2W_g(\delta_\omega)\mathbf{A}^\top \boldsymbol{\Sigma}^{-1}\{\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})\}.$$

Following Zhu et al. (2007) let $\mathbf{G}(\boldsymbol{\omega})$, the Fisher information matrix with respect to the perturbation vector $\boldsymbol{\omega}$. That is, $\mathbf{G}(\boldsymbol{\omega}) = E_\omega\{U(\boldsymbol{\omega})U^\top(\boldsymbol{\omega})\}$, where E_ω denotes the expectation with respect to $f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega})$. A perturbation $\boldsymbol{\omega}$ is appropriate if it satisfies $\mathbf{G}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$, where $c > 0$. In our case we have

$$\mathbf{G}(\boldsymbol{\omega}) = c(\boldsymbol{\omega})\mathbf{A}^\top \boldsymbol{\Sigma}^{-1}\mathbf{A}.$$

That is, $\mathbf{G}(\boldsymbol{\omega}_0) = c\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}$ with $c = c(\boldsymbol{\omega}_0)$ a positive constant, see Appendix C. Notice that usually $\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A} \neq \mathbf{I}_n$. However if $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$, then $\mathbf{G}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$ and so $\boldsymbol{\mu}(\boldsymbol{\omega}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\omega}$ is a perturbation scheme appropriate. The derivatives $\partial\boldsymbol{\Sigma}^{1/2}/\partial\phi_j$ for $j = 1, 2, 3$, are given in Appendix A.4.

A.4 Derivative of the square root $\boldsymbol{\Sigma}^{1/2}$

Corresponding to any matrix $\boldsymbol{\Sigma}$ $n \times n$ symmetric and nonnegative definite, there is a matrix symmetric nonnegative definite $\boldsymbol{\Sigma}^{1/2} = \mathbf{W}$, such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \mathbf{W}^2$. Furthermore, $\mathbf{W}$ is unique and can be expressed by

$$\mathbf{W} = \mathbf{P}\mathbf{A}^{1/2}\mathbf{P}^\top,$$

where $\mathbf{A}^{1/2} = \text{diag}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n})$, with $\alpha_1, \ldots, \alpha_n$ the eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{P}$ is a matrix $n \times n$ orthogonal ($\mathbf{P}\mathbf{P}^\top = \mathbf{I}_n$) such that $\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top = \mathbf{A}$, with $\mathbf{A} = \text{diag}(\alpha_1, \ldots, \alpha_n)$. So, derivatives of $\boldsymbol{\Sigma}$ with respect to ϕ_j is given by

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_j} = \mathbf{W}\frac{\partial \mathbf{W}}{\partial \phi_j} + \frac{\partial \mathbf{W}}{\partial \phi_j}\mathbf{W}, \quad \text{for} \quad j = 1, 2, 3. \tag{A.5}$$

This equation can be written as $\dot{\boldsymbol{\Sigma}}_j = \mathbf{W}\dot{\mathbf{W}}_j + \dot{\mathbf{W}}_j\mathbf{W}$, where $\dot{\boldsymbol{\Sigma}}_j = \dfrac{\partial \boldsymbol{\Sigma}}{\partial \phi_j}$ and $\dfrac{\partial \mathbf{W}}{\partial \phi_j} = \dot{\mathbf{W}}_j$, which has been extensively studied in the literature, see for instance (Jameson, 1968).

Note that $\dot{\boldsymbol{\Sigma}}_j$, $\mathbf{W}$ and $\dot{\mathbf{W}}_j$ are symmetric matrices. Let $\mathbf{J}_j = \mathbf{P}^\top \dot{\boldsymbol{\Sigma}}_j \mathbf{P}$ and $\mathbf{Q} = [(q_{rs})]$ symmetric matrices $n \times n$, with $q_{rs} = (\sqrt{\alpha_r} + \sqrt{\alpha_s})^{-1}$, for $r, s = 1, \ldots, n$. Then, the solution to equation (A.5) is given by

$$\frac{\partial \mathbf{W}}{\partial \phi_j} = \frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \phi_j} = \mathbf{P}(\mathbf{J}_j \odot \mathbf{Q})\mathbf{P}^\top,$$

where $\odot$ denotes the Hadamard product for $j = 1, 2, 3$.

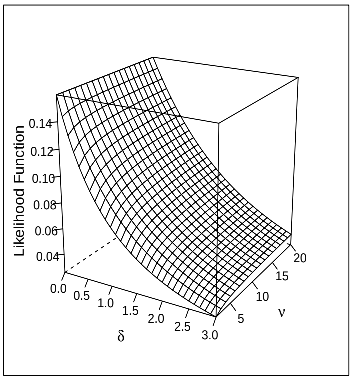
A.5 The likelihood function of the *t* model is an increasing function of ν

As noted by Zellner (1976), for the case of the usual linear regression model, “the necessary conditions on $\boldsymbol{\beta}$, $\boldsymbol{\Sigma} = \phi_1 \mathbf{I}$ and ν for a maximum of the likelihood function cannot be satisfied for $\nu \geq 1$”. In our case, the likelihood function is an increasing function of ν. For illustration, we consider the bivariate case, $\boldsymbol{t}_2(0, \mathbf{I}, \nu)$ with density function given by,

$$f(\mathbf{Y}, \nu, \delta) = \frac{\Gamma((\nu+2)/2)}{\Gamma(\nu/2)(\nu\pi)} \{1 + \delta/\nu\}^{-(\nu+2)/2},$$

with $\delta = \mathbf{Y}^\top \mathbf{Y}$. Clearly, from Fig. A.1, the likelihood function is an increasing function of ν and also of δ.

Figure A.1: Likelihood function versus ν and δ.



Source: From the author.

Chapter 3

Inference on Gaussian spatial linear models with repetitions

3.1 Resumo

Este capítulo apresenta os testes da razão de verossimilhanças usual e corrigido (via Bartlett) para os parâmetros de locação nos modelos espaciais lineares com repetições, e sua respectiva versão corrigida. Apresenta-se o enfoque de inferência para o parâmetro de suavização da família Matérn de modelos. Os estimadores de máxima verossimilhança são obtidos, e uma forma explícita da matriz de informação de Fisher é apresentada. Simulações de Monte Carlo e aplicações a dados reais são apresentadas para ilustrar a metodologia.

3.2 Introduction

Geostatistical data are data collected at known locations in space, from a process that has a value at every location in a certain (1, 2 or 3-D) domain. These data are modelled as the sum of a constant (or varying trend) and a spatially correlated residual. The geostatistics began in South Africa and reached maturity at the Ecoles des Mines at Fontainebleau near Paris. One of the earliest papers in this field developed the basis equations for optimal linear interpolation in a spatially correlated field, by Krige (1951). Then Matheron (1962, 1963, 1965, 1970) published books and used the term *krigeage* (in French), *Krigagem* (in Portuguese) or kriging (in English), which is the geostatistical prediction, finding the best linear unbiased prediction (the expected value) with its prediction error for a variable at a location, given observations and a model for their spatial variation. Given a model for the trend, and under some stationarity assumptions, geostatistical modelling involves the estimation of the spatial correlation.

The observations that can be observed may not be taken in regular grid. The spatial variability can be studied by modeling the parametric form of the covariance matrix. Common models assume a certain shape for the local process, i.e., they assume predetermined behavior. Thus, Stein (1999) and Minasny and McBratney (2005) promoted the use of the Matérn class of models (Matérn, 1960) which has great flexibility for modeling the spatial dependence structure, and it can model many local spatial process. This class has also been presented in Handcock and Wallis (1994), Diggle (2003) and Diggle and Ribeiro Jr. (2007). Gneiting et al. (2010) introduced a flexible parametric family of matrix-valued covariance functions for multivariate spatial random fields, where each constituent component is a Matérn process. Ripley (1987), Mardia and Watkins (1989), Diggle et al. (1998) and Zhang (2002) reported difficulties in likelihood estimation of covariogram parameters. Zhang (2004) has used properties of equivalence of probability measures to show that not all parameters in a spatial generalized linear mixed model are consistently estimated, but one quantity can be estimated consistently by maximum likelihood (ML) methods under asymptotics fixed-domain. Kaufman and Shaby (2013) show that asymptotic results for a Gaussian process over a fixed domain with Matérn covariance function, previously proven only in the case of a fixed range parameter, can be extended to the case of jointly estimating the range and the variance process.

Repeated measures of data have been widely analyzed in many fields, such as biology and medicine, but we can find only a few works which study repeated measures in the geostatistics field. In this case, the observations are taken from different experimental units, which is different geographical locations, where each variable is observed more than once. Considering independent realizations of the process, it is easily to present hypothesis test to verify, for instance, the explanatory variables that should be or should not be part of the model.

Often the number of observations is small, so it is important to use inference strategies that incorporate small sample corrections. The likelihood ratio test, quite often displays distortions when the sample size is small, because its null distribution is poorly approximated by the limiting χ^2 distribution, from which critical values are obtained. A strategy to improve the approximation of the test statistics approximations by the chi-squared distribution, is to alter these statistic by a correction factor. For the likelihood ratio statistic (LR), Bartlett (1937) proposed a correction factor to be multiplied by the statistic in such a way that your corrected version (LR*) presents null distribution χ^2_q with an error of order n^{-2}. More details about Bartlett correction, see (Cribari-Neto and Cordeiro, 1996) and Cordeiro and Cribari-Neto (2014).

The Bartlett correction factor used to modify the likelihood ratio test statistic bring its null distribution closer to its limiting counterpart. When the investigated model involves perturbation parameters, it is common to base the inference on the profile likelihood. Ferrari and Uribe-Opazo (2001) obtained a Bartlett correction factor for the LR statistic in the symmetric linear models class, and Cordeiro (2004) extended this results to the symmetric non-linear model class. Ferrari et al. (2004, 2005) and Cysneiros and Ferrari (2006) showed that the combined use of modified profile likelihood and Bartlett correction can deliver accurate and reliable inference in small samples. Cordeiro et al. (2006) presented Bartlett adjustments for overdispersed generalized linear models. Savalli et al. (2006) discussed the problem of testing variance components in elliptical linear mixed models. Melo et al. (2009) developed modified versions of the likelihood ratio test for fixed effects inference in mixed linear models. Cysneiros et al. (2010) obtained a correction factor Bartlett-type for the score statistic in heteroskedastic symmetric non-linear models. However, it seems that there is no work in the literature about Bartlett correction for geoestatistical models.

Our main goal is on hypothesis test for Gaussian spatial linear models (GSLM) with repetitions considering the Matérn class of geostatistics models and to present the inference approach to estimate the smooth parameter from the Matérn family class of models. We present the likelihood ratio statistic and its corrected version by using Bartlett corrected factor. The Chapter unfolds as follows. Section 3.3 presents the Gaussian spatial linear model. Section 3.4 discuss techniques for the parameters estimation. In Section 3.5 the Matérn class of covariance functions models is described and the inference for the smoothness parameter is discussed. Furthermore we present hypothesis testing for the $\boldsymbol{\beta}$ parameter vector, and the respectively corrected test, using Bartlett correction, in Section 3.7.3. In Section 3.8 we show the Monte Carlo simulations results. Section 3.9 has applications to two real data sets. Finally, Section 3.10 contains some concluding remarks. Calculations are presented in the appendices.

3.3 Gaussian spatial linear model

For a spatial process, we consider is a stochastic process $\{\mathbf{Y}_i(\mathbf{s}), \mathbf{s} \in S \subset \mathbb{R}^2\}$ (bi-dimensional Euclidean space), usually though not necessarily in $\mathbb{R}^2$, and $i = 1$ means one single realization of the process. Let

$$\boldsymbol{\mu}(\mathbf{s}) = \mathrm{E}[\mathbf{Y}_1(\mathbf{s})],$$

denote the mean value at location $\mathbf{s}$. We also assume that the variance of $\mathbf{Y}_i(\mathbf{s})$ exists for all $\mathbf{s} \in S$. The process $\mathbf{Y}_1$ is said to be Gaussian if, for any $k \geq 1$ and locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$, the vector $(Y_1(\mathbf{s}_1), \ldots, Y_1(\mathbf{s}_n))$ has a multivariate normal distribution.

The treatment so far has been based on the assumption that inference must be based on a single realization of the random field $\mathbf{Y}_i$ $(i = 1)$. As mentioned by Smith (2001), we can expect to get better estimates if there are multiple repetitions of the process. Let $\mathbf{Y} = \mathbf{Y}(\mathbf{s}) = \mathrm{vec}(\mathbf{Y}_1(\mathbf{s}), ..., \mathbf{Y}_r(\mathbf{s}))$ be an $nr \times 1$ random vector of r independently stochastic process of n elements each, that belong to the family of Gaussian distributions and depend on the position $\mathbf{s} \in S \subset \mathbb{R}^2$. The "vec" operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. It is assumed the i-th stochastic process $\mathbf{Y}_i(\mathbf{s}) = \mathrm{vec}(Y_i(\mathbf{s}_1), ..., Y_i(\mathbf{s}_n))$, $\mathbf{s} \in S \subset \mathbb{R}^2$, represents the $n \times 1$ vector , for $i = 1, \ldots, r$. Considering the matrix notation, $\boldsymbol{\mu}_i(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}$, the model can be expressed as

$$\mathbf{Y}_i(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\epsilon}_i(\mathbf{s}), \tag{3.1}$$

for $i = 1, \ldots, r$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $p \times 1$ vector of unknown parameters, $\mathbf{X}(\mathbf{s}) = \mathbf{X}$ is an t$n \times p$ matrix with vth row $\mathbf{x}_v^\top$, where $x_{vj} = x_j(s_v)$ with $\mathbf{x}_v^\top = (x_{v1}, \ldots, x_{vp})$, for $v = 1, \ldots, n$. The design matrix $\mathbf{X}$ is the same for all r repetitions, $\mathbf{s} \in S \subset \mathbb{R}^2$, and $\boldsymbol{\epsilon}_i$ is the stochastic error.

Let $\boldsymbol{\Sigma}_i = [C_i(\mathbf{s}_u, \mathbf{s}_v)]$ be the $n \times n$ covariance matrix of $\mathbf{Y}_i(\mathbf{s})$ for i-th repetition, $i = 1, \ldots, r$. The matrix $\boldsymbol{\Sigma}_i$ is non-singular, symmetric and positive defined, associated to the vector $\mathbf{Y}_i(\mathbf{s})$, where for the stationary and isotropic process, the elements $C_i(\mathbf{s}_u, \mathbf{s}_v)$ depend on the Euclidean distance between points $\mathbf{s}_u$ and $\mathbf{s}_v$. If a stationary spatial random process has the property that the dependence between any two observations depends only on the distance between their locations, $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$, irrespective of the direction, then the process is said to be isotropic. Otherwise it is said to be anisotropic. Isotropic process are convenient to deal with because there are a widely used parametric forms for $C_i(d_{uv}) = C_i(\mathbf{s}_u, \mathbf{s}_v)$.

In the same approach of Smith (2001), we shall assume a homogeneous process. We consider the same covariance structure for each repetition, i.e., the covariance matrix $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = [C(\mathbf{s}_u, \mathbf{s}_v)]$, has a structure which depends on the vector of parameters $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)^\top$ or $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3, \phi_4)^\top$, depending on the form of the covariance structure (see Section 3.5 for more details). The covariance matrix is given by

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}, \tag{3.2}$$

for $i = 1, \ldots, r$, where, $\phi_1 \geq 0$ is the parameter named as nugget effect; $\phi_2 \geq 0$ is named as sill ; $\mathbf{R} = \mathbf{R}(\phi_3, \phi_4) = [(r_{uv})]$ or $\mathbf{R} = \mathbf{R}(\phi_3) = [(r_{uv})]$ is an $n \times n$ symmetric matrix, which is function of $\phi_3 > 0$, and sometimes also function of $\phi_4 > 0$, with diagonal elements $r_{uu} = 1, (u = 1, \ldots, n)$; $r_{uv} = \phi_2^{-1} C(s_u, s_v)$ for $\phi_2 \neq 0$, and $r_{uv} = 0$ for $\phi_2 = 0$, $u \neq v = 1, \ldots, n$, where r_{uv} depends on the Euclidean distance $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$ between points $\mathbf{s}_u$ and $\mathbf{s}_v$; ϕ_3 is a function of the model range (a), ϕ_4 when exists is known as the smoothness parameter, and $\mathbf{I}_n$ is an $n \times n$ identity matrix.

We assume that $\boldsymbol{\mu}(\mathbf{s})$ is constant which we may, without loss of generality, take to be 0, and then define

$$\mathrm{var}[\mathbf{Y}_1(\mathbf{s}_u) - \mathbf{Y}_1(\mathbf{s}_v)] = 2\gamma(||\mathbf{s}_u - \mathbf{s}_v||).$$

The function $\gamma(\cdot)$ is called the semivariance, and it makes sense only if the variance between observations depends on $\mathbf{s}_u$ and $\mathbf{s}_v$ only through their difference $\mathbf{s}_u - \mathbf{s}_v$. If the process is stationary then

$$\gamma(d_{uv}) = C(0) - C(d_{uv}).$$

Statistical methods for second-order stationary processes can be considered in terms of covariances or in terms of variograms. Statisticians prefer the first way, geostatisticians the second. We concentrate on the covariance.

According to Stein (1999), for a space-time process observed at a fixed set of spatial locations at sufficiently distant points in time, it may be reasonable to assume that observations from different times are independent realizations of a random field. Goodal and Mardia (1994) says we can remove the effect of time from the data and then we can view the data as independent repeated measurements in space.

Another characteristic of the model is that it can be expressed as a linear mixed model by

$$\mathbf{Y}_i(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{b}_i(\mathbf{s}) + \boldsymbol{\tau}_i(\mathbf{s})$$

where, the deterministic term $\boldsymbol{\mu}(\mathbf{s})$ is an $n \times 1$ vector, the means of the process $\mathbf{Y}_i(\mathbf{s})$, $\mathbf{b}_i(\mathbf{s}) + \boldsymbol{\tau}_i(\mathbf{s}) = \boldsymbol{\epsilon}_i(\mathbf{s})$ and $\mathbf{b}_i \sim N_n(\mathbf{0}, \phi_2\mathbf{R})$ and $\boldsymbol{\tau}_i \sim N_n(\mathbf{0}, \phi_1\mathbf{I})$ are independents, for $i = 1, \ldots, n$.

3.4 Parameters estimation

The unknown parameters to be estimated are the β's and ϕ's. Stein (1999) demonstrated with simulated data that the method of moments is poor at describing the smooth process. It is straightforward in principle to write down the exact likelihood function and hence maximize it numerically with respect to the unknown parameters in the case we assume that we are sampling from a Gaussian process. According to Smith (2001), despite of disadvantages such as non-robustness when there are outliers in the data, or the possible multimodality of the likelihood surface, these are not reasons to abandon maximum likelihood (ML) estimation. And, we can expect to get better estimates if there are multiple repetitions of $\mathbf{Y}_i$. The maximum likelihood procedure in this case is only slightly different from that in the single-realisation case.

Common practice is first to calculate the empirical variogram by the method of moments (Matheron, 1965), and then fit the model to the empirical variogram by (weighted) nonlinear least squares.

To resume, the weighted least squares (WLS) applied to the sample variogram gives a simple and convenient method for obtaining initial estimates of variogram parameters. Our wider conclusion is that the variogram should be used as a graphical method of exploratory data analysis, rather than as a vehicle for formal parameter estimation (Diggle and Ribeiro Jr., 2007). We suggest, for example, the use of robust semivariogram proposed in Genton (1998). Diggle and Ribeiro Jr. (2007) also pointed out that an unbiased set of estimating equations could be obtained from an iteratively weighted least squares algorithm, as used in generalized linear modelling (McCullagh and Nelder, 1989).

It can be used, for instance, the algorithms Newton-Raphson or score Fisher (Little and Rubin (1987)). In practice it should be combined such approaches with additional algorithms to avoid repeated inversion of large unstructured matrices (Mardia and Marshall (1984); Zimmerman (1989)). For example, Zimmerman (1989) devised a computational acceleration for the Gaussian covariance structure over a regular parallelogram lattice.

Diggle et al. (1998); Mardia and Watkins (1989); Warnes and Ripley (1987) and Zhang (2002) reported difficulties in likelihood estimation of covariogram parameters. Zimmerman and Zimmerman (1991) numerically compare ordinary least square (OLS), WLS, restricted maximum likelihood (REML) and ML for linear and exponential variograms. They find that, for Gaussian observations, ML is only slightly better that ordinary least square and WLS. REML, which often has the best bias properties, was found to have a comparable bias to ML. This is due to the low dimension of the mean parameter vector considered. We concentrate on maximum likelihood estimation to be presented in next Section.

There are some packages availables in the free software R Core Team (2015) that can be used for geostatistical models. For instance, the package `sp` (Pebesma and Bivand,

2005) presents classes and methods for spatial data in general, more specifically the classes document where the spatial location information resides, for 2D or 3D data, and utility functions are provided, e.g. for plotting data as maps, spatial selection, as well as methods for retrieving coordinates, for subsetting, and more. The package `geoR` (Ribeiro Jr and Diggle, 2001) for geostatistical analysis includes traditional, likelihood-based and Bayesian methods, the package `geoRgml` is an extension to the package geoR, and has functions for inference in generalised linear spatial models. The `gstat` (Pebesma, 2004) package is for spatial and spatio-temporal geostatistical modelling, prediction and simulation. The package `fields` (Nychka et al., 2014) and many more have tools to analyze spatial data. More availabe packages for spatial data can be found in the CRAN task view.

3.4.1 Maximum likelihood

The maximum likelihood (ML) method estimates parameters of the model directly from the data, on the assumption that it is a multivariate normal distribution. Let $\mathbf{Y} = \text{vec}(\mathbf{Y}_1^\top(s), ..., \mathbf{Y}_r^\top(s))$ be and $nr \times 1$ random vector of r independent and identical distributed vectors. The log-likelihood for the Gaussian spatial linear model for each repetition is of the form

$$l_i(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}).$$

The log-likelihood for the Gaussian spatial linear model for the r multiple repetitions is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{r} l_i(\boldsymbol{\theta}),$$

ie,

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{rn}{2}\log(2\pi) - \frac{r}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\sum_{i=1}^{r}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}). \tag{3.3}$$

The score functions are given by:

$$\begin{aligned}\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{r} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}_i, \\ \mathbf{U}(\boldsymbol{\phi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} &= -\frac{r}{2} \frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \phi} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) \\ &+ \frac{1}{2} \sum_{i=1}^{r} \frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top \boldsymbol{\Sigma}^{-1}),\end{aligned}$$

where $\boldsymbol{\epsilon}_i = (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})$. From the solution of the score function of $\boldsymbol{\beta}$,

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{0},$$

the maximum likelihood estimator $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}}, \tag{3.4}$$

where $\bar{\mathbf{Y}} = (\bar{\mathbf{Y}}_1, \ldots, \bar{\mathbf{Y}}_n)^\top$, with $\bar{\mathbf{Y}}_v = \frac{1}{r} \sum_{i=1}^{r} Y_i(s_v)$, $v = 1, \ldots, n$.

Since we can use the GLSLM with repetitions as a linear mixed model we can follow what was mentioned by Borssoi (2014), that Demidenko (2004) and Wang and Heckman (2009) discussed the parameter identifiability for the linear mixed models. In particular, Wang and Heckman (2009) showed that there is no problem since we have that $\operatorname{var}[\boldsymbol{\epsilon}_i] = \phi_1 \mathbf{I}_n$. But one deficiency can be the sensitivity of the maximum likelihood estimates to atypical observations. For that we present the study of diagnostic techniques in next Chapters.

Unfortunately, the score equation for $\boldsymbol{\phi}$ does not lead to a closed-form solution for $\boldsymbol{\phi}$. Thus, a common practice is to maximize the profile log-likelihood, obtained by substituting the solution to the score equation for $\boldsymbol{\beta}$ [Equation (3.4)] into the log-likelihood in Equation (3.3). Then, the profile log-likelihood depends only on $\boldsymbol{\phi}$, and ignoring the constant $-\frac{rn}{2}\log(2\pi)$ we have

$$\begin{aligned}\mathcal{L}^*(\boldsymbol{\phi}) \propto\ & -\frac{r}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^{r} \left[(\mathbf{Y}_i - \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}})^{*\top} \boldsymbol{\Sigma}^{-1} \right. \\ & \left. (\mathbf{Y}_i - \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}}) \right].\end{aligned} \tag{3.5}$$

Zhang and Zimmerman (2005) compare ML estimation under increasing domain and infill asymptotics in the Matérn family of covariances. They find that finite sample behavior agrees more with the infill asymptotic inferences than with the increasing domain asymptotics. For example, the maximum likelihood estimates of covariance parameters are typically consistent and asymptotically normal when fitted by a correct model (Mardia and Marshall, 1984). In contrast, not all covariance parameters can be estimated consistently under the fixed domain asymptotic framework, even for the simple exponential covariance model in one dimension with no consideration of explanatory variables (Chen et al., 2000; Ying, 1991). The readers are refereed to Stein (1999) for more details regarding fixed domain asymptotics. Some discussion concerning which asymptotic framework is more appropriate can also be found in Zhang and Zimmerman (2005).

Ying (1991) presents asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. Chang et al. (2014) studied the asymptotic properties of generalized information criterion (GIC) for geostatistical model selection regardless of whether the covariance model is correct or wrong, and establish conditions under which GIC is consistent and asymptotically loss efficient.

3.5 Matérn family of covariance function

According to Minasny and McBratney (2005) and Stein (1999) promoted the use of the Matérn class of models, which the name comes after the Swedish forestry statistician, Bertil Matérn (Matérn, 1960). And as we mentioned, the Matérn is a covariance function particulary attractive given by

$$C(d_{uv}) = \begin{cases} \dfrac{\phi_2}{2^{\phi_4-1}\Gamma(\phi_4)} \left(d_{uv}/\phi_3\right)^{\phi_4} K_{\phi_4}\left(d_{uv}/\phi_3\right), & d_{uv} > 0, \\ \phi_1 + \phi_2, & d_{uv} = 0, \end{cases} \tag{3.6}$$

where the parameters are assumed to be non-negative, i.e., $\phi_j \geq 0$, for $j = 1, 2, 3, 4$; $K_{\phi_4}(u) = \frac{1}{2}\int_0^\infty x^{\phi_4-1}e^{-\frac{1}{2}u(x+x^{-1})}dx$ is the modified Bessel function of the third kind of order ϕ_4; see (Gradshteyn and Ryzhik, 2000). This covariance function is particularly attractive because its behavior can change according to ϕ_4 value. For instance, $\phi_4 = 0.5$

gives the exponential covariance structure, Whittle's elementary covariance is obtained with $\phi_4 = 1$, and in the limit as ϕ_4 approaches to infinity, with ϕ_3 approaching to 0 in such way that $2\phi_4^{1/2}\phi_3$ remains constant, it approaches the Gaussian covariance function. For a finite value of the ϕ_3 parameter (which is a function of the range parameter), the Matérn function represents several bounded models. In many works the notation for the smoothness parameter is κ or ν for Matérn class, but here we use ϕ_4, for simplicity of notation, since it is a parameter part of the covariance structure which can be estimated or considered as fixed. For more details about Matérn class of models, (Haskard et al., 2007; Minasny and McBratney, 2005; Stein, 1999).

A process having the Matérn covariogram (3.6) is $[\phi_4] - 1$ times mean square differentiable, where $[\phi_4]$ is the largest integer less than or equal to ϕ_4. According to Zhang (2004), other classes of covariograms do not have such a parameter to yield a preferred mean square differentiability. The practical importance is that the Matérn class of covariance models allows data to determine the smoothness (Sherman, 2011).

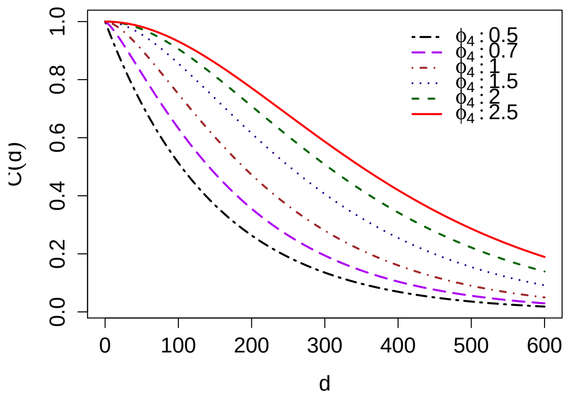
Table 3.1 presents few special cases of the Matérn class of models.

Table 3.1: Special cases of the Matérn covariance function.

smooth parameter	covariance function	model
$\phi_4 = 1/2$	$C(d_{uv}) = \phi_2 \exp(-d_{uv}/\phi_3)$	exponential
$\phi_4 = 1$	$C(d_{uv}) = \phi_2 (d_{uv}/\phi_3) K_{\phi_4}(d_{uv}/\phi_3)$	Whittle
$\phi_4 \to \infty$	$C(d_{uv}) = \phi_2 \exp(-(d_{uv}/\phi_3)^2)$	Gaussian

If ϕ_3 is large, it approximates the power function when $\phi_4 > 0$, and a log function or de Wijs function (de Wijs, 1951, 1953) when $\phi_4 \to 0$. Figure 3.1 shows some covariance models and the Matérn covariogram with varying smoothness parameter ϕ_4 and constant range and sill.

Figure 3.1: Matérn model with $\phi_1 = 0$, $\phi_2 = 1$, $\phi_3 = 150$ and different ϕ_4 values.



(a)
Source: From the author.

Matérn (1960) has shown more precisely that if a random field allows a variogram that is everywhere continuous except at the origin, then this random field is the sum of two uncorrelated random fields, one associated with a pure nugget effect and the other with an everywhere continuous variogram. Zhang's results demonstrate that, in a Matérn model with parameters ϕ_2 and ϕ_3 being $\phi_4 = 0.5$, the ratio ϕ_2/ϕ_3 is much more stably estimated than either ϕ_2 or ϕ_3 themselves. Stein (1999) strongly recommended to use the Matérn model by calculating and plot likelihood functions for unknown parameters of models for covariance structures.

For the case of the Matérn class of covariance function (or other with more than three parameters), we suggest to use the idea presented in Minasny and McBratney (2005) for one single realization, which assume ϕ_4 as fixed to find which ϕ_4 gives the maximum value for the profile log-likelihood, and then use it as a initial value and estimate it. We use the same idea, but now for a model with repetitions as we present in Section 3.9. Stein (1999) said that although he does not advocate treating $\phi_4 = 0.5$ as fixed, to keep in mind that is the same of using the exponential model.

In the case that we consider the parameter ϕ_4 as fixed, we have to choose which value, between a range of values, gives the best fit. As an alternative to select ϕ_4, we consider the cross validation criterion similar to the one presented in De Bastiani et al. (2015), defined by

$$CV = CV(\phi_4) = \frac{1}{nr}\sum_{i=1}^{r}\sum_{j=1}^{n}\left\{\frac{y_i(\mathbf{s}_j) - \hat{y}_{i(j)}(\mathbf{s}_j)}{1 - h_{jj}}\right\}^2, \tag{3.7}$$

where $\hat{y}_{i(j)}(\mathbf{s}_j) = \mathbf{x}_j^\top \hat{\boldsymbol{\beta}}_{(j)}$, with $\mathbf{x}_j^\top$ the ith row of the matrix $\mathbf{X}$, is the prediction in the location s_j without considering this observation, $(y_j, \mathbf{x}_j^\top)$, $\hat{\boldsymbol{\beta}}_{(j)}$ is the ML estimator of $\boldsymbol{\beta}$ without considering the jth observation, and h_{jj} is the jth diagonal element of the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1}$, for $j = 1, \ldots, n$. To simplify the calculations we can use the approximation $\hat{\boldsymbol{\beta}}_{(j)} \approx \hat{\boldsymbol{\beta}} + \mathbf{K}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{U}_{(j)}(\hat{\boldsymbol{\beta}})$, where

$$\mathbf{K}(\hat{\boldsymbol{\beta}}) = r\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} \quad \text{and}$$

$$\mathbf{U}_{(j)}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{r}\mathbf{X}_{(j)}^\top \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_{i(j)} - \mathbf{X}_{(j)}\hat{\boldsymbol{\beta}}), \tag{3.8}$$

with $\mathbf{X}_{(j)}$, $(n-1)\times p$, the matrix $\mathbf{X}$ without the jth row, $\mathbf{x}_j^\top$, y $\mathbf{Y}_{i(j)}$, $(n-1)\times 1$ denotes the vector $\mathbf{Y}_i$ without the response $y_i(s_j)$, for $j = 1, \ldots, n$ and $i = 1 \ldots, r$. Note that in (3.8) the matrix $\hat{\boldsymbol{\Sigma}}^{-1}$ is of order $(n-1)\times(n-1)$.

Minasny and McBratney (2005) say that the smoothness parameter, ϕ_4, from Matérn model should be determined from the spatial data.

3.5.1 Inference for ϕ_4 parameter

To estimate this parameter we used the profile likelihood method for ML. To summarize the main steps of the algorithm are:

1. Choose a set of values for ϕ_4, for instance $(0.2, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0)$.

2. For each value of ϕ_4, maximize the log-likelihood over (ϕ_1, ϕ_2, ϕ_3).

3. Plot the log-likelihood values as a functions of ϕ_4, and find the value of κ that has the log-likelihood maximum value (LMV).

4. Calculate the CV (Equation 3.7) and find the value of ϕ_4 that minimize CV.

5. Choose the values of $(\phi_1, \phi_2, \phi_3, \phi_4)$ where the log-likelihood is maximum and considered as a inicial choice for the next step.

6. Maximize the log-likelihood over all parameters $(\phi_1, \phi_2, \phi_3, \phi_4)$.

The derivatives of the Matérn model with respect to ϕ_4 parameter are given in Appendix B.3.

3.6 Asymptotic standard error estimation

The observed information matrix $I(\boldsymbol{\theta})$ for the GLSM with multiple replications is $I(\boldsymbol{\theta}) = -L(\boldsymbol{\theta})$, evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where $L(\boldsymbol{\theta}) = \partial^2\mathcal{L}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top = \ddot{\mathcal{L}}(\boldsymbol{\theta})$. Furthermore, we notice that the $I(\boldsymbol{\theta})$, which is evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, has the same equations as the expected information matrix given below. The derivatives of first and second-order of the scale matrix $\boldsymbol{\Sigma}$, with respect to ϕ_j, for $j = 1, 2, 3$ and $j = 4$ for some models; for some covariance functions are presented in Uribe-Opazo et al. (2012). And some useful constructions are given in Appendices A.4, B.1, and B.3.

Asymptotic standard errors can be calculated by inverting either the observed information matrix or the expected information matrix. The expected information matrix, $\mathrm{E}[-\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top] = \mathrm{E}[(\partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta})\ (\partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top)] = \mathrm{E}(\mathbf{U}_\theta\mathbf{U}_\theta^\top)$, is given by, (Lange et al., 1989; Mitchell, 1989; Waller and Gotway, 2004))

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{F} = \mathrm{E}[\mathbf{U}\mathbf{U}^\top] = \begin{pmatrix} \mathrm{E}(\mathbf{U}_\beta\mathbf{U}_\beta^\top) & \mathrm{E}(\mathbf{U}_\beta\mathbf{U}_\phi^\top) \\ \mathrm{E}(\mathbf{U}_\phi\mathbf{U}_\beta^\top) & \mathrm{E}(\mathbf{U}_\phi\mathbf{U}_\phi^\top) \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{\beta\beta} & \mathbf{F}_{\beta\phi} \\ \mathbf{F}_{\phi\beta} & \mathbf{F}_{\phi\phi} \end{pmatrix},$$

where U is the score function and

$$\begin{aligned} \mathbf{F}_{\beta\beta} &= r\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}, \\ \mathbf{F}_{\beta\phi} &= \mathbf{0}, \\ \mathbf{F}_{\phi\beta} &= \mathbf{0}, \\ \mathbf{F}_{\phi\phi} &= \frac{r}{2}\frac{\partial\,\mathrm{vec}^\top(\boldsymbol{\Sigma})}{\partial\boldsymbol{\phi}}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\frac{\partial\,\mathrm{vec}(\boldsymbol{\Sigma})}{\partial\boldsymbol{\phi}^\top}. \end{aligned}$$

See Appendix B.2 for the details. We used $F_{\beta\beta}^{-1}$ and $F_{\phi\phi}^{-1}$ to estimate the dispersion matrices for the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\phi}}$, respectively.

3.7 Hypothesis testing for the β parameter vector

The greater the number of perturbation parameter, less is the quality of asymptotic approximations. To overcome such problems, some modifications for the log-likelihood were proposed, as the one proposed by Barndorff-Nielsen (1983, 1994); Cox and Reid (1987); McCullagh and Tibishirani (1990) and Stern (1997), which are described in Pace and Savan (1997).

We can see our model defined in (3.1) as a particular case of the model given in Equation (1) of Melo et al. (2009) to be able to use some of their results. Next Chapter gives more details about this particular case. So $\mathcal{L}(\boldsymbol{\theta})$ is the total log-likelihood as defined in (3.3), given $\mathbf{Y} = (\mathbf{Y}_1^\top(s), ..., \mathbf{Y}_r^\top(s))^\top$, which depends on the parameters vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$, with $p+m$ components [m is the length of $\boldsymbol{\phi}$]. The parameters vector $\boldsymbol{\beta}$ of p components can be partitioned $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ where, $\boldsymbol{\beta}_1^\top = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q)^\top$ is the interest vector with dimension q and $\boldsymbol{\beta}_2^\top = (\boldsymbol{\beta}_{q+1}, \ldots, \boldsymbol{\beta}_p)^\top$ and the perturbations parameter vector with dimension $(p-q)$. Consequently we have $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$, where $\mathbf{X}_1$ is an $n \times q$ matrix formed by the first q columns of $\mathbf{X}$ and $\mathbf{X}_2$ and $n \times (p-q)$ matrix formed by the $(p-q)$ columns of $\mathbf{X}$.

3.7.1 Parameters orthogonalization

Let $\boldsymbol{\beta}_1$ be the parameter of interest, similarly to Zucker et al. (2000) and Melo et al. (2009) we transform the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\phi}^\top)^\top$ in $\boldsymbol{\vartheta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\xi}^\top, \boldsymbol{\phi}^\top)^\top$ with $\boldsymbol{\xi} = \boldsymbol{\beta}_2 + (\mathbf{X}_2^\top \boldsymbol{\Sigma}^{-1} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \boldsymbol{\Sigma}^{-1} \mathbf{X}_1 \boldsymbol{\beta}_1$.

Using this transformation presented, we have that $\boldsymbol{\beta}_1^\top$ and $\boldsymbol{\nu} = (\boldsymbol{\xi}^\top, \boldsymbol{\phi}^\top)^\top$ are orthogonal, since

$$\mathrm{E}\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\xi}^\top}\right] = \mathbf{0} \quad \text{and} \quad \mathrm{E}\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\phi}^\top}\right] = \mathbf{0}.$$

We have $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$, which after the orthogonalization it can be written as

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1^*\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\xi},$$

where $\mathbf{X}_1^* = \mathbf{X}_1 - \mathbf{X}_2(\mathbf{X}_2^\top \boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2^\top \boldsymbol{\Sigma}^{-1}\mathbf{X}_1$. Thus the log-likelihood can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\vartheta}) &= -\tfrac{rn}{2}\log(2\pi) - \tfrac{r}{2}log|\boldsymbol{\Sigma}| \\ &\quad -\tfrac{1}{2}\sum_{i=1}^{r}(\mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1 - \mathbf{X}_2\boldsymbol{\xi})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1 - \mathbf{X}_2\boldsymbol{\xi}).\end{aligned}$$

Let $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ be the hypothesis of interest versus the alternative hypothesis $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$, where $\boldsymbol{\beta}_1^0$ is a fixed vector of dimension $q(\leq p)$, where, $\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\xi}}^\top, \hat{\boldsymbol{\phi}}^\top)^\top$ is the unrestricted ML estimator for $\boldsymbol{\vartheta}$, and denote with a tilde the restricted ML estimator. So, $\tilde{\boldsymbol{\vartheta}} = (\boldsymbol{\beta}_1^{(0)\top}, \tilde{\boldsymbol{\xi}}^\top, \tilde{\boldsymbol{\phi}}^\top)^\top$ is the restricted ML estimator of $\boldsymbol{\vartheta}$, where $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\phi}}$ are the restricted ML estimators of $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ under H_0.

3.7.2 Likelihood ratio statistic

The likelihood ratio statistic (LR) to test $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ versus $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$, where $\boldsymbol{\beta}_1^0$ is a fixed vector of dimension $q(\leq p)$, is defined by

$$LR = 2(\mathcal{L}(\hat{\boldsymbol{\vartheta}}) - \mathcal{L}(\tilde{\boldsymbol{\vartheta}}))$$

where $\mathcal{L}(\hat{\boldsymbol{\vartheta}})$ is given by

$$\begin{aligned}\mathcal{L}(\hat{\boldsymbol{\vartheta}}) &= -\tfrac{rn}{2}\log(2\pi) - \tfrac{r}{2}log|\hat{\boldsymbol{\Sigma}}| \\ &\quad -\tfrac{1}{2}\sum_{i=1}^{r}(\mathbf{Y}_i - \mathbf{X}_1^*\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\xi}})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}_1^*\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\xi}}).\end{aligned}$$

and $\mathcal{L}(\tilde{\boldsymbol{\vartheta}})$ is given by

$$\begin{aligned}\mathcal{L}(\tilde{\boldsymbol{\vartheta}}) &= -\tfrac{rn}{2}\log(2\pi) - \tfrac{r}{2}log|\tilde{\boldsymbol{\Sigma}}| \\ &\quad -\tfrac{1}{2}\sum_{i=1}^{r}(\mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1^{(0)} - \mathbf{X}_2\tilde{\boldsymbol{\xi}})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1^{(0)} - \mathbf{X}_2\tilde{\boldsymbol{\xi}}).\end{aligned}$$

Thus, the likelihood ratio statistic has the form:

$$LR = \frac{r}{2}\log\left(\frac{\hat{\boldsymbol{\Sigma}}}{\tilde{\boldsymbol{\Sigma}}}\right) + \frac{1}{2}\sum_{i=1}^{r}(\tilde{\boldsymbol{\delta}}_i - \hat{\boldsymbol{\delta}}_i),$$

where, $\tilde{\boldsymbol{\delta}}_i = (\mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1^{(0)} - \mathbf{X}_2\tilde{\boldsymbol{\xi}})^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1^{(0)} - \mathbf{X}_2\tilde{\boldsymbol{\xi}})$, is evaluated in $\tilde{\boldsymbol{\vartheta}}$, and $\hat{\boldsymbol{\delta}}_i = (\mathbf{Y}_i - \mathbf{X}_1^*\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\xi}})^\top \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_i - \mathbf{X}_1^*\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\xi}})$, is evaluated in $\hat{\boldsymbol{\vartheta}}$.

The log-likelihood given in (3.5) is a profile likelihood, which can be written considering the partition of the parameters vector $\boldsymbol{\beta}$ how we presented in Section 3.7. Thus we can define the profile log-likelihood by

$$\mathcal{L}_p(\boldsymbol{\beta}_1) = \mathcal{L}(\boldsymbol{\beta}_1, \hat{\boldsymbol{\xi}}^*, \hat{\boldsymbol{\phi}}^*), \tag{3.9}$$

where $\hat{\boldsymbol{\xi}}^*$ and $\hat{\boldsymbol{\phi}}^*$ are the maximum likelihood estimates considering $\boldsymbol{\beta}_1$ as fixed.

For the hypothesis test under $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$ *versus* $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^{(0)}$, with $\boldsymbol{\beta}_1^{(0)}$ as fixed, the likelihood ratio statistic based on (3.9) is the same as the one based on the non profile log-likelihood given in (3.5), (Ferrari et al., 2005), i.e,

$$LR = 2(\mathcal{L}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\phi}}) - \mathcal{L}(\boldsymbol{\beta}_1^{(0)}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\phi}})) = 2(\mathcal{L}_p(\hat{\boldsymbol{\beta}}_1) - \mathcal{L}_p(\boldsymbol{\beta}_1^{(0)})).$$

Asymptotically and under the null hypothesis, the LR statistic is distributed as a chi-squared random variable, with degrees of freedom equal to $(p - q)$.

3.7.3 Improved likelihood ratio tests - Bartlett correction

For small sample size the first order approximation in general may be not satisfactory. Looking for improving this approximation, Bartlett (1937) proposed a new statistics which consist of the multiplication of the LR statistics by a constant $(1 + C/q)^{-1}$, obtaining the corrected version of the LR statistics, LR*, given by

$$LR^* = \frac{LR}{1 + C/q},$$

where C is a constant of order n^{-1}, $O(n^{-1})$, chosen such that, under the null hypothesis

$$E[LR^*] = q + O(n^{-3/2}).$$

In regular problems and under the null hypothesis, LR* has a chi-squared distribution, χ^2_q less than an error n^{-2}, see (Barndorff-Nielsen and Hall, 1988). Thus, while $P(LR > x_\alpha) = \alpha + O(n^{-1})$, $P(LR^* > x_\alpha) = \alpha + O(n^{-2})$, where x_α is the quantile $(1 - \alpha)$ of the distribution χ^2_q. In other words, the error of approximation by χ^2_q for the null distribution of LR is of order n^{-1}, while the error of this approximation for the distribution of LR^* is reduced to order n^{-2}.

Several methods to obtain the Bartlett correction factor are presented in the literature, among them is the one presented in Lawley (1956) that consider, for regular problems, to obtain a general formula for the constant C in terms of the cumulants and your derivatives of the log-likelihood. Then, the Bartlett correction factor for the GSLM with repetitions is obtained. For simplicity, here we only give the expression for C when $\boldsymbol{\beta}_1^{(0)} = \mathbf{0}$.

$$C = \operatorname{tr}\left(\frac{1}{r}\mathbf{D}^{-1}\left(-\frac{1}{2}\mathbf{M} + \frac{1}{4}\mathbf{P} - \frac{1}{2}(\boldsymbol{\varsigma} + \boldsymbol{\nu})\boldsymbol{\tau}^\top\right)\right),$$

where $\mathbf{D}$, $\mathbf{M}$ and $\mathbf{P}$ are $m \times m$ matrices given by

$$\begin{aligned}
\mathbf{D} =& \frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_k}\right),\\
\mathbf{M} =& \operatorname{tr}\left[\left(\mathbf{X}_1^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^*\right)^{-1}\left(\mathbf{X}_1^{*\top}\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_j \partial \phi_k}\mathbf{X}_1^* + 2\dot{\mathbf{X}}_{1k}^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_1^*\right)\right],\\
\mathbf{P} =& \operatorname{tr}\left[\left(\dot{\mathbf{X}}_1^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_1^*\right)\left(\mathbf{X}_1^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^*\right)^{-1}\left(\dot{\mathbf{X}}_1^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_k}\mathbf{X}_1^*\right)\left(\mathbf{X}_1^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^*\right)^{-1}\right],
\end{aligned}$$

and $\boldsymbol{\varsigma}$, $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$ are (m)-vector whose jth elements are

$$\begin{aligned}
&\operatorname{tr}\left(D^{-1}\mathbf{A}_j\right)\\
&\operatorname{tr}\left[\left(\mathbf{X}_2^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\right)^{-1}\left(\mathbf{X}_2^\top\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_2\right)\right]\\
&\operatorname{tr}\left[\left(\mathbf{X}_1^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^*\right)^{-1}\left(\mathbf{X}_1^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_1^*\right)\right],
\end{aligned}$$

and

$$\mathbf{A}_j = \frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_l}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \phi_j \phi_k}\right) - \frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_k}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \phi_j \phi_l}\right) - \frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \phi_k \phi_l}\right).$$

Appendix B.4 shows the details.

3.8 Monte Carlo Simulations

In general, Monte Carlo methods are all methods that examinate the properties of a probability distribution, by generating a sample from the distribution and then studying the statistical properties in this sample.

Let $\mathcal{R}$ be rejection area. To estimate the P[error type I] for small values of r by the Monte Carlo simulation the main steps are:

1- Generate N independent sample according the null hypothesis,

$$(\mathbf{Y}_1^{(k)}, \ldots, \mathbf{Y}_r^{(k)}),$$

for $k = 1, \ldots, N$ and each $\mathbf{Y}_i$ is a random vector of lenght n,

2- calculate the statistic,

$$T^{(k)} = T(\mathbf{Y}_1^{(k)}, \ldots, \mathbf{Y}_r^{(k)}),$$

for $k = 1, \ldots, N$,

3- determine the percentage of samples of H_0 that are rejected (wrongly), that is

$$P[\text{error type I}] = P[T \in \mathcal{R}] \approx \frac{1}{N}\sum_{k=1}^{N} 1_{T^{(k)} \in \mathcal{R}}.$$

All simulations were performed using the Ox matrix programming language (Doornik, 2006). We consider 10000 Monte Carlo replications, $n = 25$ the sample size in a regular grid of 5×5 units.

3.8.1 Simulation 1

We present the results of Monte Carlo simulation experiments in which we evaluate the finite sample performances of the likelihood ratio test and its Bartlett-corrected version.

The simulations were based on the following linear model $\mathbf{Y}_i = \beta_0 + \beta_1\mathbf{x}_1 + \boldsymbol{\epsilon}$, for $i = 1, \ldots, 5$. The values of $\mathbf{x}_1$ were obtained as random draws from the standard uniform distribution $\mathcal{U}(0, 1)$, and $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \phi_1\mathbf{I}_n + \phi_2\mathbf{R}(\phi_3)$ considering the exponential geostatistical model. We test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. The parameters values are $\beta_0 = 0.5$, $\beta_1 = 0.5$, $\phi_1 = 3.0$, we varied $\phi_2 = 0.5$ and $\phi_2 = 3.0$, and $\phi_3 = 3.0$ treated as fixed in a sample grid of 5×5. All tests were carried out at the following nominal levels: $\alpha = 1\%$, $\alpha = 5\%$ and $\alpha = 10\%$. The results are presented in Table 3.2. Note that, apart from the first situation, all the statistics present the rejection rates above the nominal level, but the corrected version of the likelihood statistic presents slightly better results.

Table 3.2: Null rejection rates of the tests of $H_0 : \beta_1 = 0$. For $r = 5$ repetitions of the exponential geostatistical model with $q = 1$ in an regular grid with $n = 25$.

ϕ_2	$\alpha(\%)$	LR	LR*
0.5	1.00	1.40	0.94
0.5	5.00	11.49	8.89
0.5	10.00	23.17	19.72
3.0	1.00	2.66	1.89
3.0	5.00	12.11	10.28
3.0	10.00	21.24	19.08

Table 3.3 presents the mean and variance of the χ^2_1, LR and LR* statistics. The results showed that the LR* statistics present mean and variance closest to the mean and variance of the χ^2_1 then the mean and variance of LR, but still they exceed the values of the χ^2_1 statistics.

Table 3.3: Mean and variance of the χ^2_1, LR and LR*. For $r = 5$ repetitions of the exponential geostatistical model in a regular grid with $n = 25$.

ϕ_2	Mean χ^2_1	Var χ^2_1	Mean LR	Var LR	Mean LR*	Var LR*
0.5	1.00	2.00	1.8739	19.009	1.7131	16.476
3.0	1.00	2.00	1.6310	7.3866	1.5142	6.7151

3.8.2 Simulation 2

This simulation is similar to the first one, but now we increased the number of parameters to be tested. The simulation were based on the linear model $\mathbf{Y}_i = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \boldsymbol{\epsilon}$, for $i = 1, \ldots, 5$. The values of $\mathbf{x}_1$ were obtained as random draws from the standard uniform distribution $\mathcal{U}(0, 1)$, $\mathbf{x}_2$ is and indicator function and $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} =$

$\phi_1\mathbf{I}_n+\phi_2\mathbf{R}(\phi_3)$ considering the exponential geostatistical model. We test $H_0 : \beta_1 = \beta_2 = 0$ against $H_1 : \beta_1 \neq 0$ and $\beta_2 \neq 0$. The parameters values are $\beta_0 = 0.5$, $\beta_1 = 0.5$, $\phi_1 = 3.0$, we varied $\phi_2 = 0.5$ and $\phi_2 = 3.0$, and $\phi_3 = 3.0$ treated as fixed in a sample grid of 5×5. All tests were carried out at the following nominal levels: $\alpha = 1\%$, $\alpha = 5\%$ and $\alpha = 10\%$. The null rejection rates of the test under evaluation is displayed in Table 3.4. Note that all the statistics present the rejection rates above the nominal level. As expected, the corrected version of the likelihood ratio statistic shows null rejection rates closer to the nominal level compared to the no corrected version.

Table 3.4: Null rejection rates of the tests of $H_0 : \beta_1 = \beta_2 = 0$. For $r = 5$ repetitions of the exponential geostatistical model with $q = 2$ in an regular grid with $n = 25$.

ϕ_2	$\alpha(\%)$	LR	LR*
0.5	1.00	2.00	1.82
0.5	5.00	7.47	6.80
0.5	10.00	14.08	13.21
3.0	1.00	28.91	26.35
3.0	5.00	16.1	13.96
3.0	10.00	28.91	26.35

The mean and variance of the χ^2_2, LR and LR* statistics are presented in Table 3.5. The results showed that the LR* statistics present mean and variance closer to the mean and variance of the χ^2_2 compared the mean and variance of LR. On the other hand the values exceed the one of the χ^2_2 statistics, specially the variance of LR and LR* statistics.

Table 3.5: Mean and variance of the χ^2_2, LR and LR*. For $r = 5$ repetitions of the exponential geostatistical model in an regular grid with $n = 25$.

ϕ_2	Mean χ^2_2	Varχ^2_2	Mean LR	Var LR	Mean LR*	Var LR*
0.5	2.00	4.00	1.5940	18.897	1.5361	17.917
3.0	2.00	4.00	2.1248	4.6723	1.9869	4.1811

3.9 Application

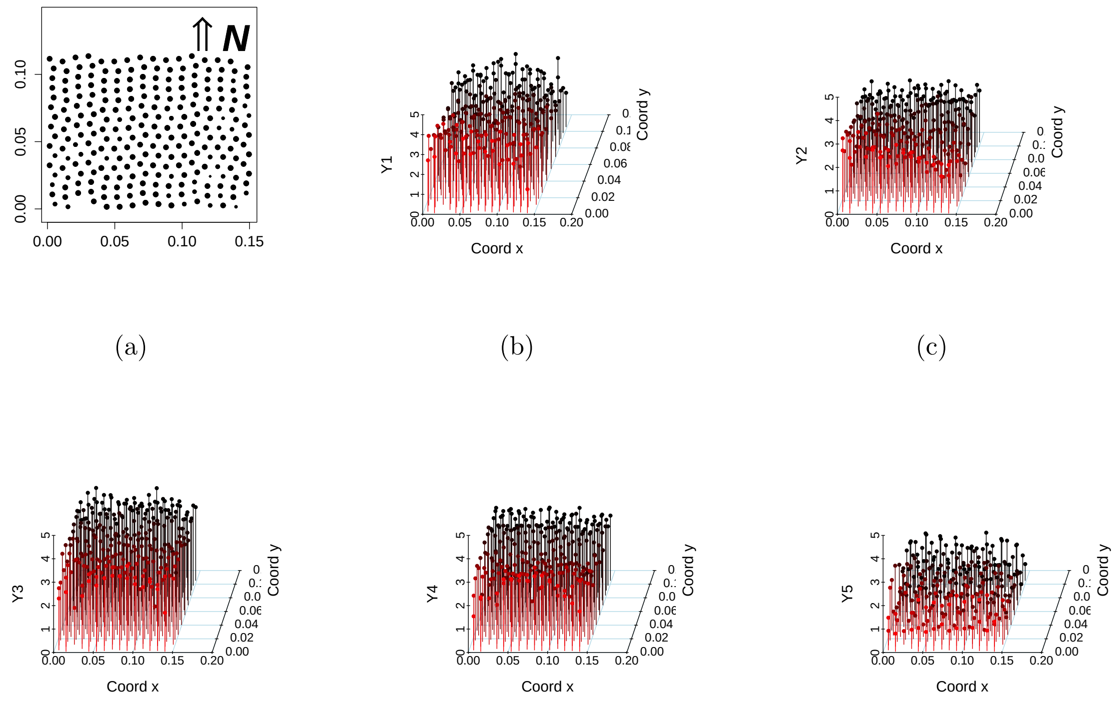
The data set were collected in a grid of 7.20×7.20 *m* in an experimental area with 1.33 *ha* at Eloy Gomes Researh Center at *Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda* (COODETEC), in Cascavel city at Paraná State - Brazil, with Oxisol soil. It were collected soybean productivity data and four chemical contents during April in the years 1998 (Y_1) or year 1, 1999 (Y_2) year 2, 2000 (Y_3) year 3, 2001 (Y_4) year 4 and 2002 (Y_5) year 5 with 253 observations each. To explain the expectation value of the productivity, it were considered as explanatory variables in the model these chemical contents of soil: phosphorus (P)[mg dm^{-3}], potassium (K)[cmolc dm^{-3}], calcium (Ca)[cmolc dm^{-3}] and magnesium (Mg)[cmolc dm^{-3}].

Table 3.6 presents a descriptive analysis of productivity dataset. The smallest value for the mean of the productivity is for Prod 2002 (Y_5). This year has the biggest variance coefficient (var. coef.) and beyond that, in general the lower values as shown in Figures 3.2 and 3.3. The observations were taken at the same site for each repetition and Figure 3.2(a) show how the data are spread in the area. Figures 3.2(b), 3.2(c), 3.2(d), 3.2(e) and 3.2(f) are the scatterplot for each repetition. The values of the productivity have been decreasing along the years. Figure 3.3 show the usual boxplots for the productivity, where only the observations taken in the harvest year 2002 do not have outliers.

Table 3.6: Descriptive analysis of productivity dataset.

Prod	Min.	1st Quartil	Median	Mean	3rd Quartil	Max.	var. coef.
1998	1.190	2.420	2.740	2.755	3.050	4.140	0.18
1999	0.690	1.960	2.190	2.154	2.390	3.610	0.18
2000	1.260	2.930	3.150	3.106	3.330	3.980	0.12
2001	1.310	2.520	2.720	2.644	2.880	3.340	0.14
2002	0.190	0.770	1.220	1.283	1.820	2.340	0.44

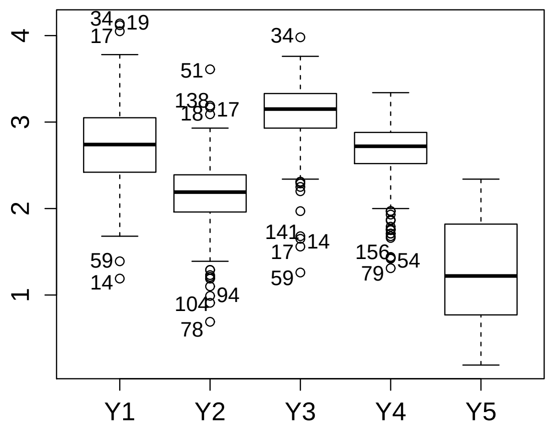
Figure 3.2: (a) Plot of how the data is spread in the area; each point has five repetitions, (b), (c), (d), (e) and (f) scatterplot for each repetition, for years 1998, 1999, 2000, 2001 and 2002, respectively.



(d) (e) (f)

Source: From the author.

Figure 3.3: Boxplot for soybean productivity in the years 1998 (Y_1), 1999 (Y_2), 2000 (Y_3), 2001 (Y_4) and 2002 (Y_5).



Source: From the author.

In other to choose the variance covariance structure that best describe the spatial dependence of the soybean productivity, we used cross validation (CV), the log-likelihood maximum value (LMV) and the trace of the asymptotic covariance matrix of an estimated mean (Tr) criteria presented in Table 4.1 and in Figure 4.2 of Chapter 4. Based on this criteria we selected the Gaussian covariance structure to model the spatial variability.

In Table 3.7 are presented the results to test the fixed effects, considering the Gaussian covariance structure. We observe that in all situations the null hypothesis is rejected when considering 10% of significant level, and the only hypothesis that is not rejected when considering 5% and 1% of significant is the first one, $\boldsymbol{\beta}_1 = 0$.

Table 3.7: Profile Likelihood ratio test and Bartlett correction version with the respectives p-values, for the soybean productivity data set.

H_0	LR	LR-p-value	LR^*	LR^*-p-value
$\beta_1 = 0$	2.248747	0.0864	2.213197	0.0887
$\beta_2 = 0$	22.65191	0.0000	22.28766	0.0000
$\beta_3 = 0$	12.35896	0.0002	12.04922	0.0003
$\beta_4 = 0$	17.10885	0.0000	16.80962	0.0000
$\beta_1 = 0$ and $\beta_2 = 0$	24.10884	0.0000	23.44502	0.0000
$\beta_3 = 0$ and $\beta_4 = 0$	23.35274	0.0000	22.30249	0.0000
$\beta_1 = 0$ and $\beta_3 = 0$	16.57818	0.0001	15.97536	0.0000
$\beta_1 = 0$ and $\beta_4 = 0$	16.57818	0.0001	16.10108	0.0000
$\beta_2 = 0$ and $\beta_3 = 0$	31.26882	0.0000	30.14944	0.0000
$\beta_2 = 0$ and $\beta_4 = 0$	36.73298	0.0000	35.68228	0.0000
$\beta_2 = 0$; $\beta_3 = 0$ and $\beta_4 = 0$	42.5386	0.0000	40.33767	0.0000
$\beta_1 = 0$; $\beta_3 = 0$ and $\beta_4 = 0$	27.86659	0.0000	26.41378	0.0000
$\beta_1 = 0$; $\beta_2 = 0$ and $\beta_4 = 0$	39.56427	0.0000	38.15563	0.0000
$\beta_1 = 0$; $\beta_2 = 0$ and $\beta_3 = 0$	33.53977	0.0000	32.10161	0.0000
$\beta_1 = 0$; $\beta_2 = 0$; $\beta_3 = 0$ and $\beta_4 = 0$	45.02532	0.0000	42.55362	0.0000

Based on the likelihood ratio test and its respective Bartlett corrected version, we considered all the explanatory variables in the analysis.

Table 3.8 shows the parameters estimates considering the Gaussian covariance function, and the respective asymptotic standard errors (se) in parenthesis. We note that P and Mg has an inverse proportional relationship with the mean of the productivity, that is, after a determined value as we increase the contents of P and Mg, the mean of productivity decrease. The opposite happen with K and Ca. For the Gaussian covariance structure, the range of the model is given by $\sqrt{3}\phi_3$, so we conclude that under this scenario, the range of spatial dependence is approximately 0.0705 m.

Table 3.8: Parameters estimates for Coodetec data considering the Gaussian covariance function, the asymptotic standard errors (in parenthesis) and the p-values.

Intercept	P	K	Ca	Mg	nugget	sill	f(range)
$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
2.4013	-0.0017	0.3270	0.0118	-0.0592	0.1927	0.0579	0.0407
(0.1020)	(0.0109)	(0.1746)	(0.0392)	(0.0515)	(0.0178)	(0.0219)	(0.0098)

3.10 Conclusions

We performed likelihood-based testing inference on the parameters of a spatial linear model. We had written the model as a random effect model, so is the same to say that we addressed the issue to test the fixed effects parameters of mixed linear models when the sample contains a small number of observations. We allow to test joint restrictions on one or more fixed effects parameters.

For the spatial linear model the covariance matrix of the random effects is not necessarily linear when deriving the Bartlett correction to the profile likelihood ratio test. We obtained the likelihood ratio statistic and the corrected version by using Bartlett corrected factor.

The standard likelihood ratio test is liberal, as evidenced by our Monte Carlo results. The simulation study clearly show that the proposed tests slightly outperform the standard likelihood ratio test. It needs a deeper study to obtain more results.

We also present some useful derivatives for the selection and estimation procedure of the covariance matrix that define the spatial dependence structure, considering the Matérn family of geostatistical models.

Appendix B

B.1 Covariance matrix derivatives

Let us consider the covariance matrix $\mathbf{\Sigma} = \phi_1 \mathbf{I} + \phi_2 \mathbf{R}$ given in 3.2 , where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$ are unknown parameters, $\mathbf{I}$ is an $n \times n$ identity matrix and $\mathbf{R}$ is defined according the chosen geostatistical model (Uribe-Opazo et al., 2012). Without loss of generality, for $q = 3$ we have

$$\operatorname{vec}^\top(\mathbf{\Sigma}) = \operatorname{vec}^\top(\phi_1 \mathbf{I} + \phi_2 \mathbf{R}) = \phi_1 \operatorname{vec}^\top(\mathbf{I}) + \phi_2 \operatorname{vec}^\top(\mathbf{R}),$$

consequently,

$$\frac{\partial \operatorname{vec}^\top(\mathbf{\Sigma})}{\partial \boldsymbol{\phi}} = \frac{\partial \phi_1 \operatorname{vec}^\top(\mathbf{I})}{\partial \boldsymbol{\phi}} + \frac{\partial \phi_2 \operatorname{vec}^\top(\mathbf{R})}{\partial \boldsymbol{\phi}}, \tag{B.1}$$

where the "vec" operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. The first part of the Equation (B.1) is given by

$$\frac{\partial \phi_1 \operatorname{vec}^\top(\mathbf{I})}{\partial \boldsymbol{\phi}} = \begin{bmatrix} \dfrac{\partial \phi_1 \operatorname{vec}^\top(\mathbf{I})}{\partial \phi_1} \\ \dfrac{\partial \phi_1 \operatorname{vec}^\top(\mathbf{I})}{\partial \phi_2} \\ \dfrac{\partial \phi_1 \operatorname{vec}^\top(\mathbf{I})}{\partial \phi_3} \end{bmatrix} = \begin{bmatrix} \operatorname{vec}^\top(\mathbf{I}) \\ 0 \\ 0 \end{bmatrix}$$

and the second part of the Equation (B.1) is given by

$$\frac{\partial \phi_2 \operatorname{vec}^\top(\mathbf{R})}{\partial \boldsymbol{\phi}} = \begin{bmatrix} \dfrac{\partial \phi_2 \operatorname{vec}^\top(\mathbf{R})}{\partial \phi_1} \\ \dfrac{\partial \phi_2 \operatorname{vec}^\top(\mathbf{R})}{\partial \phi_2} \\ \dfrac{\partial \phi_2 \operatorname{vec}^\top(\mathbf{R})}{\partial \phi_3} \end{bmatrix} = \begin{bmatrix} 0 \\ \operatorname{vec}^\top\left(\dfrac{\partial \mathbf{R}}{\partial \phi_2}\right) \\ \phi_2 \operatorname{vec}^\top\left(\dfrac{\partial \mathbf{R}}{\partial \phi_3}\right) \end{bmatrix}.$$

Thus, the solution for Equation (B.1) is an $3 \times n^2$ matrix given by

$$\frac{\partial \operatorname{vec}^\top(\Sigma)}{\partial \boldsymbol{\phi}} = \begin{bmatrix} \operatorname{vec}^\top(\mathbf{I}) \\ \operatorname{vec}^\top(\mathbf{R}) \\ \phi_2 \operatorname{vec}^\top\left(\dfrac{\partial \mathbf{R}}{\partial \phi_3}\right) \end{bmatrix}.$$

B.2 Expected information matrix

Asymptotic standard errors can be calculated by inverting either the observed information matrix or the expected information matrix. The expected information matrix is given by

$$\mathbf{F}(\theta) = \mathbf{F} = \begin{pmatrix} \mathrm{E}(\mathbf{U}_\beta \mathbf{U}_\beta^\top) & \mathrm{E}(\mathbf{U}_\beta \mathbf{U}_\phi^\top) \\ \mathrm{E}(\mathbf{U}_\phi \mathbf{U}_\beta^\top) & \mathrm{E}(\mathbf{U}_\phi \mathbf{U}_\phi^\top) \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{\beta\beta} & \mathbf{F}_{\beta\phi} \\ \mathbf{F}_{\phi\beta} & \mathbf{F}_{\phi\phi} \end{pmatrix},$$

where $\mathbf{U}_\beta = \mathbf{U}(\beta)$ and $\mathbf{U}_\phi = \mathbf{U}(\phi)$ are the score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, respectively.

Let define $\mathbf{Z}_i = \Sigma^{-1/2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})$, where $\mathbf{Y}_i \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ then $\mathbf{Z}_i \sim N_n(\mathbf{0}, \mathbf{I})$, where $\mathbf{0}$ is an $n \times p$ null matrix and $\mathbf{I}$ is an $n \times n$ identity matrix. Thus,

$$\begin{aligned}
\mathbf{U}_\beta \mathbf{U}_\beta^\top &= \sum_{i=1}^{r} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \mathbf{Y}_i^\top \boldsymbol{\Sigma}^{-1/2} \mathbf{X}, \\
\mathbf{U}_\beta \mathbf{U}_\phi^\top &= -\frac{1}{2} \sum_{i=1}^{r} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top} \\
&\quad + \frac{1}{2} \sum_{i=1}^{r} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \mathbf{Y}_i^\top \boldsymbol{\Sigma}^{-1/2}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top}, \\
\mathbf{U}_\phi \mathbf{U}_\phi^\top &= T_1 + T_2 + T_3 + T_4,
\end{aligned}$$

where,

$$
\begin{aligned}
T_1 &= \frac{r}{4}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top},\\
T_2 &= -\frac{1}{4}\sum_{i=1}^{r}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1/2}\mathbf{Y}_i\mathbf{Y}_i^\top\boldsymbol{\Sigma}^{-1/2}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top},\\
T_3 &= -\frac{1}{4}\sum_{i=1}^{r}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1/2}\mathbf{Y}_i\mathbf{Y}_i^\top\boldsymbol{\Sigma}^{-1/2}) \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top},\\
T_4 &= \frac{1}{4}\sum_{i=1}^{r}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1/2}\mathbf{Y}_i\mathbf{Y}_i^\top\boldsymbol{\Sigma}^{-1/2}) \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1/2}\mathbf{Y}_i\mathbf{Y}_i^\top\boldsymbol{\Sigma}^{-1/2}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top},
\end{aligned}
$$

which implies that $\mathbf{F}_{\beta\beta} = r\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\mathbf{F}_{\beta\phi} = \mathbf{0}$, $\mathbf{F}_{\phi\beta} = \mathbf{0}$, and

$$
\begin{aligned}
\mathbf{F}_{\phi\phi} &= \mathrm{E}[T_1] + \mathrm{E}[T_2] + \mathrm{E}[T_3] + \mathrm{E}[T_4]\\
&= -\frac{r}{4}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1}) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top}\\
&\quad +\frac{1}{4}\sum_{i=1}^{r}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}}\Big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}) + (\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2})K_n(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{\Sigma}^{-1/2})\\
&\quad + \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) \operatorname{vec}^\top(\boldsymbol{\Sigma}^{-1})\Big]\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top}\\
&= \frac{r}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top},
\end{aligned}
$$

where $\otimes$ denote the kronecker product and K_n is the commutation matrix.

B.3 Derivative for smoothness parameter of Matérn class of models

For the Matérn class of covariance structure there is an extra parameter, known as smooth parameter, that give more flexibility for this class. To get the asymptotic standard error of such parameter we need the first derivative of the log-likelihood in respect of it. For that, we used results presented on Mencía and Sentana (2005) and Harkard (2007), based on Abramowitz and Stegun (1965) as we present in the following.

The modified Bessel function of the third kind with order ϕ_4, which we denote as $K_{\phi_4}(.)$, is closely related to the modified Bessel function of the first kind $I_{\phi_4(.)}$, as

$$
K_{\phi_4}\left(\frac{d}{\phi_3}\right) = \frac{\pi}{2}\frac{I_{-\phi_4}(\frac{d}{\phi_3}) - I_{\phi_4}(\frac{d}{\phi_3})}{\sin(\pi\phi_4)}. \qquad \text{(B.2)}
$$

Mencía and Sentana (2005) present some comments according to their experience in relation to the stability of the modified Bessel functions derivative. For example, for $\phi_4 > 0$ and $(d/\phi_3) > 12$, the derivative of (B.2) with respect to ϕ_4 gives a better approximation than the direct derivative of $K_{\phi_4}(\frac{d}{\phi_3})$, which is in fact very unstable.

To evaluate the derivative with respect to ϕ_4 at $\phi_4 \neq 0, \pm 1, \pm 2, \ldots$, use

$$\frac{\partial K_{\phi_4}\left(\frac{d}{\phi_3}\right)}{\partial \phi_4} = \frac{1}{2}\pi \csc(\phi_4 \pi)\left[\frac{\partial I_{-\phi_4}\left(\frac{d}{\phi_3}\right)}{\partial \phi_4} - \frac{\partial I_{\phi_4}\left(\frac{d}{\phi_3}\right)}{\partial \phi_4}\right] - \pi \cot(\phi_4 \pi) K_{\phi_4}\left(\frac{d}{\phi_3}\right),$$

where

$$\frac{\partial I_{\phi_4}\left(\frac{d}{\phi_3}\right)}{\partial \phi_4} = I_{\phi_4}\left(\frac{d}{\phi_3}\right)\ln\left(\frac{1}{2}\frac{d}{\phi_3}\right) - \left(\frac{1}{2}\frac{d}{\phi_3}\right)^{\phi_4}\sum_{k=0}^{\infty}\frac{\psi(\phi_4+k+1)[\frac{1}{4}(\frac{d}{\phi_3})^2]^k}{\Gamma(\phi_4+k+1)k!},$$

and

$$\frac{\partial I_{-\phi_4}\left(\frac{d}{\phi_3}\right)}{\partial \phi_4} = -I_{-\phi_4}\left(\frac{d}{\phi_3}\right)\ln\left(\frac{1}{2}\frac{d}{\phi_3}\right) + \left(\frac{1}{2}\frac{d}{\phi_3}\right)^{-\phi_4}\sum_{k=0}^{\infty}\frac{\psi(-\phi_4+k+1)[\frac{1}{4}(\frac{d}{\phi_3})^2]^k}{\Gamma(-\phi_4+k+1)k!},$$

where ψ is the digamma function, $\psi = \frac{d}{dx}\{\ln\Gamma(x)\} = \Gamma'(x)\Gamma(x)$ and

$$I_{-\phi_4}(\frac{d}{\phi_3}) = \frac{2}{\pi}\sin(\phi_4\pi)K_{\phi_4}(\frac{d}{\phi_3}) + I_{\phi_4}(\frac{d}{\phi_3}).$$

For evaluating at integer values of ϕ_4,

$$\left.\frac{\partial K_{\phi_4}\left(\frac{d}{\phi_3}\right)}{\partial \phi_4}\right|_{\phi_4=m} = \frac{m!(\frac{1}{2}\frac{d}{\phi_3})^{-m}}{2}\sum_{k=0}^{m-1}\frac{(\frac{1}{2}\frac{d}{\phi_3})^k K_k(\frac{d}{\phi_3})}{(m-k)k!},$$

for $m = 1, 2, 3 \ldots$.

Continuing to find $\partial(\mathbf{R})/\partial\phi_4$ for Matérn model we have that

$$\frac{\partial\left(\frac{d}{\phi_3}\right)^{\phi_4}}{\partial \phi_4} = \left(\frac{d}{\phi_3}\right)^{\phi_4}\ln\left(\frac{d}{\phi_3}\right)$$

and

$$\frac{\partial\left[2^{\phi_4-1}\Gamma(\phi_4)\right]^{-1}}{\partial\phi_4} = -[2_{\phi_4-1}\Gamma(\phi_4)]^{-1}[\ln(2)+\psi(\phi_4)],$$

using $\Gamma'(\phi_4) = \psi(\phi_4)\Gamma(\phi_4)$.

Then, using the product rule and the results above we have for Matérn covariance structure that

$$\frac{\partial r_{ij}}{\partial\phi_4} \;=\; \frac{(d_{ij}/\phi_3)^{\phi_4}}{2^{\phi_4-1}\Gamma(\phi_4)}\left[K_{\phi_4}(\tfrac{d_{ij}}{\phi_3})\left(\log\left(\tfrac{d_{ij}}{\phi_3}\right)-\log(2)-\psi(\phi_4)\right)+\frac{\partial K_{\phi_4}(\frac{d_{ij}}{\phi_3})}{\partial\phi_4}\right],$$

B.4 Factor for Bartlett correction

The factor for Bartlett correction factor is obtained for the Gaussian spatial linear models with multiple repetitions. The vector of parameters is $\boldsymbol{\vartheta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\xi}^\top, \boldsymbol{\phi}^\top)^\top$, where $\boldsymbol{\vartheta}_r$ is the rth element of $\boldsymbol{\vartheta}$. We adopt the the tensor notation for the log-likelihood cumulants: $\lambda_{rs} = \mathrm{E}(\partial^2\mathcal{L}(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}_r\partial\boldsymbol{\vartheta}_s)$, $\lambda_{rst} = \mathrm{E}(\partial^3\mathcal{L}(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}_r\partial\boldsymbol{\vartheta}_s\partial\boldsymbol{\vartheta}_t)$, ..., and the following for the derivatives of cumulants: $(\lambda_{rs})_t = \partial\lambda_{rs}/\partial\vartheta_t$, $(\lambda_{rs})_{tu} = \partial^2\lambda_{rs}/\partial\vartheta_t\partial\vartheta_u$, ... The Fisher's information matrix has elements $-\lambda_{rs}$, where $-\lambda^{rs}$ is the element of its inverse. Additionaly $\tau^{rs} = \lambda^{ra}\lambda^{sb}\sigma_{ab}$, where (σ_{ab}) is the inverse of the matrix λ^{ab} of dimension $q \times q$. We use indices a, b, c, d in reference to the components of $\boldsymbol{\beta}_1$, indices f, g for the components of $\boldsymbol{\xi}$, and indices j, k, l, o for the elements of $\boldsymbol{\phi}$.

Lawley's (1956) formula for the Bartlett correction factor, C, is given by

$$C = \sum_{\boldsymbol{\beta}_1,\boldsymbol{\xi},\boldsymbol{\phi}}(l_{rstu} - l_{rstuvw}) - \sum_{\boldsymbol{\xi},\boldsymbol{\phi}}(l_{rstu} - l_{rstuvw}),$$

with

$$\lambda_{rstu} = \lambda^{rs}\lambda^{tu}\left[\frac{1}{4}\lambda_{rstu} - (\lambda_{rst})_u - (\lambda_{rt})_{su}\right],$$

and

$$\lambda_{rstuvw} = \lambda^{rs}\lambda^{tu}\lambda^{vw}\left\{\lambda_{rtv}\left(\frac{1}{6}\lambda_{suw} - (\lambda_{sw})_u\right) + \left(\frac{1}{4}\lambda_{svw} - (\lambda_{sw})_v\right)\right.$$
$$\left. + (\lambda_{rt})_v(\lambda_{sw})_u + (\lambda_{rt})_u(\lambda_{sw})_v\right\}.$$

The following notation and results will be used to obtain C.

$$\begin{aligned}
\mathbf{z}_i &= \mathbf{Y}_i - \mathbf{X}_1^*\boldsymbol{\beta}_1 - \mathbf{X}_2\boldsymbol{\xi},\\
\frac{\partial \mathbf{X}_1^*}{\partial \boldsymbol{\beta}_1} &= \mathbf{0},\\
\frac{\partial \mathbf{X}_1^*}{\partial \boldsymbol{\xi}} &= \mathbf{0},\\
\frac{\partial \mathbf{X}_1^*}{\partial \phi_j} = \dot{\mathbf{X}}_{1j}^* &= -\mathbf{X}_2(\mathbf{X}_2^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2^\top\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_1^*,\\
\frac{\dot{\mathbf{X}}_{1j}^*}{\partial \phi_k} = \ddot{\mathbf{X}}_{1jk}^* &= 2\mathbf{X}_2(\mathbf{X}_2^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2^\top\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_k}\mathbf{X}_2(\mathbf{X}_2^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2^\top\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_1^*\\
&\quad - \mathbf{X}_2(\mathbf{X}_2^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2^\top\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_j\partial \phi_k}\mathbf{X}_1^*,\\
\dot{\mathbf{X}}_{1j}^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^* &= \mathbf{0},\\
\mathbf{X}_2^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^* &= \mathbf{0},\\
\dot{\mathbf{X}}_{1k}^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_l}\dot{\mathbf{X}}_{1j}^* &= \mathbf{0},
\end{aligned}$$

The first-order derivatives of the log-likelihood function in Equation (3.9) are

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1} &= \sum_{i=1}^{r}\mathbf{X}_1^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{z}_i,\\
\frac{\partial \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\xi}} &= \sum_{i=1}^{r}\mathbf{X}_2^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{z}_i,\\
\frac{\partial \mathcal{L}(\boldsymbol{\vartheta})}{\partial \phi_j} &= -\frac{r}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_j}\right) + \sum_{i=1}^{r}\boldsymbol{\beta}_1^\top\dot{\mathbf{X}}_{1j}^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{z}_i - \frac{1}{2}\sum_{i=1}^{r}\mathbf{z}_i^\top\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{z}_i.
\end{aligned}$$

Note that when r is not a subindex, it is the number of repetitions. The second-order derivatives are

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} &= -r\mathbf{X}_1^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_1^*, \\
\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \xi} &= \mathbf{0}, \\
\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \phi_j} &= \sum_{i=1}^{r} \dot{\mathbf{X}}_{1j}^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{z}_i + -\sum_{i=1}^{r} \mathbf{X}_1^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{z}_i, \\
\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \xi_1 \partial \xi_1} &= -r\mathbf{X}_2^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_2, \\
\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \xi_1 \partial \phi_j} &= \sum_{i=1}^{r} \mathbf{X}_2^{\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}(\mathbf{Y}_i - \mathbf{X}_2\boldsymbol{\xi}), \\
\frac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \phi_j \partial \phi_k} &= -\frac{r}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_k}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_j}\right) - \frac{r}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}^2}{\partial \phi_j \partial \phi_k}\right) + \sum_{i=1}^{r}\boldsymbol{\beta}_1^{\top}\ddot{\mathbf{X}}_{1jk}^{*\top}\boldsymbol{\Sigma}^{-1}\mathbf{z}_i \\
&\quad + \sum_{i=1}^{r}\boldsymbol{\beta}_1^{\top}\dot{\mathbf{X}}_{1j}^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_k}\mathbf{z}_i - \sum_{i=1}^{r}\boldsymbol{\beta}_1^{\top}\dot{\mathbf{X}}_{1j}^{*\top}\boldsymbol{\Sigma}^{-1}\dot{\mathbf{X}}_{1k}^{*}\boldsymbol{\beta}_1 + \sum_{i=1}^{r}\boldsymbol{\beta}_1^{\top}\dot{\mathbf{X}}_{1k}^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{z}_i \\
&\quad - \frac{1}{2}\sum_{i=1}^{r}\mathbf{z}_i^{\top}\frac{\partial \boldsymbol{\Sigma}^2}{\partial \phi_j \partial \phi_k}\mathbf{z}_i.
\end{aligned}
$$

The third-order derivatives are

$$
\begin{aligned}
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^{\top} \partial \boldsymbol{\beta}} &= \mathbf{0}, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^{\top} \partial \xi} &= \mathbf{0}, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^{\top} \partial \phi_j} &= -r\mathbf{X}_1^{*\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_1^*, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\xi}^{\top} \partial \boldsymbol{\xi}} &= \mathbf{0}, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\xi}^{\top} \partial \phi_j} &= \mathbf{0}, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}_1 \partial \phi_j \partial \phi_k} &= \sum_{i=1}^{r}\left(\ddot{\mathbf{X}}_{1jk}^{*}\boldsymbol{\Sigma}^{-1} + \dot{\mathbf{X}}_{1j}^{*}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_k} + \dot{\mathbf{X}}_{1k}^{*}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j} + \mathbf{X}_1^{*\top}\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_j \partial \phi_k}\right)\mathbf{z}_i, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \xi_1 \partial \xi^{\top} \partial \phi_j} &= -r\mathbf{X}_2^{\top}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi_j}\mathbf{X}_2, \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \xi_1 \partial \phi_j \partial \phi_k} &= \sum_{i=1}^{r}\mathbf{X}_2^{\top}\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_j \partial \phi_k}(\mathbf{Y}_i - \mathbf{X}_2\boldsymbol{\xi}), \\
\frac{\partial^3 \mathcal{L}(\boldsymbol{\vartheta})}{\partial \phi_j \partial \phi_k \partial \phi_l} &= -\frac{r}{2}\operatorname{tr}\left(\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_k \partial \phi_l}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_j}\right) - \frac{r}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_k}\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_j \partial \phi_l}\right) - \frac{r}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_l}\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi_j \partial \phi_k}\right)
\end{aligned}
$$

$$-\frac{r}{2}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\frac{\partial^3\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k\partial\phi_l}\right)+\sum_{i=1}^{r}\boldsymbol{\beta}_1^\top\left(\frac{\partial\ddot{\mathbf{X}}_{1jk}^*}{\partial\phi_l}\mathbf{\Sigma}^{-1}+\ddot{\mathbf{X}}_{1jk}^*\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_l}+\right.$$
$$\left.\ddot{\mathbf{X}}_{1jl}^*\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_k}+\mathbf{X}_{1j}^{*\top}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_k\partial\phi_l}+\ddot{\mathbf{X}}_{1kl}^*\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}+\mathbf{X}_{1k}^{*\top}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_l}+\mathbf{X}_{1l}^{*\top}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\right)\mathbf{z}_i$$
$$-\frac{1}{2}\sum_{i=1}^{r}\frac{\partial^3\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k\phi_l}.$$

The third-order derivatives not shown above are equal to $\mathbf{0}$. Finally, the fourth-order derivatives are

$$\frac{\partial^4\mathcal{L}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1^\top\partial\phi_j\partial\phi_k}=-2r\dot{\mathbf{X}}_{1k}^{*\top}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_1^*-r\mathbf{X}_1^{*\top}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\mathbf{X}_1^*,$$
$$\frac{\partial^4\mathcal{L}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1^\top\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^\top}=\mathbf{0},$$
$$\frac{\partial^4\mathcal{L}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1^\top\partial\boldsymbol{\xi}\partial\phi_j}=\mathbf{0}.$$

Taking expected values of second, third and fourth derivatives, we obtain

$$\Lambda_{\beta_1\beta_1}=\mathrm{E}\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}_1\boldsymbol{\beta}_1}\right)=-r\mathbf{X}_1^{*\top}\mathbf{\Sigma}^{-1}\mathbf{X}_1^*,$$

and in similar way we have

$$\Lambda_{\beta_1\xi}=\mathbf{0},$$
$$\Lambda_{\beta_1\phi_j}=\mathbf{0},$$
$$\Lambda_{\xi\xi}=-r\mathbf{X}_2^\top\mathbf{\Sigma}^{-1}\mathbf{X}_2,$$
$$\Lambda_{\xi\phi_j}=r\mathbf{X}_2^\top\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_1^*\boldsymbol{\beta}_1,$$
$$\Lambda_{\beta_1\beta_1\phi_j}=-r\mathbf{X}_1^{*\top}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_1^*,$$
$$\Lambda_{\beta_1\phi_j\phi_k}=\mathbf{0},$$
$$\Lambda_{\xi\xi\phi_j}=-r\mathbf{X}_2^\top\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_2,$$
$$\Lambda_{\xi\phi_j\phi_k}=r\mathbf{X}_2^\top\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\mathbf{X}_1^*\boldsymbol{\beta}_1,$$

$$
\begin{aligned}
\Lambda_{\beta_1\beta_1\xi\xi} &= \mathbf{0}, \\
\Lambda_{\beta_1\beta_1\xi\phi_k} &= \mathbf{0}, \\
\Lambda_{\beta_1\beta_1\phi_j\phi_k} &= -2r\dot{\mathbf{X}}_{1k}^{*\top}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_1^* - r\mathbf{X}_1^{*\top}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\mathbf{X}_1^*, \\
\lambda_{jk} &= -\frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\frac{\partial\mathbf{\Sigma}}{\partial\phi_k}\right) - r\boldsymbol{\beta}_1\dot{\mathbf{X}}_{1j}^{*\top}\mathbf{\Sigma}^{-1}\dot{\mathbf{X}}_{1k}^*\boldsymbol{\beta}_1, \\
\lambda_{ljk} &= -2r\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_l}\frac{\partial\mathbf{\Sigma}}{\partial\phi_k}\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_j}\right) + \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_l\partial\phi_k}\right) + \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_k}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_l\partial\phi_j}\right) \\
&\quad + \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_l}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\right).
\end{aligned}
$$

The expected Fisher's information matrix out of minus is given by

$$
\begin{pmatrix}
\Lambda_{\beta\beta} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \Lambda_{\xi\xi} & \Lambda_{\xi\phi} \\
\mathbf{0} & \Lambda_{\xi\phi}^\top & \Lambda_{\phi\phi}
\end{pmatrix}.
$$

Using inverse matrix rule from Rao (1973), the matrices formed out of minus Fisher's information matrix inverse are: $\Lambda^{\beta\beta} = \Lambda_{\beta\beta}^{-1}$, $\Lambda^{\phi\phi} = (\Lambda_{\phi\phi} - \Lambda_{\xi\phi}^\top\Lambda_{\xi\xi}^{-1}\Lambda_{\xi\phi})^{-1}$, $\Lambda^{\xi\xi} = \Lambda_{\xi\xi}^{-1} + \Lambda_{\xi\xi}^{-1}\Lambda_{\xi\phi}\Lambda^{\phi\phi}\Lambda_{\xi\phi}^\top\Lambda_{\xi\xi}^{-1}$ and $\Lambda^{\xi\phi} = -\Lambda_{\xi\xi}^{-1}\Lambda_{\xi\phi}\Lambda^{\phi\phi\top}$, where the jth column of $\Lambda_{\xi\phi}$ is $\Lambda_{\xi\phi_j}$ and the (j,k)th element of $\Lambda_{\phi\phi}$ is λ_{jk}. The derivatives of cumulants are

$$
\begin{aligned}
(\Lambda_{\beta_1\beta_1})_j &= -r\mathbf{X}_1^{*\top}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_1^*, \\
(\Lambda_{\beta_1\beta_1})_{jk} &= -2r\dot{\mathbf{X}}_{1k}^{*\top}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_1^*\boldsymbol{\beta}_1 - r\mathbf{X}_1^{*\top}\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\mathbf{X}_1^*, \\
(\Lambda_{\xi\xi})_j &= -r\mathbf{X}_2^\top\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_2, \\
(\Lambda_{\xi\phi_j})_k &= r\mathbf{X}_2^\top\frac{\partial^2\mathbf{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\mathbf{X}_1^*\boldsymbol{\beta}_1 + r\mathbf{X}_2^\top\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\dot{\mathbf{X}}_{1k}^*\boldsymbol{\beta}_1, \\
(\lambda_{jl})_k &= -r\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_l}\frac{\partial\mathbf{\Sigma}}{\partial\phi_j}\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_k}\right) + \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_l}\frac{\partial^2\mathbf{\Sigma}}{\partial\phi_j\partial\phi_k}\right) \\
&\quad + \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\phi_j}\frac{\partial^2\mathbf{\Sigma}}{\partial\phi_k\partial\phi_l}\right).
\end{aligned}
$$

From the orthogonality between $\boldsymbol{\beta}_1$ and $(\xi^\top, \boldsymbol{\phi}^\top)^\top$ that $\lambda^{af} = \lambda^{aj} = (\lambda_{af})_{jb} = \lambda_{jfa} = \lambda_{jfab} = 0$. And also it is possible to show that $l_{abcd} = l_{abfg} = l_{abjf} = l_{fgab} = 0$ and that

$\Lambda_{abjk} = (\Lambda_{abj})_k = \Lambda_{jkab}$. After tiresome algebraic manipulations, the Bartlett correction factor reduces to

$$C = \sum \left[-\frac{1}{2}\lambda^{ab}\lambda^{jk}\lambda_{abjk} + \frac{1}{4}\lambda^{ab}\lambda^{cd}\lambda^{jk}\lambda_{abj}\lambda_{cdk} - \frac{1}{2}\lambda^{ab}\lambda^{jk}\lambda^{lo}\lambda_{abj}(\lambda_{lok} - 2(\lambda_{lo})_k) \right.$$
$$\left. + \lambda^{ab}\lambda^{fk}\lambda^{jl}\lambda_{abj}\left(2(\lambda_{fk})_l - \frac{3}{2}\lambda_{fkl}\right) - \frac{1}{2}\lambda^{ab}\lambda^{fg}\lambda^{jk}\lambda_{abj}\lambda_{fgk} \right].$$

And in matrix notation it is given by

$$C = \mathrm{tr}\left\{ \Lambda^{\phi\phi}\left[-\frac{1}{2}\mathbf{M} + \frac{1}{4}\mathbf{P} - \left(\frac{1}{2}\boldsymbol{\varsigma} - \boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\nu}\right)\tau^{\top} \right] \right\},$$

where $\boldsymbol{\varsigma}$, $\boldsymbol{\delta}$and $\boldsymbol{\nu}$are m-vectors whose j-th elements are, respectively, $\mathrm{tr}\left(\Lambda^{\phi\phi}\mathbf{A}_j\right)$, $\mathrm{tr}\left(\Lambda^{\xi\phi^{\top}}\mathbf{B}_j\right)$ and $\mathrm{tr}\left[-\Lambda^{\xi\xi}\left(\mathbf{X}_2^{\top}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\phi_j}\mathbf{X}_2\right)\right]$.

The matrix $\mathbf{B}_j$ contains the m column vectors $\left(\frac{1}{2}\mathbf{X}_2^{\top}\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\phi_j\partial\phi_k}\mathbf{X}_1^{*} + 2\mathbf{X}_2^{\top}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\phi_j}\dot{\mathbf{X}}_{1k}^{*}\right)\boldsymbol{\beta}_1$. This results are similar to the ones obtained by Melo et al. (2009).

Chapter 4

Local influence on Gaussian spatial linear models with repetitions

4.1 Resumo

Neste capítulo, são apresentadas técnicas de diagnósticos de influência local em modelos espaciais lineares Gaussianos com repetições. Conceitos de influência local sob dois enfoques diferentes e alavanca generalizada são revisados. Além disso, foi desenvolvido medidas de diagnóstico sob esquemas de perturbação apropriados para a resposta e ponderação de casos. Uma aplicação a dados reais ilustra a metodologia desenvolvida neste capítulo.

4.2 Introduction

Geostatistical began in South Africa and reached maturity at the *Ecoles des Mines at Fontainebleau* near Paris. Geostatistical data are data collected at known sites in space, from a process that has a value at every site in a certain domain. The data are modelled as the sum of a constant or varying trend and a spatially correlated residual. Given a model for the trend, and under some stationarity assumptions, geostatistical modelling involves the estimation of the spatial correlation, Pebesma (2004).

Recent proposals have discussed Gaussian spatial linear models (GSLM) to study the structure of dependence in spatially referenced data. For more details about estimation, inference methods and applications of these models, see, for example, (Cressie, 1993; Isaaks and Srisvastava, 1989; Mardia and Marshall, 1984; Schabenberger and Gotway, 2005; Waller and Gotway, 2004; Webster and Oliver, 2007).

To assess the effect of small perturbations in the model (or data) on the parameter estimates, Cook (1986) proposed an interesting method, named local influence. This analysis does not involve recomputing the parameter estimates for each case deletion, so it is often computationally simpler. Influence diagnostics based on the likelihood displacement have been developed for multivariate elliptical linear models by Galea et al. (1997) and Liu (2000). Cadigan and Farell (2002) considered local influence diagnostics for a statistical model that is fully parametric, and where estimation involves a fit function that is second order differentiable with respect to the parameters. Zhu et al. (2007) constructed influence measures by assessing local influence of perturbations to a statistical model. Chen and Zhu (2009) proposed a perturbation selection method for selecting an appropriate perturbation with desirable properties then, developed a second-order local influence measure on the basis of the selected perturbation, in the context of general latent variable models. Vasconcellos and Zea Fernandez (2009) presented influence analysis with homogeneous linear restrictions. Giménez and Galea (2013) applied the approach of Zhu et al. (2007) in functional heteroskedasticity measurement error models.

There are only a few works in the literature about influence diagnostics in geostatistical analysis, see (Christensen et al., 1992a; Diamond and Armstrong, 1984; Warnes, 1986). For the situation we have no repetition, Uribe-Opazo et al. (2012) used diagnostic techniques to assess the sensitivity of the maximum likelihood estimators, covariance functions and linear predictor to small perturbations in the data and/or in the Gaussian spatial linear model assumptions. Assumpção et al. (2014) presented the generalized leverage to evaluate the influence of a vector on its own predicted value for different approaches for t-Student spatial linear models. De Bastiani et al. (2015) presented local influence on elliptical spatial linear models under the appropriated perturbation scheme. Garcia-Papani et al. (2016) presented diagnostics tools for the Birbaum-Saunders distribution within the geostatistical framework.

Waternaux et al. (1989) suggested several practical procedures to detect outliers for the repeated measurements model based on the global influence approach. Lesaffre and Verbeke (1998) extended the local influence methodology to normal linear mixed models in repeated-measurement context and under the case-weight perturbation scheme.

Smith (2001) presented the model of multiple repetitions of a spatial process and parameter estimation process. The observations are taken from different experimental units, which is different geographical site for our case, where each variable are observed more than once.

In this work we propose the study of repeated measures in the geostatistical field. We focus on the GSLM with repetitions and we present diagnostic studies. We discuss maximum likelihood estimation and diagnostic tools such as local influence and generalized leverage. Moreover, we consider appropriated perturbation scheme in the response variable and case weight perturbation, proposed by Zhu et al. (2007). The Chapter unfolds as follows. Section 4.3 presents the GSLM with repetitions. In Section 4.4, the maximum likelihood estimators are obtained and in Section 4.5 we present a procedure to select the covariance structure model. In Section 4.6 an explicit expression for the Fisher information matrix is presented. Section 4.7 reviews concepts of local influence for two different approaches and generalized leverage. We discuss the selection of an appropriate perturbation scheme by using the methodology proposed by Zhu et al. (2007) and present the results for the GSLM with repetitions. Section 4.8 contains an application with real data to illustrate the methodology developed in this paper. Finally, Section 4.9 contains some concluding remarks. Calculations are presented in the appendices.

4.3 Gaussian spatial linear model with repetitions

Let $\mathbf{Y} = \mathbf{Y}(\mathbf{s}) = \text{vec}(\mathbf{Y}_1(\mathbf{s}), ..., \mathbf{Y}_r(\mathbf{s}))$ be an $nr \times 1$ random vector of r independently stochastic process of n elements each, that belong to the family of Gaussian distributions and depend on the sites $\mathbf{s}_j \in S \subset \mathbb{R}^2$ for $j = 1, \ldots, n$, where $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)^\top$. The "vec" operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. It is assumed the i-th stochastic process $\mathbf{Y}_i(\mathbf{s}) = \text{vec}(Y_i(\mathbf{s}_1), ..., Y_i(\mathbf{s}_n))$, represents the $n \times 1$ vector, for $i = 1, \ldots, r$, which can be expressed as a linear mixed model by

$$\mathbf{Y}_i(\mathbf{s}) = \boldsymbol{\mu}_i(\mathbf{s}) + \mathbf{b}_i(\mathbf{s}) + \tau_i(\mathbf{s})$$

where, the deterministic term $\boldsymbol{\mu}_i(\mathbf{s})$ is an $n \times 1$ vector, the means of the process $\mathbf{Y}_i(\mathbf{s})$, $\mathbf{b}_i(\mathbf{s})$ and $\boldsymbol{\tau}_i(\mathbf{s})$ are independents and together form the stochastic error, i.e. $\mathbf{b}_i(\mathbf{s})+\boldsymbol{\tau}_i(\mathbf{s}) = \boldsymbol{\epsilon}_i(\mathbf{s})$ is an $n \times 1$ vector of a stationary process with zero mean vector, $\mathrm{E}[\boldsymbol{\epsilon}_i(\mathbf{s})] = \mathbf{0}$, and covariance $\boldsymbol{\Sigma}_i$. The mean vector $\boldsymbol{\mu}_i(\mathbf{s})$ can be written as a spatial linear model by

$$\boldsymbol{\mu}_i(\mathrm{s}) = \mathbf{X}(\mathrm{s})\boldsymbol{\beta},$$

where, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a $p \times 1$ vector of unknown parameters, $\mathbf{X} = \mathbf{X}(\mathbf{s}) = [\mathbf{x}_{j1}(\mathbf{s}) \dots \mathbf{x}_{jp}(\mathbf{s})]$ is an $n \times p$ matrix of p explanatory variables, for $j = 1, \dots, n$, i.e, the design matrix $\mathbf{X}$ is the same for all r repetitions.

Goodal and Mardia (1994) argue that we can remove the effect of time from the data and then we can view the data as independent repeated measurements in space. According Stein (1999), for a space-time process observed at a fixed set of spatial sites at sufficiently distant points in time, it may be reasonable to assume that observations from different times are independent realizations of a random field.

As in Smith (2001), the GSLM for the i-th independent stochastic process, assuming a homogeneous process, can be written in matrix form by

$$\mathbf{Y}_i(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_i(\mathbf{s}). \tag{4.1}$$

for $i = 1, \dots, r$. The covariance matrix $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = [C(s_u, s_v)]$ is an $n \times n$ covariance matrix of $\mathbf{Y}_i(\mathbf{s})$ for the i-th repetition, $i = 1, \dots, r$. The matrix $\boldsymbol{\Sigma}$ is non-singular, symmetric and positive defined, associated to the vector $\mathbf{Y}_i(\mathbf{s})$, where for the stationary and isotropic process, the elements $C(\mathbf{s}_u, \mathbf{s}_v)$ depend on the Euclidean distance $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$ between points $\mathbf{s}_u$ and $\mathbf{s}_v$.

So, from what was previously written we have that $\mathbf{Y} \sim N_{nr}(\mathbf{1}_r \otimes \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_r \otimes \boldsymbol{\Sigma})$, with probability density function (pdf) given by

$$\begin{aligned} f(\mathbf{Y}, \boldsymbol{\theta}) &= \textstyle\prod_{i=1}^r f(\mathbf{Y}_i, \boldsymbol{\theta}) \\ &= \textstyle\prod_{i=1}^r (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})\right], \end{aligned} \tag{4.2}$$

where $\otimes$ denote the kronecker product and $(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}) = \delta_i$ is the Mahalanobis distance.

The covariance matrix $\boldsymbol{\Sigma}$ has a structure which depends on parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$ as given in Equation (4.3) (Mardia and Marshall, 1984; Uribe-Opazo et al., 2012):

$$\boldsymbol{\Sigma} = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}, \tag{4.3}$$

where, $\phi_1 \geq 0$ is the parameter known as nugget effect; $\phi_2 \geq 0$ is known for sill ; $\mathbf{R} = \mathbf{R}(\phi_3, \phi_4) = [(r_{uv})]$ or $\mathbf{R} = \mathbf{R}(\phi_3) = [(r_{uv})]$ is an $n \times n$ symmetric matrix, which is function of $\phi_3 > 0$, and sometimes also function of $\phi_4 > 0$, with diagonal elements $r_{uu} = 1, (u = 1, \ldots, n)$; $r_{uv} = \phi_2^{-1} C(s_u, s_v)$ for $\phi_2 \neq 0$, and $r_{uv} = 0$ for $\phi_2 = 0$, $u \neq v = 1, \ldots, n$, where r_{uv} depends on the Euclidean distance $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$ between points $\mathbf{s}_u$ and $\mathbf{s}_v$; ϕ_3 is a function of the model range (a), ϕ_4 when exists is known as the smoothness parameter, and $\mathbf{I}_n$ is an $n \times n$ identity matrix.

The Matérn (Matérn, 1960; Minasny and McBratney, 2005; Stein, 1999) is a covariance function particularly attractive given by

$$C(d_{uv}) = \begin{cases} \dfrac{\phi_2}{2^{\phi_4 - 1}\Gamma(\phi_4)} \left(d_{uv}/\phi_3\right)^{\phi_4} K_{\phi_4}\left(d_{uv}/\phi_3\right), & d_{uv} > 0, \\ \phi_1 + \phi_2, & d_{uv} = 0, \end{cases}$$

where the parameters are assumed to be non-negative, i.e., $\phi_q \geq 0$, for $q = 1, 2, 3, 4$; $K_{\phi_4}(u) = \frac{1}{2}\int_0^\infty x^{\phi_4 - 1} e^{-\frac{1}{2}u(x + x^{-1})} dx$ is the modified Bessel function of the third kind of order ϕ_4; see (Gradshteyn and Ryzhik, 2000). This covariance function is particularly attractive because its behavior can change according ϕ_4 value. For instance, $\phi_4 = 0.5$ gives the exponential covariance structure, Whittle's elementary covariance ($\phi_4 = 1$), and in the limit as ϕ_4 approaches to infinity, with ϕ_3 approaching to 0 in such way that $2\phi_4^{1/2}\phi_3$ remains constant, it approaches the Gaussian covariance function. For a finite value of the ϕ_3 parameter (which is function of the range parameter), the Matérn function represents several bounded models. In many works the notation for the smoothness parameter is κ or ν for Matérn class, but for simplicity of notation we use ϕ_4, since it is a parameter part of the covariance structure which can be estimated or considered as fixed.

4.4 Maximum likelihood estimation

According to Smith (2001), despite of disadvantages such as non-robustness when there are outliers in the data, or the possible multimodality of the likelihood surface, they are not reason to abandon maximum likelihood (ML) estimation. Of course, we can expect to get better estimates if there are repetitions of $\mathbf{Y}_i$. The maximum likelihood procedure in this case is only slightly different from that in the single case.

The log-likelihood for the GLSM for the r independent repetitions is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{r} \mathcal{L}_i(\boldsymbol{\theta}),$$

where,

$$\mathcal{L}_i(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}). \tag{4.4}$$

The score functions are given by

$$\begin{aligned}
\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{r} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}_i, \\
\mathbf{U}(\boldsymbol{\phi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} &= -\frac{r}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \phi} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\sum_{i=1}^{r} \frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top \boldsymbol{\Sigma}^{-1}),
\end{aligned}$$

where $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}$. From the solution of the score function of $\boldsymbol{\beta}$, $\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$, the maximum likelihood estimator $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}}, \tag{4.5}$$

where $\bar{\mathbf{Y}} = (\bar{\mathbf{Y}}_1, \ldots, \bar{\mathbf{Y}}_n)^\top$, with $\bar{\mathbf{Y}}_j = \frac{1}{r}\sum_{i=1}^{r} Y_i(\mathbf{s}_j)$, $j = 1, \ldots, n$.

Since we can see the GSLM with repetitions as a linear mixed model we can follow what Demidenko (2004) and Wang and Heckman (2009) said about parameter identifiability problems for the linear mixed models. In particular, Wang and Heckman (2009) showed that there is no problem since we have that $\operatorname{Var}[\boldsymbol{\epsilon}_i] = \phi_1 \mathbf{I}_n$. But one deficiency can be the sensibility of the maximum likelihood estimates for atypical observations. For that we present the study of diagnostic techniques in Section 4.7.

Unfortunately, the score equations for $\boldsymbol{\phi}$ do not lead to a closed-form solution for $\boldsymbol{\phi}$. Thus, a common practice is to maximize the concentrated log-likelihood, obtained substituting the solution to the score equation for $\boldsymbol{\beta}$ [Equation (4.5)] into the log-likelihood in Equation (4.4). The concentrated log-likelihood then depends only on $\boldsymbol{\phi}$, and ignoring the constant $-\frac{rn}{2}\log(2\pi)$ it is given by

$$\begin{aligned}\mathcal{L}^*(\boldsymbol{\phi}) \propto & -\frac{r}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{r}(\mathbf{Y}_i - \mathbf{X}(\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}})^\top\boldsymbol{\Sigma}^{-1} \\ & \times(\mathbf{Y}_i - \mathbf{X}(\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}}).\end{aligned}$$

The package `geoR` (Ribeiro Jr and Diggle, 2001) from software `R` has already implemented procedures for the parameters estimation. The ML estimates of covariance parameters are typically consistent and asymptotically normal when fitted by a correct model (Mardia and Marshall, 1984). The readers are refereed to Stein (1999) for more details regarding fixed domain asymptotics. Some discussion concerning which asymptotic framework is more appropriate can also be found in Zhang and Zimmerman (2005).

4.5 Covariance Structure Selection

For the case of the Matérn class of covariance function (or other with more than three parameters), we suggest to use the idea presented in Minasny and McBratney (2005) for one single realization, which assume ϕ_4 as fixed to find which ϕ_4 gives the maximum value for the profile log-likelihood, and then use it as a initial value and estimate it. We use the same idea, but now for a model with multiple repetitions. Stein (1999) says that although he do not advocate treating $\phi_4 = 0.5$ as fixed, it is important to keep in mind that is the same of using the exponential model.

In the case that we consider the parameter ϕ_4 as fixed, we have to choose which value, between a range of values, gives the best fit. Despite of the maximum value of the log-likelihood, we consider the cross validation criterion similar to the one presented in

(De Bastiani et al., 2015), defined by

$$CV = CV(\phi_4) = \frac{1}{nr}\sum_{i=1}^{r}\sum_{j=1}^{n}\left\{\frac{y_i(\mathbf{s}_j) - \hat{y}_{i(j)}(\mathbf{s}_j)}{1-h_{jj}}\right\}^2,$$

where $\hat{y}_{i(j)}(\mathbf{s}_j) = \mathbf{x}_j^\top \hat{\boldsymbol{\beta}}_{(j)}$, with $\mathbf{x}_j^\top$ the i-th row of the matrix $\mathbf{X}$, is the prediction in the site $\mathbf{s}_j$ without considering this observation, $(y_j, \mathbf{x}_j^\top)$, $\hat{\boldsymbol{\beta}}_{(j)}$ is the ML estimator of $\boldsymbol{\beta}$ without considering the j-th observation, and h_{jj} is the j-th diagonal element of the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}$, for $j = 1, \ldots, n$, see Section 4.7.4. To simplify the calculations we can use the approximation $\hat{\boldsymbol{\beta}}_{(j)} \approx \hat{\boldsymbol{\beta}} + \mathbf{K}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{U}_{(j)}(\hat{\boldsymbol{\beta}})$, where

$$\mathbf{K}(\hat{\boldsymbol{\beta}}) = r\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} \quad \text{and}$$

$$\mathbf{U}_{(j)}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{r}\mathbf{X}_{(j)}^\top\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_{i(j)} - \mathbf{X}_{(j)}\hat{\boldsymbol{\beta}}), \tag{4.6}$$

with $\mathbf{X}_{(j)}$, $(n-1)\times p$, the matrix $\mathbf{X}$ without the j-th row, $\mathbf{x}_j^\top$, and $\mathbf{Y}_{i(j)}$, $(n-1)\times 1$ denotes the vector $\mathbf{Y}_i$ without the response $y_i(\mathbf{s}_j)$, for $j = 1, \ldots, n$ and $i = 1 \ldots, r$. Note that in (4.6) the matrix $\hat{\boldsymbol{\Sigma}}^{-1}$ is of order $(n-1)\times(n-1)$.

Alternatively, another criterion is to use the trace of the asymptotic covariance matrix of an estimated mean as a criterion in selecting a better model (Kano et al., 1993). Let $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. From the results of Section 4.6, it follows that the trace of the asymptotic covariance matrix of $\boldsymbol{\mu}$ is given by,

$$\operatorname{tr}(\phi_4) = (1/r)\operatorname{tr}[(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}].$$

4.6 Asymptotic standard error estimation

The observed information matrix, say $I(\boldsymbol{\theta})$, is $I(\boldsymbol{\theta}) = -\mathbf{L}(\boldsymbol{\theta})$, evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where $\mathbf{L}(\boldsymbol{\theta}) = \partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. Furthermore, we notice that the $I(\boldsymbol{\theta})$, which is evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, has the same equations as the expected information matrix given below. The derivatives of first and second-order of the scale matrix $\boldsymbol{\Sigma}$, with respect to ϕ_q, for some covariance functions are presented in Uribe-Opazo et al. (2012). Some useful constructions are given in Appendix B.1.

Asymptotic standard errors can be calculated by inverting either the observed information matrix or the expected information matrix. The expected information matrix, $\mathbf{F}(\boldsymbol{\theta})$, is given by, (Waller and Gotway, 2004)

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{F} = \begin{pmatrix} \mathrm{E}(\mathbf{U}_\beta \mathbf{U}_\beta^\top) & \mathrm{E}(\mathbf{U}_\beta \mathbf{U}_\phi^\top) \\ \mathrm{E}(\mathbf{U}_\phi \mathbf{U}_\beta^\top) & \mathrm{E}(\mathbf{U}_\phi \mathbf{U}_\phi^\top) \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{\beta\beta} & \mathbf{F}_{\beta\phi} \\ \mathbf{F}_{\phi\beta} & \mathbf{F}_{\phi\phi} \end{pmatrix},$$

where $\mathbf{F}_{\beta\beta} = r\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\mathbf{F}_{\beta\phi} = \mathbf{0}$, $\mathbf{F}_{\phi\beta} = \mathbf{0}$ and $\mathbf{F}_{\phi\phi} = \dfrac{r}{2}\dfrac{\partial\,\mathrm{vec}^\top(\boldsymbol{\Sigma})}{\partial\boldsymbol{\phi}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\dfrac{\partial\,\mathrm{vec}(\boldsymbol{\Sigma})}{\partial\boldsymbol{\phi}^\top}$.

See Appendix B.2 for the details. We used $\mathbf{F}_{\beta\beta}^{-1}$ and $\mathbf{F}_{\phi\phi}^{-1}$ to estimate the dispersion matrices for the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\phi}}$, respectively.

4.7 Local Influence

Detecting influential observations is an important step in the analysis of a data set. There are different approaches to assess the influence of perturbations in a data set and in the model given the estimated parameters. Cook (1977) gave a starting point to the development of case-deletion diagnostics for a broad class of statistical models, that is to asses the effect of an observation by completely removing it. Cook (1986) presented the local influence approach, that is, a weight ω_i is given for each case and the effect on the parameter estimation is measured by perturbing around these weights. Choosing weights equal zero or one corresponds to the global case-deletion approach. In general, perturbation measures do not depend on the data directly, but rather on its structure via the model.

Lesaffre and Verbeke (1998) extended the local influence methodology to normal linear mixed models in repeated-measurement context and under the case-weight perturbation scheme. Cerioli and Riani (1999), Militino et al. (2006) showed that case deletion diagnostics do suffer from masking and suggest robust procedures based on subsets of data free from outliers. Borssoi et al. (2011) and Uribe-Opazo et al. (2012) discussed diagnostic techniques, using local influence methodology, to evaluate the sensitivity of maximum likelihood estimators, the covariance functions and the linear predictor under small perturbations in the data and/or spatial linear model with normal distribution.

The method of local influence was introduced by Cook (1986) as a general tool for assessing the influence of local departures from the assumptions underlying the statistical models. A perturbation scheme is introduced into the postulated model through a perturbation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)^\top$ ($\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^k$), generating the perturbed model $\mathcal{M} = \{f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$, where $f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega})$ is the pdf of $\mathbf{Y}$ given in (4.2), perturbed by $\boldsymbol{\omega}$ and $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \log f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega})$ is the corresponding loglikelihood function. Then, an influence measure is constructed using the basic geometric idea of curvature of the likelihood displacement given by

$$LD(\boldsymbol{\omega}) = 2[\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_\omega)],$$

where $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ in the postulated model, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$ and $\hat{\boldsymbol{\theta}}_\omega$ is the ML estimator of $\boldsymbol{\theta}$ in the perturbed model $\mathcal{M}$.

Cook (1986) proposed to study the local behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$ and shows that the normal curvature C_l of $LD(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\boldsymbol{l}$, is given by $C_l = C_l(\boldsymbol{\theta}) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|$, with $||\boldsymbol{l}|| = 1$, where $-\mathbf{L} = -\mathbf{L}(\boldsymbol{\theta})$ is the observed information matrix (given in Section 3.6), evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\phi^\top)^\top$, where $\boldsymbol{\Delta}_\beta = \partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top$ and $\boldsymbol{\Delta}_\phi = \partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\omega}^\top$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, where $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})$ is the perturbed log-likelihood and $\boldsymbol{\omega}_0$ is the non perturbation vector.

Zhu et al. (2007) developed a rigorous differential-geometrical framework of a perturbation model. First and second-order influence measures based on the observed data likelihood function of the related statistical model are proposed in order to assess the local influence of small perturbations on any statistics of interest and the important issue about the selection of and appropriate perturbation vector is also addressed. Essentially, to verify if the perturbation scheme is appropriated, Zhu et al. (2007) proposed to use the

Fisher information matrix of $\boldsymbol{\omega}$ in the model $\mathcal{M}$ considering the vector $\boldsymbol{\theta}$ as fixed. Below we briefly describe the methodology.

Let $\mathbf{G}(\boldsymbol{\omega})$ be the Fisher information matrix with respect to the perturbation vector $\boldsymbol{\omega}$. That is, $\mathbf{G}(\boldsymbol{\omega}) = \mathrm{E}_\omega[\mathbf{U}(\boldsymbol{\omega})\mathbf{U}^\top(\boldsymbol{\omega})]$, where E_ω denotes the expectation with respect to $f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega})$, and $\mathbf{U}(\boldsymbol{\omega})$ the perturbed score function. A perturbation $\boldsymbol{\omega}$ is appropriate if it satisfies $\mathbf{G}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$, where $c > 0$. According to Zhu et al. (2007), see also (Giménez and Galea, 2013), this condition determines orthogonality between the different components of $\boldsymbol{\omega}$, ensures the amounts of perturbation introduced by the components of $\boldsymbol{\omega}$ are uniform and avoids redundant components of $\boldsymbol{\omega}$. In Sections 4.7.2 and 4.7.3 we present the appropriate perturbation for case weight perturbation and perturbation of the response variable.

The plot of the elements $|l_{max}|$ versus i (order of data) can reveal what type of perturbation has more influence on $LD(\boldsymbol{\omega})$, in the neighborhood of $\boldsymbol{\omega}_0$, Cook (1986). Even considering $C_i = 2|j_{ii}|$, where j_{ii} are the elements of the main diagonal of the matrix $\mathbf{J} = \boldsymbol{\Delta}^\top\mathbf{L}^{-1}\boldsymbol{\Delta}$, can be used the index plot of C_i to evaluate the presence of influential observations.

Since C_l is not invariant under uniform change of scale, Poon and Poon (1999) proposed the conformal normal curvature $B_l = C_l/\operatorname{tr}(2\mathbf{J})$, see (Zhu and Lee, 2001). An interesting property of conformal curvature is that for any direction unit $\boldsymbol{l}$, it follows that $0 \leq B_l \leq 1$. We denote by $B_i = 2|j_{ii}|/\operatorname{tr}(2\mathbf{J})$ the conformal curvature in the unit direction with i-th entry 1 and all other entries 0. According to Zhu and Lee (2001), the i-th observation is potentially influential if $B_i > \bar{B} + 2\mathrm{sd}(B)$, where $\bar{B} = \sum_{i=1}^n B_i/n$ and $\mathrm{sd}(B)$ is the standard deviation of $B_1, \ldots, B_n$.

4.7.1 Influence measures based on the likelihood function

Since the log-likelihood is used in the estimation process for inference and we are using it as a criterion to select the covariance matrix structure that best describe the spatial dependence, it is of concern to analyze the sensitivity of this statistic to a particular perturbation scheme. More details can be seen in (Billor and Loynes, 1993; Cadigan and Farell, 2002; Tsai, 1986; Zhu et al., 2007). In fact, following Cadigan and Farell (2002) notation, let $g(\hat{\boldsymbol{\theta}})$ be the objective function. We are interested in evaluate the influence of the perturbed objective function $g_\omega = g_\omega(\hat{\boldsymbol{\theta}}_\omega)$. The first order influence of a perturbation scheme $\boldsymbol{\omega}$ is measured through the slope in the direction d given by, (Cadigan and Farell, 2002), $S(d) = d^\top \dot{g}_0$, where

$$\dot{g}_0 = \frac{\partial g_\omega(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\omega}}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} - \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

The maximum slope $s_{max} = \sqrt{\dot{g}_0^\top \dot{g}_0}$ and the corresponding direction vector $d_{max} = \dot{g}_0 / s_{max}$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, are useful to detect local influence.

4.7.2 Case weight perturbation

The case weights perturbation scheme is often the basis of the study of influence. A perturbed log-likelihood function, allowing different weights for each repetition can be defined as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{r} \omega_i \mathcal{L}_i(\boldsymbol{\theta}),$$

and $\mathcal{L}_i(\boldsymbol{\theta})$ is the contribution of the i-th repetition of the log-likelihood and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_r)^\top$ is the vector of perturbation which is assumed to belong to an open subset Ω of $\mathbb{R}^r$, with $0 \leq \omega_i \leq 1$. For this case we have that $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ is the non-perturbed vector, such that $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \mathcal{L}(\boldsymbol{\theta})$ and $f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega}_0) = f(\mathbf{Y}, \boldsymbol{\theta})$ for all $\boldsymbol{\theta}$.

The density of the perturbed model is given by

$$f(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^{r} (2\pi)^{-n/2} |\omega_i^{-1} \boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{\omega_i}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})\right].$$

Then, the i-th contribution of the perturbed log-likelihood is given by

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\theta}|\omega_i) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}\omega_i^{-1}| - \frac{\omega_i}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}) \\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \frac{n}{2}\log\omega_i - \frac{1}{2}\omega_i(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}).
\end{aligned}$$

As showed by Zhu et al. (2007), for the case weight pertubation scheme, in general $\mathbf{G}(\boldsymbol{\omega}) = \text{diag}\{Var_\omega[\mathcal{L}_1(\boldsymbol{\theta}), \ldots, \mathcal{L}_r(\boldsymbol{\theta})]\}$ is of the form $\mathbf{G}(\omega_0) = c\mathbf{I}_r$ when we have the same number of observation for each repetition. In our case we have that

$$
\begin{aligned}
\mathrm{Var}_{\omega_0}[\mathcal{L}_i(\boldsymbol{\theta})] &= \mathrm{Var}_{\omega_0}\left[-\frac{n}{2}-\frac{1}{2}\log|\boldsymbol{\Sigma}|-\frac{1}{2}(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})\right] \\
&= \frac{1}{4}Var_{\omega_0}\left[(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})\right] = \frac{n}{2},
\end{aligned}
$$

because $(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})$ has $\chi^2_{(n)}$ distribution.

The model given in (4.1) is assumed to be homoscedastic, that is, the covariance matrix of the random errors is assumed to be the same, $\boldsymbol{\Sigma}$. If we consider the perturbation $\omega_i^{-1}\boldsymbol{\Sigma}$, for $i=1,\ldots,r$, we can analyze the sensitivity of the log-likelihood estimates under a possibility of heteroskedasticity in the model. The log-likelihood function of the perturbed model and the metric matrix $\mathbf{G}(\boldsymbol{\omega}_0)$ take the same form as for case weights perturbation. Thus, the perturbation of the covariance matrix is an appropriate too.

Based on the likelihood displacement

Considering the appropriated perturbation scheme, for the likelihood displacement methodology $\boldsymbol{\Delta}=(\boldsymbol{\Delta}_\beta^\top,\boldsymbol{\Delta}_\phi^\top)^\top$, where $\boldsymbol{\Delta}_\beta=(\boldsymbol{\Delta}_{1\beta}^\top,\ldots,\boldsymbol{\Delta}_{r\beta}^\top)^\top$ with dimension $p\times r$ and $\boldsymbol{\Delta}_\phi=(\boldsymbol{\Delta}_{1\phi}^\top,\ldots,\boldsymbol{\Delta}_{r\phi}^\top)$ with dimension $3\times r$, where

$$
\boldsymbol{\Delta}_\beta=\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\quad\text{and}\quad\boldsymbol{\Delta}_\phi=\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\omega}^\top},
$$

with elements

$$
\boldsymbol{\Delta}_{i\beta}=\frac{\partial^2\mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\omega_i}=\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\epsilon}_i\quad\text{and}
$$

$$
\boldsymbol{\Delta}_{i\phi}=\frac{\partial^2\mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\phi}\partial\omega_i}=\frac{1}{2}\frac{\partial\,\mathrm{vec}^\top(\hat{\boldsymbol{\Sigma}})}{\partial\boldsymbol{\phi}}\,\mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i^\top\hat{\boldsymbol{\Sigma}}^{-1}\right),
$$

where $\hat{\boldsymbol{\epsilon}}_i=(\mathbf{Y}_i-\mathbf{X}\hat{\boldsymbol{\beta}})$, evaluated at $\boldsymbol{\omega}=\boldsymbol{\omega}_0$ and $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, for $i=1,\ldots,n$.

Based on the likelihood function

Let $g(\hat{\boldsymbol{\theta}})=\mathcal{L}(\hat{\boldsymbol{\theta}})$ be the objective function and $g_\omega(\hat{\boldsymbol{\theta}}_\omega)=\sum_{i=1}^{r}\omega_i\mathcal{L}_i(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})$ be the perturbed objective function. In this case, $\left.\frac{\partial g_\omega(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}=\left[\mathcal{L}_1(\hat{\boldsymbol{\theta}}),\ldots,\mathcal{L}_r(\hat{\boldsymbol{\theta}})\right]^\top=\dot{g}_0$, then $d_{max}=\dot{g}_0/\sqrt{\dot{g}_0^\top\dot{g}_0}$. Notice that $\dot{g}_0^\top\dot{g}_0=\sum_{i=1}^{r}\mathcal{L}_i^2(\hat{\boldsymbol{\theta}})$ such that the i-th component of d_{max}

in absolut value is $|d_{imax}| = |\mathcal{L}_i(\hat{\boldsymbol{\theta}})|/\sqrt{\sum_{k=1}^{r}\mathcal{L}_k^2(\hat{\boldsymbol{\theta}})}$, for $i = 1, ..., r$. As a result, $|d_{imax}|$ can be interpreted as a measure of the relative importance of the i-th contribution to the total likelihood, which in turn is very simple to calculate.

4.7.3 Perturbation on the response variable

Let us consider as a perturbation scheme the model shift in mean, i.e. $\mathbf{Y}_i = \boldsymbol{\mu}(\boldsymbol{\omega}_i) + \boldsymbol{\epsilon}_i$, with $\boldsymbol{\mu}(\boldsymbol{\omega}_i) = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\omega}_i$ where $\mathbf{A}$, $n \times n$, is a matrix that does not depend on $\boldsymbol{\beta}$ or on $\boldsymbol{\omega}_i$, for $i = 1, \ldots, r$. In this case $\boldsymbol{\omega}_0 = \mathbf{0}$. Equivalently we can write $\mathbf{Y}_{i\omega} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, with $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + (-1)\mathbf{A}\boldsymbol{\omega}_i$, that corresponds to a perturbation scheme of the response vector.

In this case we have that $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{r} \mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})$ where $\mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})$ is given by

$$\mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\delta_{i\omega}, \tag{4.7}$$

where $\delta_{i\omega} = [\mathbf{Y}_i - \boldsymbol{\mu}(\boldsymbol{\omega}_i)]^\top \boldsymbol{\Sigma}^{-1}[\mathbf{Y}_i - \boldsymbol{\mu}(\boldsymbol{\omega}_i)]$.

To select an adequate matrix $\mathbf{A}$ we can use the methodology proposed by Zhu et al. (2007). In effect, the i-th score function for $\boldsymbol{\omega}$ in the perturbed log-likelihood function (4.7) is given by

$$\mathbf{U}_i(\boldsymbol{\omega}) = \frac{\partial \mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega}_i)}{\partial \boldsymbol{\omega}_i} = \mathbf{A}^\top \boldsymbol{\Sigma}^{-1}[\mathbf{Y}_i - \boldsymbol{\mu}(\boldsymbol{\omega}_i)],$$

for $i = 1, \ldots, r$. Let $\mathbf{G}(\boldsymbol{\omega}) = E_\omega[\mathbf{U}(\boldsymbol{\omega})\mathbf{U}^\top(\boldsymbol{\omega})] = \text{diag}[\mathbf{g}_{11}(\boldsymbol{\omega}_1), \ldots, \mathbf{g}_{rr}(\boldsymbol{\omega}_r)]$ be the Fisher information matrix with respect to the perturbation vector $\boldsymbol{\omega}$. That is, $\mathbf{g}_{ii}(\boldsymbol{\omega}_i) = E_{\omega_i}[\mathbf{U}_i(\boldsymbol{\omega}_i)\mathbf{U}_i^\top(\boldsymbol{\omega}_i)]$. A perturbation $\boldsymbol{\omega}_i$ is appropriate if it satisfies $\mathbf{g}_{ii}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$, where $c > 0$, for $i = 1, \ldots, r$. In our case we have $\mathbf{g}_{ii}(\boldsymbol{\omega}_{i0}) = c\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}$, with $c = 1$. Notice that usually $\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A} \neq \mathbf{I}_n$. However if $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$, then $\mathbf{g}_{ii}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$ and so $\boldsymbol{\mu}(\boldsymbol{\omega}_i) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\omega}_i$ is a perturbation scheme appropriate for $i = 1, \ldots, r$, as shown in De Bastiani et al. (2015) and Borssoi (2014).

Based on the likelihood displacement

Considering the appropriated perturbation scheme, for the likelihood displacement methodology, where $\boldsymbol{\Delta}_\beta = (\boldsymbol{\Delta}_{1\beta}^\top, \ldots, \boldsymbol{\Delta}_{r\beta}^\top)$ is an $p \times nr$ matrix, with $\boldsymbol{\Delta}_{i\beta}$ an $p \times n$ matrix, and $\boldsymbol{\Delta}_{\phi} = (\boldsymbol{\Delta}_{i\phi}^{\top}, \ldots, \boldsymbol{\Delta}_{i\phi}^{\top})$ is an $3 \times nr$ matrix, with $\boldsymbol{\Delta}_{i\phi}$ an $3 \times n$ matrix. We obtain

$$\boldsymbol{\Delta}_{i\beta} = \frac{\partial^2 \mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^{\top}} = -\mathbf{X}^{\top} \hat{\boldsymbol{\Sigma}}^{-1/2} \quad \text{and}$$

$$\boldsymbol{\Delta}_{i\phi} = \frac{\partial^2 \mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\omega}_i^{\top}} = -\sum_{i=1}^{r} \frac{\partial \operatorname{vec}^{\top}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1/2}) \operatorname{vec}(\boldsymbol{\epsilon}_i \otimes \mathbf{1}^{\top})$$

with elements

$$\boldsymbol{\Delta}_{i\boldsymbol{\phi}_j} = \frac{\partial^2 \mathcal{L}_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\phi}_j \partial \omega_i^{\top}} = \hat{\boldsymbol{\epsilon}}_i^{\top} \left(\hat{\boldsymbol{\Sigma}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}^{1/2}}{\partial \phi_j} - \hat{\boldsymbol{\Sigma}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}}{\partial \phi_j} \hat{\boldsymbol{\Sigma}}^{-1/2} \right),$$

evaluated in $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\epsilon}}_i = (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}})$ and $\mathbf{1}$ is an $n \times 1$ vector of ones, for $j = 1, 2, 3$ and $i = 1, \ldots, r$.

Based on the likelihood function

Let $g(\hat{\boldsymbol{\theta}}) = \mathcal{L}(\hat{\boldsymbol{\theta}})$ be the objective function and $g_{\omega}(\hat{\boldsymbol{\theta}}_{\omega}) = \sum_{i=1}^{r} \mathcal{L}_i(\hat{\boldsymbol{\theta}}_{\omega}|\boldsymbol{\omega})$ the perturbed function. For this case

$$\frac{\partial g_{\omega}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\omega}}\bigg|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \dot{g}_0 = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{\boldsymbol{\epsilon}}_1 \\ \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{\boldsymbol{\epsilon}}_2 \\ \vdots \\ \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{\boldsymbol{\epsilon}}_r \end{pmatrix}$$

and $\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0}$, so $\dot{g}_0 = \frac{\partial g_{\omega}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\omega}}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}$ and $d_{max} = \dot{g}_0 / \sqrt{\dot{g}_0^{\top} \dot{g}_0}$, with $\dot{g}_0^{\top} \dot{g}_0 = \sum_{i=1}^{r} \hat{\boldsymbol{\epsilon}}_i^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\epsilon}}_i$.

4.7.4 Generalized Leverage

The general concept of a generalized leverage is related to a certain value observed y_j, over the corresponding fitted value $\hat{y}_j$, see for example (Hoaglin and Welsh, 1978; Ross, 1987; St. Laurent and Cook, 1992).

The generalized leverage is defined by $\mathbf{GL} = [(\partial \hat{\mathbf{Y}}_i / \partial \mathbf{Y}_k^{\top})] = [(\mathbf{GL}_{ik})]$, where $\mathbf{GL}_{ik}$ is an $n \times n$ matrix, for $i, k = 1, \ldots, r$, where $\hat{\mathbf{Y}} = \mathbf{1} \otimes \hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. According to Wei et al. (1998), it can also been written by

$$\mathbf{GL} = \mathbf{D}_{\boldsymbol{\theta}}[-\mathbf{L}(\boldsymbol{\theta})]^{-1}\mathbf{L}_{\boldsymbol{\theta}\mathbf{Y}},$$

at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where under the model defined in (4.1) $\mathbf{D}_{\boldsymbol{\theta}} = \partial(\mathbf{1}_r \otimes \boldsymbol{\mu})/\partial\boldsymbol{\theta}^\top = \mathbf{1}_r \otimes [\mathbf{X}\ \mathbf{0}_{n\times 3}]$, the observed information matrix $-\mathbf{L}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ coincides to the expected information matrix, i.e., $-\mathbf{L}(\hat{\boldsymbol{\theta}}) = \mathbf{F}(\boldsymbol{\theta})$ given in Section 3.6, and $\mathbf{L}_{\boldsymbol{\theta}\mathbf{Y}} = \partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\mathbf{Y}^\top = (\mathbf{L}_{\boldsymbol{\theta}\mathbf{Y}_1}, \ldots, \mathbf{L}_{\boldsymbol{\theta}\mathbf{Y}_r})^\top$ with $\mathbf{L}_{\boldsymbol{\theta}\mathbf{Y}_i} = (\mathbf{L}_{\boldsymbol{\beta}\mathbf{Y}_i}, \mathbf{L}_{\boldsymbol{\phi}\mathbf{Y}_i})^\top$ where

$$\begin{aligned} \mathbf{L}_{\boldsymbol{\beta}\mathbf{Y}_i} &= \left(\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\right)_{p\times n} \text{ and} \\ \mathbf{L}_{\boldsymbol{\phi}\mathbf{Y}_i} &= \left(\frac{\partial\,\mathrm{vec}^\top(\boldsymbol{\Sigma})}{\partial\boldsymbol{\phi}}\mathrm{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_i)\right)_{3\times n}, \end{aligned}$$

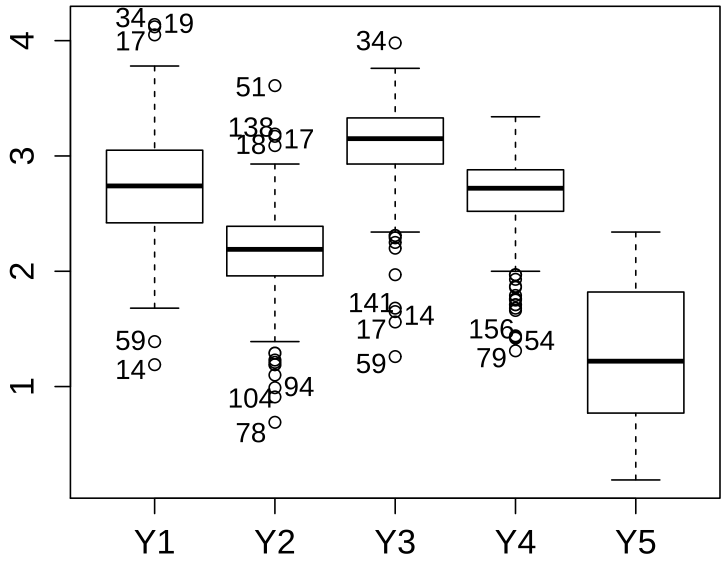
for $i = 1, \ldots, r$. Since the model assumes the same design matrix, $\mathbf{X}$, for each repetition, $\mathbf{GL}_{ii} = \frac{1}{r}\mathbf{H}$ (Galea et al., 2005). The leverage are the same for each repetition and the elements of h_{jj} for $j = 1, \ldots, n$ (Section 4.5) are used as diagnostic tools for the influence in the vector $\hat{\mathbf{Y}}_i$. The j-th response is potentially influential if $\frac{1}{r}h_{jj} > \frac{1}{r}\overline{h} + \frac{2}{r}\mathrm{sd}(\mathbf{H})$, where $\overline{h} = \sum_{j=1}^n h_{jj}/n$ and $\mathrm{sd}(\mathbf{H})$ is the standard deviation of $h_{11}, \ldots, h_{nn}$.

4.8 Application

4.8.1 The data set

The data set were collected in a grid of 7.20×7.20 m in an experimental area with 1.33 ha at Eloy Gomes Researh Center at *Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda* (COODETEC), in Cascavel city at Paraná State - Brazil, with Oxisol soil. It were collected soybean productivity data and four chemical contents during April in the years 1998 (Y_1) or year 1, 1999 (Y_2) year 2, 2000 (Y_3) year 3, 2001 (Y_4) year 4 and 2002 (Y_5) year 5 with 253 observations each. To explain the expectation value of the productivity, it were considered as explanatory variables in the model these chemical contents of soil: phosphorus (P)[mg dm^{-3}], potassium (K)[cmolc dm^{-3}], calcium (Ca)[cmolc dm^{-3}] and magnesium (Mg)[cmolc dm^{-3}]. A detailed descriptive analysis of the data set is presented in Section 3.9. Figure 4.1 shows the usual boxplots.

Figure 4.1: Boxplots for soybean productivity in the years 1998 (Y_1), 1999 (Y_2), 2000 (Y_3), 2001 (Y_4) and 2002 (Y_5).



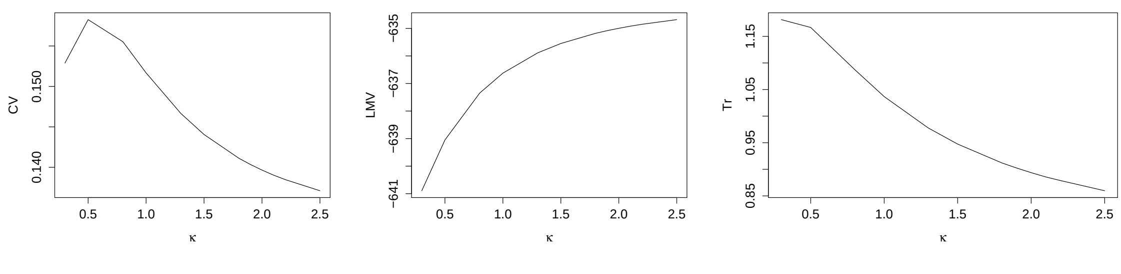
Source: From the author.

In order to choose the variance covariance structure that best describe the spatial dependence of the soybean productivity, we used cross validation (CV), the log-likelihood maximum value (LMV) and the trace of the asymptotic covariance matrix of an estimated mean (Tr) criteria presented in Table 4.1 and in Figure 4.2. These criteria lead us to select the Gaussian covariance function.

Table 4.1: CV, LMV and Tr criteria with all the observations.

κ	CV	LMV	Tr
0.3	0.1529	-640.8929	1.1814
0.5	0.1583	-639.0491	1.1667
0.8	0.1555	-637.3470	1.0873
1.0	0.1517	-636.6239	1.0367
1.3	0.1467	-635.8890	0.9777
1.5	0.1441	-635.5485	0.9474
1.8	0.1411	-635.1785	0.9121
1.9	0.1404	-635.0824	0.9025
2.0	0.1397	-634.9966	0.8937
2.1	0.1390	-634.9200	0.8856
2.2	0.1385	-634.8512	0.8788
2.5	0.1371	-634.6827	0.8597
∞(Gauss)	0.1284	-633.8891	0.8597

Figure 4.2: Criteria to choose the variance covariance structure (a) Cross validation (CV), (b) The log-likelihood maximum value (LMV) and (c) the trace of the asymptotic covariance matrix of an estimated mean (Tr).



(a) (b) (c)

Source: From the author.

Table 4.2 shows the parameters estimates considering the Gaussian covariance function, and the respective asymptotic standard errors (se) in parenthesis. We used the Z-test, which has normal asymptotic distribution, to test the hypothesis of the form H_0: $\beta_k = 0$ versus H_1: $\beta_k \neq 0$, where β_k is any of the parameters of the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$. The test statistic is given by $Z = \hat{\beta}_k/\mathrm{se}(\hat{\beta}_k)$, where $\hat{\beta}_k$ is the ML estimator of β_k and $\mathrm{se}(\hat{\beta}_k)$

the corresponding se. The level of significance of each parameter tested individually was fixed at 5% of significance. According to Table 4.2, $\hat{\beta}_0$ is significant. Using the same idea to test the $\phi's$, we find that their significants. The macronutrients P and K are needed in relatively large quantities in the soil compared to others to prevent plant deficiencies, but after a critical level is reached there is no additional yield increment.

Table 4.2: Parameters estimates for COODETEC data considering the Gaussian covariance function, the asymptotic standard errors (in parenthesis) and the p-values.

Intercept	P	K	Ca	Mg	nugget	sill	f(range)
$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
2.4013**	-0.0017	0.3270	0.0118	-0.0592	0.1927**	0.0579**	0.0407**
(0.1020)	(0.0109)	(0.1746)	(0.0392)	(0.0515)	(0.0178)	(0.0219)	(0.0098)

** significant for 5%

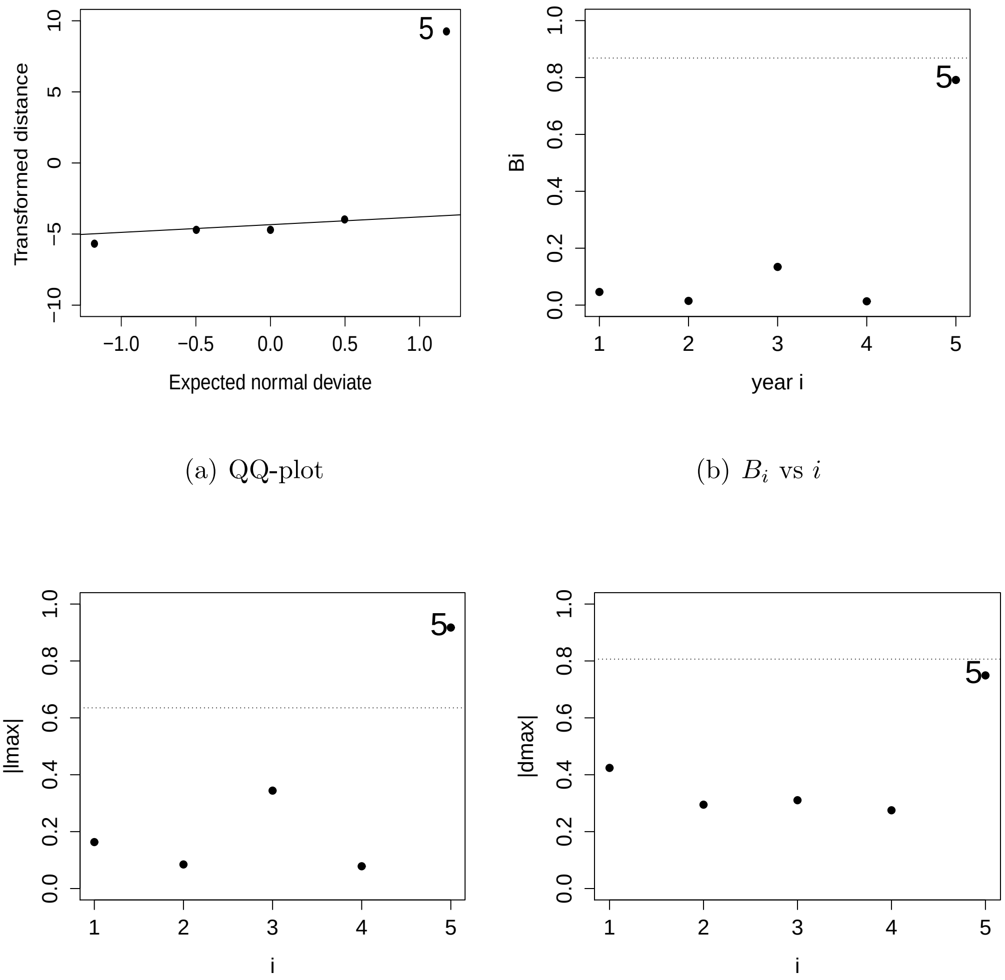
A critical analysis of the model's assumptions was carried out. Probability normal graphics of the transformed distances d_i given in Equation (4.8) were used to evaluate the goodness of fit and also to identify outliers, according Lange et al. (1989). As we can note, in Figure 4.3(a) the graphic identified an atypical behaviour of a point, which is above the line. The other transformed distance are aligned over the line with 45 degree.

$$d_i = \left\{ \left(\frac{\hat{\delta}_i}{n} \right)^{1/3} - \left[1 - \left(\frac{2}{9n} \right) \right] \right\} \Big/ \sqrt{\left(\frac{2}{9n} \right)}. \tag{4.8}$$

4.8.2 Case weight perturbation

Figure 4.3 presents influential plots for case weight perturbation that allow to evaluate the individual contribution of each repetition in the process of estimation. In this case each repetition, corresponding to a year, receive different weight. We can see that the year 2002 has more influence on the estimation process. The details are presented in Table 4.4 and Figure 4.5 show the difference between the the estimated mean of productivity when we consider and do not consider the year 2002.

Figure 4.3: The data for five repetitions (five years) from 1998 until 2002 (a) expected normal deviate vs transformed distance, (b) B_i vs year i, (c)$|l_{max}|$ vs year i and (d) $|d_{max}|$ vs year i, considering the case-weight perturbation.



(a) QQ-plot

(b) B_i vs i

(c) $|l_{max}|$ vs i

(d) $|d_{max}|$ vs i

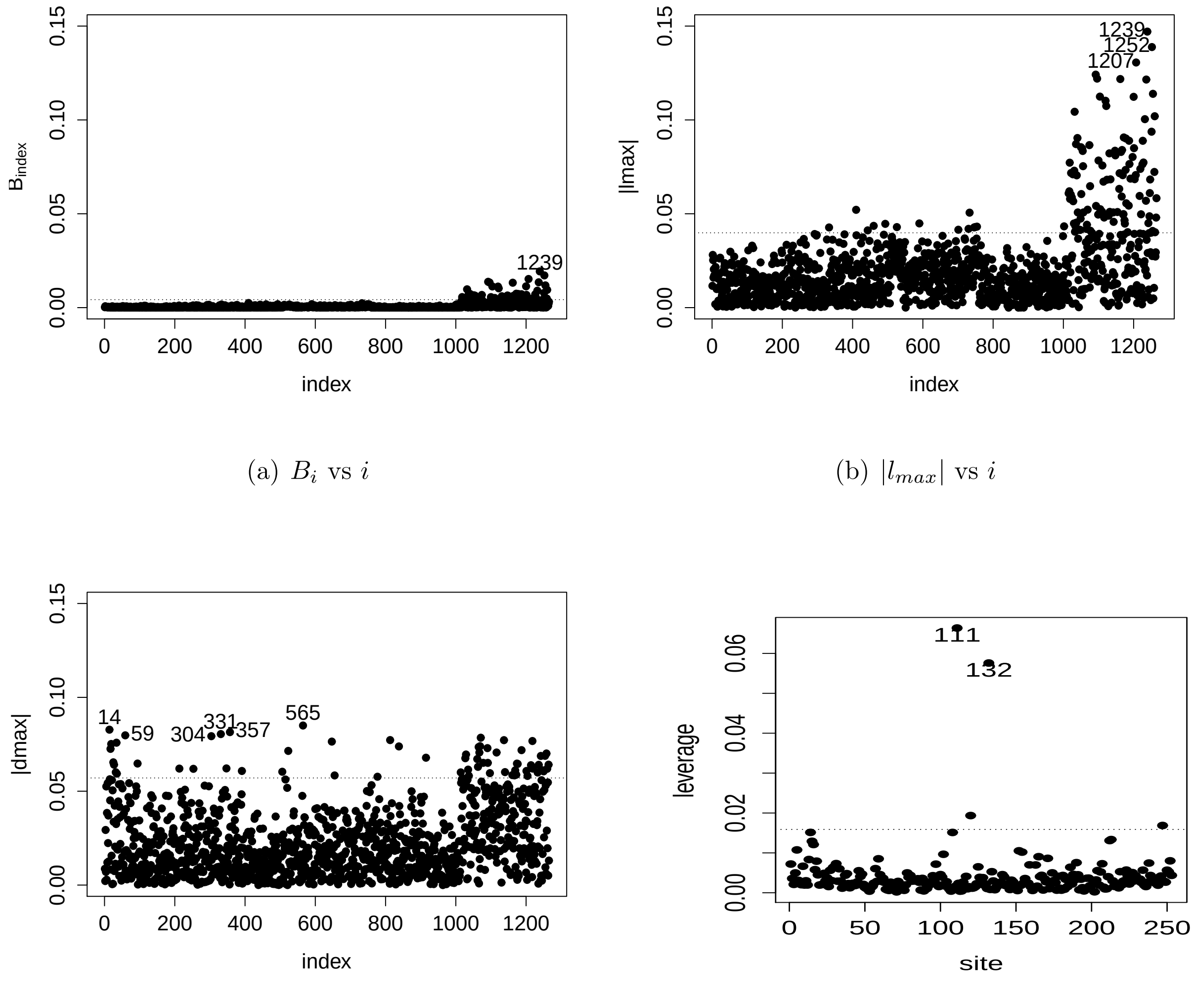
Source: From the author.

4.8.3 Perturbation on the response variable and generalized leverage

Figure 4.4(a), 4.4(b) and 4.4(c) present the graphics for the analysis of local influence in the response variable. The first two graphics agree that the observations belonging to the data collected in the year 2002 should be treated carefully. The third graphic pointed out observations #14 and #59 which appear in the first and third bloxplots of Figure 4.1. Figure 4.4(d) presents the leverage vs sites for each repetition, where the years have the same sites as potentially influent over the correspondent fitted value, as we mentioned in

Section 4.7.4. The analysis of the data set without sites #111 and #132 do not implie changes in the estimation of the parameters and inference. The choice of the covariance structure is still the Gaussian one and the results are quite similar with the one presented in Tables 4.1 and 4.2 and Figure 4.2.

Figure 4.4: The data for five repetitions (five years) from 1998 until 2002 (a) B_{index} vs $index$, (b) $|l_{max}|$ vs $index$, c) $|d_{max}|$ vs $index$, considering the perturbation on the response variable and d) leverage vs $index$ for a year. The index vary according the number of observations.



(a) B_i vs i

(b) $|l_{max}|$ vs i

(c) $|d_{max}|$ vs i

(d) Leverage vs i

Source: From the author.

4.8.4 Analysis without subject #5

We reanalyzed the data without subject #5, that correspond to the whole year 2002. The results for criteria to select the model is given in Table 4.3. The chosen model for the covariance structure is again the Gaussian covariance function.

Table 4.3: CV, LMV and Tr criteria for the data set without subject #5.

κ	CV	LMV	Tr
0.3	0.1468	-424.7226	1.2141
0.5	0.1535	-423.5043	1.1692
0.8	0.1421	-422.4204	1.0685
1.0	0.1324	-421.9738	1.0121
1.3	0.1210	-421.5312	0.9492
1.5	0.1153	-421.3311	0.9175
1.8	0.1090	-421.1188	0.8808
1.9	0.1073	-421.0647	0.8709
2.0	0.1059	-421.0170	0.8618
2.1	0.1045	-420.9747	0.8536
2.2	0.1033	-420.9372	0.8460
2.5	0.1004	-420.8470	0.8268
∞(Gauss)	0.0809	-420.5483	0.8268

In Tables 5.1 and 4.4 we note that the covariates Ca and Mg are still not significant in a level of 5%. The covariates P and K changed the signal but are still not significant. Be aware that the Z-score test verify individually the significance of the parameters, i.e., even they are not significant by their own they can be in the presence of the other. The parameters estimates and the asymptotic standard errors (in parenthesis) are given in Table 4.4.

Comparing Tables 4.2 and 4.4 we note that the covariates Ca and Mg are now significant in a level of 5%. The covariate P and K changed the signal but are still not significant. Be aware that the Z-score test verify individually the significance of the parameters, i.e., even they are not significant by their own they can be in the presence of the other.

Table 4.4: Parameters estimates for Coodetec data without subject #5 and the asymptotic standard errors (in parenthesis), considering the Gaussian covariance function.

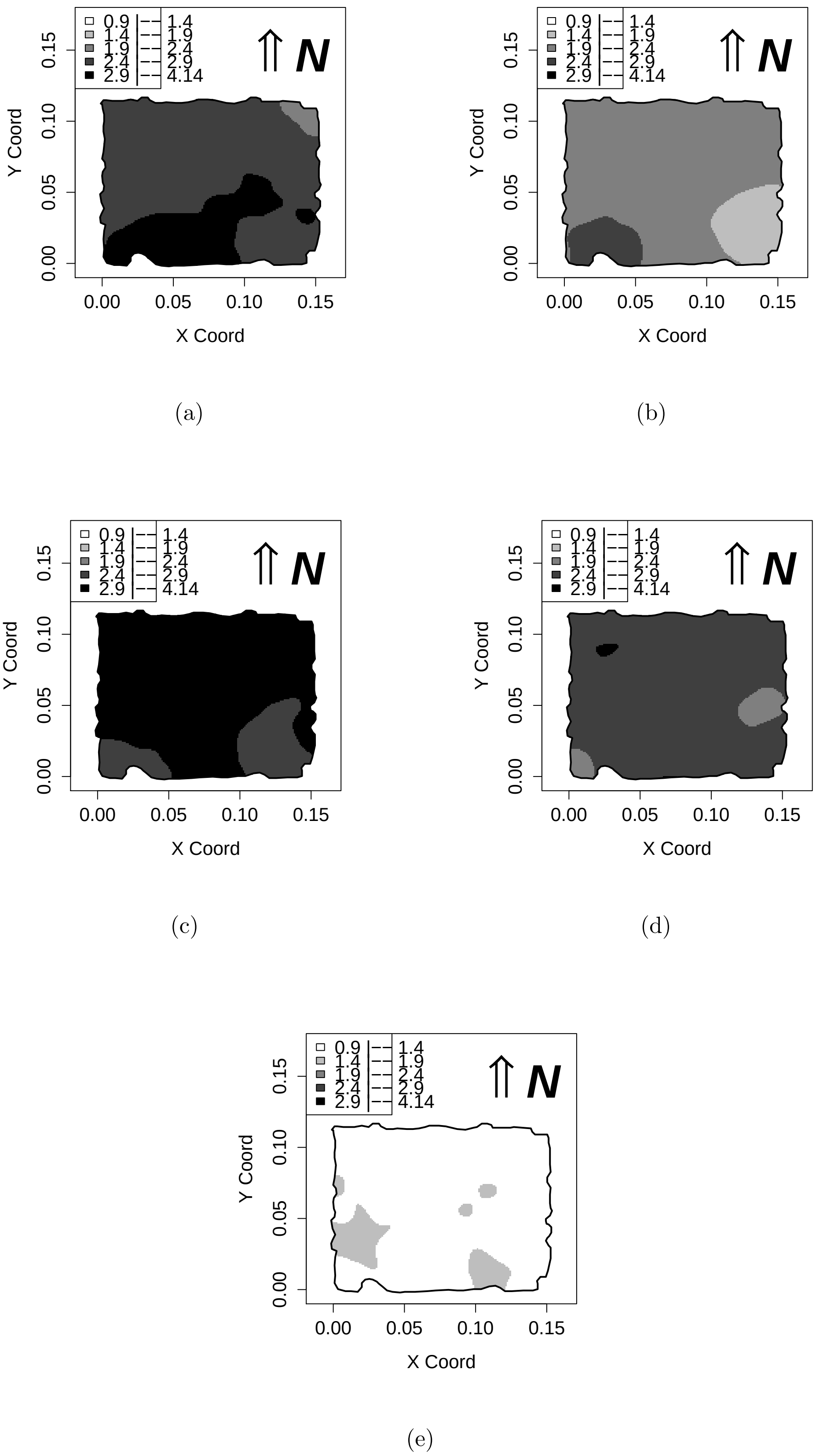
Intercept	P	K	Ca	Mg	nugget	sill	f(range)
$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
2.7512**	0.0047	-0.1920	0.0809**	-0.1528**	0.1313**	0.0490**	0.0388**
(0.0966)	(0.0101)	(0.1640)	(0.0364)	(0.0477)	(0.0122)	(0.0173)	(0.0083)

** significant for 5%

4.8.5 Maps

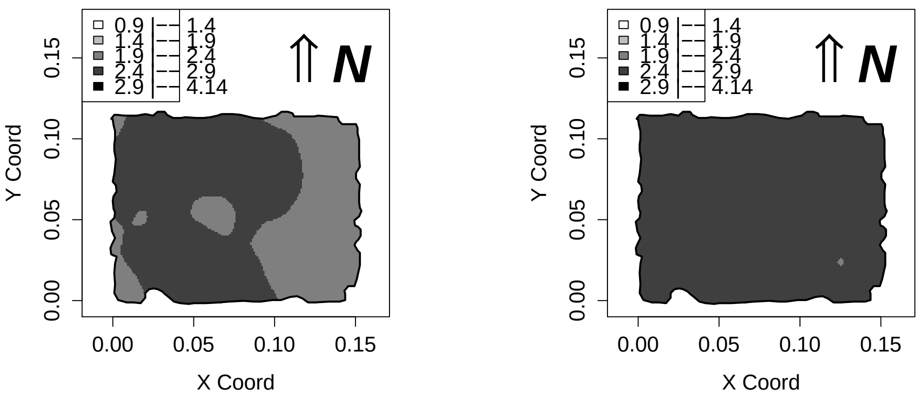
Figure 4.5 presents the maps for the response variable for each year and for the mean of the years. For all cases the same explanatory variables were considered, P, K, Ca and Mg, as well we considered the same covariance function, i.e., the Gaussian covariance structure, given in Table 4.2. Yorinori et al. (2005) commented that on May of 2001, rust surveys showed spread throughout most of Paraguay and into western and northern Parana, Brazil, and that in the 2001-2002 season, rust was spread in Brazil to more than 60% of the soybean acreage, causing field losses estimated of 0.1 million metric tons. The smallest values of the productivity are clear is the latest two years (Figures 4.5(d) and 4.5(e)), and even worst in the latest one. And also is evident the difference between the maps considering all the years and the map without the fifth year in Figures 4.6(a) and 4.6(b), respectively. The map without the fifth year show greater value for the mean of the productivity than the one considering the whole information.

Figure 4.5: Maps for productivity in the year (a) 1998, (b) 1999, (c) 2000, (d) 2001 and (e) 2002.



Source: From the author.

Figure 4.6: Maps for productivity in the year (a) for the mean of the whole years and (b) for the mean of the whole years (without year 2002).



(a) (b)

Source: From the author.

4.9 Conclusion

The purpose of diagnostic techniques is to evaluate the role of observations in the fit of the proposed model and identify influential cases that may affect the values of statistical interest. The evaluation of the stability of the fitted model in a data set should be part of all statistical analysis, since a few observations can influence/distort the values of the statistic interest (estimators and hypothesis test) and lead to wrong inference.

The goal of this work was to propose local influence measures to detect observations that can distort some statistics of interest in the Gaussian spatial linear model with independent repetitions. We proposed measures based on the likelihood displacement and some first order influence measures to assess the stability of the likelihood function, using the case weight perturbation scheme and perturbation on the response variable. We applied the methodology developed in the paper to a data set of soybean productivity collected in the West region of Paraná State, Brazil.

We obtained explicit expressions for the matrices necessary for implementing these diagnostic techniques using the Matérn family of covariance functions. We would like to emphasize that first order local influence measures to assess the stability of the likelihood function, are very simple to calculate, as can be seen in section 4.7. This type of results is not very common in the literature on the topic. Thus we extended the work of Uribe-Opazo et al. (2012) to the case of independent repetitions.

To select the spatial dependence structure we use the likelihood function, cross validation and the trace of the asymptotic covariance matrix of the ML estimator of the mean of the process. In the application considered, the three criteria suggest to select the same covariance structure.

As in other statistical models, based on the normality assumption, the influence of the outliers on statistics of interest is considerable. So, there are important effects, for example, in the soybean productivity maps, as shown in Figure 4.5. Non-normal alternative distributions, as the t multivariate distribution, and the framework of spatio-temporal models for this data set are in progress.

Chapter 5

Global Diagnostics on Gaussian spatial linear models with repetitions

5.1 Abstract

Neste capítulo são apresentadas técnicas de diagnósticos de influência global nos modelos espaciais lineares Gaussianos com repetições. São revisados os conceitos de distância de Cook baseada na função de verossimilhança e na função Q. A principal contribuição é este novo enfoque para os modelos geoestatísticos. Aplicação a conjunto de dados reais ilustra a metodologia desenvolvida. Extensivos cálculos são apresentados no Apêndice.

5.2 Introduction

Geostatistics refers to the sub-branch of spatial statistics in which the data consist of a finite sample of measured values relating to an underlying spatially continuous phenomenon, Diggle and Ribeiro Jr. (2007). Some proposals have discussed Gaussian spatial linear models to study the structure of dependence in spatially referenced data. For more details about estimation, inference methods and applications of these models, see (De Bastiani et al., 2016).

Most diagnostic measures were originally developed under linear regression models. Cook's (Cook (1977)) distance is one of the most known diagnostic tools for detecting influential individual or subsets of observations in linear regression. Cook and Weisberg (1982) give a comprehensive account of a variety of methods for the study of influence on linear regression. Chatterjee and Hadi (1988) treat linear regression diagnostics as a tool for application of linear regression models to real-life data. Haslett and Dillane (2004) extended previous work of deletion diagnostics for estimates of variance components obtained by restricted maximum likelihood estimation for the linear mixed model.

Davison and Tsai (1992) reviewed various diagnostics for generalized linear models and extended to more general models, such as models for censored and grouped data for nonlinear regression. Wei (1998) presented a comprehensive introduction to exponential family nonlinear models, giving attention to regression diagnostics and influence analysis.

Some proposals have discussed diagnostics tools for spatial linear models. In geostatistics, an atypical observation can cause changes in environmental and geological patterns. Christensen et al. (1992a) measured the effect of the observations on prediction using diagnostics based on case-deletion. Christensen et al. (1993) proposed diagnostics to detect influential observations on the estimation of the covariance function. Zewotir and Galpin (2005) provided routine diagnostic tools for linear mixed models, which are computationally inexpensive. Militino et al. (2006) proposed influence measures for multivariate spatial linear models. Uribe-Opazo et al. (2012) presented local influence for Gaussian spatial linear models.

De Gruttola et al. (1987) described measures of influence and leverage for a generalized least squares (GLS) estimator of the regression coefficients in a class of multivariate linear models for repeated measurements. Waternaux et al. (1989) suggested several pratical procedures to detect outliers for the repeated measurements model based on the global influence approach. For example, for longitudinal data, Preisser and Qaqish (1996) developed Cook's distance for generalized estimation equations. Christensen et al. (1992b), Banerjee (1998), Demidenko and Stukel (2005) and Pan et al. (2014) considered case deletion and subject deletion diagnostics for linear mixed models. Assumpção et al. (2014) present the generalized leverage to evaluate the influence of a vector on its own predicted value for different approach for spatial linear models.

In this work we propose the study of several global influence measures for quantifying the effects of perturbing spatial linear models with repeated measures. The chapter unfolds as follows. Section 5.3 presents the Gaussian spatial linear model with repetition, the complete loglikelihood and the Q function, and present a brief description of the maximum likelihood estimation and asymptotic standard erros calculations. Section 5.4 reviews concepts of global influence based on the likelihood and on the Q-function. Section 5.5 contains an application with real data, to illustrate the methodology developed in this paper. Finally, Section 5.6 contains some concluding remarks. Calculations are presented in the appendices.

5.3 The Gaussian Spatial Linear Model with Repetitions

For a spatial process, the basic object we consider is a stochastic process $\{\mathbf{Y}_i(\mathbf{s}), \mathbf{s} \in S \subset \mathbb{R}^2\}$ (bi-dimensional Euclidean space), usually though not necessarily in $\mathbb{R}^2$, and $i = 1$ means one single realization of the process. We also assume that the variance of $\mathbf{Y}_i(\mathbf{s})$ exists for all $\mathbf{s} \in S$. The process $\mathbf{Y}_i$ is said to be Gaussian if, for any $k \geq 1$ and locations $\mathbf{s}_1, \ldots, \mathbf{s}_k$, the vector $(Y_1(\mathbf{s}_1), \ldots, Y_1(\mathbf{s}_k))$ has a multivariate normal distribution.

Let $\mathbf{Y} = \mathbf{Y}(\mathbf{s}) = \text{vec}(\mathbf{Y}_1(\mathbf{s}), \ldots, \mathbf{Y}_r(\mathbf{s}))$ be an $nr \times 1$ random vector of r independently stochastic process of n elements each, that belong to the family of Gaussian distributions and depend on the sites $\mathbf{s}_j \in S \subset \mathbb{R}^2$ for $j = 1, \ldots, n$, where $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)^\top$. The "vec" operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. It is assumed the i-th stochastic process $\mathbf{Y}_i(\mathbf{s}) = \text{vec}(Y_i(\mathbf{s}_1), \ldots, Y_i(\mathbf{s}_n))$, represents the $n \times 1$ vector, for $i = 1, \ldots, r$, which can be expressed as a linear mixed model by

$$\mathbf{Y}_i(\mathbf{s}) = \mu_i(\mathbf{s}) + \mathbf{b}_i(\mathbf{s}) + \tau_i(\mathbf{s}) \tag{5.1}$$

where, the deterministic term $\mu_i(\mathbf{s})$ is an $n \times 1$ vector, the means of the process $\mathbf{Y}_i(\mathbf{s})$, $\mathbf{b}_i(\mathbf{s})$ and $\tau_i(\mathbf{s})$ are independents and together form the stochastic error, i.e., $\mathbf{b}_i(\mathbf{s}) + \tau_i(\mathbf{s}) = \epsilon_i(\mathbf{s})$ is an $n \times 1$ vector of a stationary process with zero mean vector, $\text{E}[\epsilon_i(\mathbf{s})] = \mathbf{0}$, and covariance $\boldsymbol{\Sigma}_i$. For a linear model, the mean vector $\mu_i(\mathbf{s})$ can be written as $\mu_i(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}$, where, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $p \times 1$ vector of unknown parameters, $\mathbf{X} = \mathbf{X}(\mathbf{s}) = [\mathbf{x}_{j1}(\mathbf{s}) \ldots \mathbf{x}_{jp}(\mathbf{s})]$ is an $n \times p$ matrix of p explanatory variables, for $j = 1, \ldots, n$, i.e, the design matrix $\mathbf{X}$ is the same for all r repetitions.

As in Smith (2001), the GSLM for the i-th independent stochastic process, assuming a homogeneous process, can be written in matrix form by $\mathbf{Y}_i(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta} + \epsilon_i(\mathbf{s})$.

The covariance matrix $\mathbf{\Sigma}_i = \mathbf{\Sigma} = [C(s_u, s_v)]$ is an $n \times n$ covariance matrix of $\mathbf{Y}_i(\mathbf{s})$ for the i-th repetition, $i = 1, \ldots, r$. The matrix $\mathbf{\Sigma}$ is non-singular, symmetric and positive defined, associated to the vector $\mathbf{Y}_i(\mathbf{s})$, where for the stationary and isotropic process, the elements $C(\mathbf{s}_u, \mathbf{s}_v)$ depend on the Euclidean distance $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$ between points $\mathbf{s}_u$ and $\mathbf{s}_v$.

So, from what was previously written we have that $\mathbf{Y} \sim N_{nr}(\mathbf{1}_r \otimes \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_r \otimes \mathbf{\Sigma})$, with probability density function (pdf) given by

$$\begin{aligned} f(\mathbf{Y}, \boldsymbol{\theta}) &= \textstyle\prod_{i=1}^{r} f(\mathbf{Y}_i, \boldsymbol{\theta}) \\ &= \textstyle\prod_{i=1}^{r} (2\pi)^{-n/2} |\mathbf{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})\right], \end{aligned}$$

where $\otimes$ denote the kronecker product and $(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}) = \delta_i$ is the Mahalanobis distance.

The covariance matrix $\mathbf{\Sigma}$ has a structure which depends on parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$ as given in Equation (5.2) (Mardia and Marshall, 1984; Uribe-Opazo et al., 2012):

$$\mathbf{\Sigma} = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}, \tag{5.2}$$

where, $\phi_1 \geq 0$ is the parameter known as nugget effect; $\phi_2 \geq 0$ is known for sill ; $\mathbf{R} = \mathbf{R}(\phi_3, \phi_4) = [(r_{uv})]$ or $\mathbf{R} = \mathbf{R}(\phi_3) = [(r_{uv})]$ is an $n \times n$ symmetric matrix, which is function of $\phi_3 > 0$, and sometimes also function of $\phi_4 > 0$, with diagonal elements $r_{uu} = 1, (u = 1, \ldots, n)$; $r_{uv} = \phi_2^{-1} C(s_u, s_v)$ for $\phi_2 \neq 0$, and $r_{uv} = 0$ for $\phi_2 = 0$, $u \neq v = 1, \ldots, n$, where r_{uv} depends on the Euclidean distance $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$ between points $\mathbf{s}_u$ and $\mathbf{s}_v$; ϕ_3 is a function of the model range (a), ϕ_4 when exists is known as the smoothness parameter, and $\mathbf{I}_n$ is an $n \times n$ identity matrix.

A number of different structures are available, and the question is which one is best. There are a large number of covariance structures to choose from. For example, exponential, Gaussian and Matérn model. To select one of them, beyond the maximum value of the log-likelihood, we used the cross validation and the trace of the asymptotic covariance matrix (Kano et al., 1993).

The maximum likelihood estimation were already presented in previous chapters. The unknown parameters to be estimated are the β's and ϕ's. The log-likelihood for the GLSM for the r independent repetitions is given by $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{r} \mathcal{L}_i(\boldsymbol{\theta})$,

$$\mathcal{L}_i(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}),$$

and the score functions are given by

$$\begin{aligned}
\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{r} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}_i, \\
\mathbf{U}(\boldsymbol{\phi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} &= -\frac{r}{2}\frac{\partial\, \mathrm{vec}^\top(\boldsymbol{\Sigma})}{\partial \phi}\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\sum_{i=1}^{r}\frac{\partial\, \mathrm{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}}\mathrm{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top\boldsymbol{\Sigma}^{-1}),
\end{aligned}$$

where $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}$. The maximum likelihood estimator $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}},$$

where $\bar{\mathbf{Y}} = (\bar{\mathbf{Y}}_1, \ldots, \bar{\mathbf{Y}}_n)^\top$, with $\bar{\mathbf{Y}}_j = \frac{1}{r}\sum_{i=1}^{r} Y_i(\mathbf{s}_j)$, $j = 1, \ldots, n$.

The expected information matrix, $\mathbf{F}(\boldsymbol{\theta})$, is given by, (De Bastiani et al., 2016; Waller and Gotway, 2004)

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{F} = \begin{pmatrix} \mathrm{E}[\mathbf{U}(\beta)\mathbf{U}^\top(\beta)] & \mathrm{E}[\mathbf{U}(\beta)\mathbf{U}^\top(\phi)] \\ \mathrm{E}[\mathbf{U}(\phi)\mathbf{U}^\top(\beta)] & \mathrm{E}[\mathbf{U}(\phi)\mathbf{U}^\top(\phi)] \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{\beta\beta} & \mathbf{F}_{\beta\phi} \\ \mathbf{F}_{\phi\beta} & \mathbf{F}_{\phi\phi} \end{pmatrix},$$

where $\mathbf{F}_{\beta\beta} = r\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\mathbf{F}_{\beta\phi} = \mathbf{0}$, $\mathbf{F}_{\phi\beta} = \mathbf{0}$ and $\mathbf{F}_{\phi\phi} = \frac{r}{2}\frac{\partial\, \mathrm{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\frac{\partial\, \mathrm{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top}$.

We used $\mathbf{F}_{\beta\beta}^{-1}$ and $\mathbf{F}_{\phi\phi}^{-1}$ to estimate the dispersion matrices for the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\phi}}$, respectively.

According to Osorio (2006) the conditional distribution is given by

$$\mathbf{b}_i|\mathbf{Y}_i \sim N\left(\mathbf{G}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}), \Omega_i\right), \; i = 1\ldots, r,$$

where $\Omega_i = \mathbf{G} - \mathbf{G}\mathbf{Z}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i\mathbf{G}$. Then the complete loglikelihood is given by

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{r} \mathcal{L}_{ic}(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ is the vector of unknown parameters and

$$\begin{aligned}\mathcal{L}_{ic}(\boldsymbol{\theta}) &= -\frac{1}{2}\log|\mathbf{D}_i| - \frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{D}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\\ &\quad -\frac{1}{2}\log|\mathbf{G}| - \frac{1}{2}\mathbf{b}_i^\top\mathbf{G}^{-1}\mathbf{b}_i\end{aligned}$$

Let consider $\phi_1 = \phi_1$, $\eta = (\eta_1, \eta_2, \eta_3)^\top = (\phi_2, \phi_3, \phi_4)^\top$, so $\mathbf{R} = \mathbf{R}(\eta_2, \eta_3)$ and $\boldsymbol{\Sigma} = \phi_1\mathbf{I} + \eta_1\mathbf{R}$, thus $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi_1, \eta^\top)^\top$, more details are given on Appendix C.1. The Q -function, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \mathrm{E}[\mathcal{L}_c(\boldsymbol{\theta})|\mathbf{Y}, \boldsymbol{\theta} = \boldsymbol{\theta}^*]$, for the model presented in (5.1) is of the form

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) &= -\frac{nr}{2}\log\phi_1 - \frac{r}{2\phi_1}\mathrm{tr}(\Omega^*) - \frac{1}{2\phi_1}\sum_{i=1}^{r}\mathbf{r}_i^{*\top}\mathbf{r}_i^*\\ &\quad -\frac{nr}{2}\log\eta_1 - \frac{r}{2}\log|\mathbf{R}| - \frac{1}{2\eta_1}\sum_{i=1}^{r}\mathrm{tr}[\mathbf{R}^{-1}(\Omega^* + \mathbf{b}_i^{*\top}\mathbf{b}_i^*)],\end{aligned}$$

where

$$\begin{aligned}\mathbf{r}_i^* &= \mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta} - \mathbf{b}_i^*,\\ \mathbf{b}_i^* &= \mathrm{E}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}^*) = \eta_1^*\mathbf{R}^*\boldsymbol{\Sigma}^{*-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}^*),\\ \Omega^* &= \mathrm{Cov}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}^*) = \eta_1^*\mathbf{R}^* - \eta_1^{*2}\mathbf{R}^*\boldsymbol{\Sigma}^{*-1}\mathbf{R}^*,\end{aligned}$$

and the $*$ symbol means that the objects are in function of $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

5.4 Diagnostic Techniques

Detecting influential observations is an important step in the analysis of a data set. There are different approaches to assess the influence of perturbations in a data set and in the model given the estimated parameters. Cook (1977) gives a starting point to the development of case-deletion diagnostics for all sorts of statistical models. Case-deletion is an example of a global influence analysis, that is to asses the effect of an observation by completely removing it. Cook (1986) presents the local influence approach, that is, a weight ω_i is given for each case and the effect on the parameter estimation is measured by perturbing around these weights. Choosing weights equal zero or one corresponds to the global case-deletion approach. In general, perturbation measures do not depend on the data directly, but rather on its structure via the model.

There are some papers in the literature on diagnostic for spatial linear models. Diamond and Armstrong (1984) and Warnes (1986) observed the sensitivity of predictions to perturbations in the covariance function. Christensen et al. (1992a) discussed case deletion diagnostics for detecting observations that are influential for prediction based on universal kriging, while Christensen et al. (1993) considered diagnostic with the deletion of points to estimate the parameters of the covariance function by the method of restricted maximum likelihood. More recently, Cerioli and Riani (1999), Militino et al. (2006) showed that case deletion diagnostics do suffer from masking and suggest robust procedures based on subsets of data free from outliers. In this direction, Filzmoser et al. (2014) proposed the Mahalanobis distance to identify multivariate outliers.

5.4.1 Global influence

Case-deletion is a diagnostic technique that evaluate the impact on the parameters estimates given the model, by eliminating one or more observations from the data set. This kind of diagnostics techniques have been discussed by Cook and Weisberg (1982) and Chatterjee and Hadi (1988), Pan et al. (2014), among others. We can eliminate a subject, each observation or an entire location, by using the idea of the technique proposed by Cook (1977) known as Cook's distance which has a typical measure defined by

$$D_{i\theta} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]})^{\top}\mathbf{M}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]}),$$

where $\mathbf{M}$ is an appropriately chosen positive definite matrix, for instance, the inverse of the asymptotic covariance matrix and $\hat{\boldsymbol{\theta}}_{[i]}$ is the estimate of $\boldsymbol{\theta}$ refitting the model with the same principles used to obtain $\hat{\boldsymbol{\theta}}$, but without the i-th observation.

Likelihood-based diagnostics

In the Cook's Distance case-deletion we eliminate a subject to evaluate it influences on the analysis. For the GLSM it is given by

$$D_{i\theta} \;=\; (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]})^{\top}[-\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})](\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]}),$$

for $i = 1, \ldots, r$, where $\mathcal{L}_{[i]}(\boldsymbol{\theta})$ is the loglikelihood with the i-th subject deleted, $\ddot{\mathcal{L}}_{[i]}(\boldsymbol{\theta})$ is the second derivative of $\mathcal{L}_{[i]}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_{[i]}$ is the maximum likelihood estimate under $\mathcal{L}_{[i]}(\boldsymbol{\theta})$. Here $\mathbf{M} = [-\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})]$, because we have that $[-\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})]^{-1}$ is an estimator of the of the asymptotic covariance matrix and as mentioned above, $\mathbf{M}$ commonly is the inverse of this estimator. Since the computation of $[\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})]$ may become cumbersome when r is too large, it is used $\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})$ to replace $\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})$. In general, the asymptotic covariance matrix $[-\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})]^{-1}$ is not block-diagonal, so to decompose the Cook's statistic into different components of interest an alternative is to use the expectation of $[-\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})]^{-1}$, $\mathrm{E}[-\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})]^{-1}$, which is the inverse of the expected information matrix. For the GLSM the $[-\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})]^{-1} = \mathrm{E}[-\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})]^{-1} = \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}$, thus both give the same measure given by

$$D_{i\theta} \;=\; (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]})^\top \mathbf{F}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]}).$$

To calculate $\hat{\boldsymbol{\theta}}_{[i]}$, we may use a one-step Newton Raphson approximation at $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}_{[i]} = \hat{\boldsymbol{\theta}} + [-\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})]^{-1}\dot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})$, where the dots over the functions denote derivatives with respect to $\boldsymbol{\theta}$, and

$$\mathcal{L}_{[i]}(\boldsymbol{\theta}) = -\frac{n(r-1)}{2}\log(2\pi) - \frac{(r-1)}{2}|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{\substack{k=1 \\ k\neq i}}^{r} (\mathbf{Y}_k - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_k - \mathbf{X}\boldsymbol{\beta}),$$

for $i = 1, \ldots, r$. And again, to overcome some computational difficulty, using $\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})$ to replace $\ddot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})$, and so on $\mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}$, the approximated Cook's distance becomes

$$\begin{aligned} D^1_{i\theta} &= [\dot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})]^\top \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}[\dot{\mathcal{L}}_{[i]}(\hat{\boldsymbol{\theta}})] \\ &= [U_{[i]}(\hat{\boldsymbol{\theta}})]^\top \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}[U_{[i]}(\hat{\boldsymbol{\theta}})], \end{aligned}$$

for $i = 1, \ldots, r$, where $U_{[i]}(\hat{\boldsymbol{\theta}}) = (U_{[i]}(\hat{\boldsymbol{\beta}}), U_{[i]}(\hat{\boldsymbol{\phi}}))^\top$ is the score function formula without the i-th subject evaluated at $\hat{\boldsymbol{\theta}}$ (given below). We can decompose $D^1_{i\theta} = D^1_{i\beta} + D^1_{i\phi}$, which means that the diagnostic measure of $\boldsymbol{\theta}$ is the sum of the diagnostic measures of the fixed effects $\boldsymbol{\beta}$ and the variance components $\boldsymbol{\phi}$ in terms of $D^1_{i\theta}$, for $i = 1, \ldots, r$, where

$D^1_{i\beta} = [U_{[i]}(\hat{\boldsymbol\beta})]^\top \mathbf{F}^{-1}_{\beta\beta}[U_{[i]}(\hat{\boldsymbol\beta})]$ and $D^1_{i\phi} = [U_{[i]}(\hat{\boldsymbol\phi})]^\top \mathbf{F}^{-1}_{\phi\phi}[U_{[i]}(\hat{\boldsymbol\phi})]$.

$$Furthermore, U_{[i]}(\hat{\boldsymbol\beta}) = -\mathbf{X}^\top \hat{\boldsymbol\Sigma}^{-1}\hat{\epsilon}_i,$$

and

$$U_{[i]}(\hat{\phi}_j) \quad = \quad \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol\Sigma}^{-1}\frac{\partial\hat{\boldsymbol\Sigma}}{\partial\boldsymbol\phi_j}\right) - \frac{1}{2}\hat{\epsilon}_i^\top\frac{\partial\hat{\boldsymbol\Sigma}^{-1}}{\partial\boldsymbol\phi_j}\hat{\epsilon}_i,$$

or in vec notation is given by

$$U_{[i]}(\hat{\boldsymbol\phi}) \quad = \quad \frac{1}{2}\frac{\partial\,\mathrm{vec}^\top(\hat{\boldsymbol\Sigma})}{\partial\phi}\mathrm{vec}(\hat{\boldsymbol\Sigma}^{-1}) - \frac{1}{2}\frac{\partial\,\mathrm{vec}^\top(\hat{\boldsymbol\Sigma})}{\partial\boldsymbol\phi}\mathrm{vec}(\boldsymbol\Sigma^{-1}\hat{\epsilon}_i\hat{\epsilon}_i^\top\hat{\boldsymbol\Sigma}^{-1}),$$

where $\hat{\epsilon}_i = (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol\beta})$, for $i = 1, \ldots, r$.

Q-function-based diagnostics

Pan et al. (2014) also proposed a case-deletion approach to identify influential subjects and influential observations in linear mixed models, based on the Q-function, the conditional expectation of the logarithm of the joint-likelihood between responses and random effects. Let consider $\phi_1 = \phi_1$, $\eta = (\eta_1, \eta_2, \eta_3)^\top = (\phi_2, \phi_3, \phi_4)^\top$, so $\mathbf{R} = \mathbf{R}(\eta_2, \eta_3)$ and $\boldsymbol\Sigma = \phi_1\mathbf{I} + \eta_1\mathbf{R}$, thus $\boldsymbol\theta = (\boldsymbol\beta^\top, \phi_1, \eta^\top)^\top$, more details are given on Appendix C.1. The Q-function, $Q(\boldsymbol\theta|\boldsymbol\theta^*) = \mathrm{E}[\mathcal{L}(\boldsymbol\theta)|\mathbf{Y}, \boldsymbol\theta = \boldsymbol\theta^*]$, for the model presented in (5.1) is of the form

$$\begin{aligned} Q(\boldsymbol\theta|\boldsymbol\theta^*) \quad = \quad & -\frac{nr}{2}\log\phi_1 - \frac{r}{2\phi_1}\operatorname{tr}(\Omega^*) - \frac{1}{2\phi_1}\sum_{i=1}^r \mathbf{r}_i^{*\top}\mathbf{r}_i^* \\ & -\frac{nr}{2}\log\eta_1 - \frac{r}{2}\log|\mathbf{R}| - \frac{1}{2\eta_1}\sum_{i=1}^r \operatorname{tr}[\mathbf{R}^{-1}(\Omega^* + \mathbf{b}_i^{*\top}\mathbf{b}_i^*)], \end{aligned}$$

where

$$\begin{aligned} \mathbf{r}_i^* \quad &= \quad \mathbf{Y}_i - \mathbf{X}\boldsymbol\beta - \mathbf{b}_i^*, \\ \mathbf{b}_i^* \quad &= \quad \mathrm{E}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol\theta^*) = \eta_1^*\mathbf{R}^*\boldsymbol\Sigma^{*-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol\beta^*), \\ \Omega^* \quad &= \quad \mathrm{Cov}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol\theta^*) = \eta_1^*\mathbf{R}^* - \eta_1^{*2}\mathbf{R}^*\boldsymbol\Sigma^{*-1}\mathbf{R}^*, \end{aligned}$$

and the $*$ symbol means that the objects are in function of $\boldsymbol\theta = \boldsymbol\theta^*$.

To calculate the case-deletion estimate $\hat{\boldsymbol\theta}_{[i]}$ of $\boldsymbol\theta$, Zhu and Lee (2001) proposed the use of the following one-step approximation based on the Q-function

$$\hat{\boldsymbol{\theta}}_{[i]} = \hat{\boldsymbol{\theta}} + [-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^{-1}\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}),$$

where $Q_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = Q_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ be the Q-function formed without the i-th subject but evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, and $\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = [\partial Q_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\ddot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = [\partial^2 Q_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. The Q-function-based Cook's statistic from Zhu and Lee (2001) and presented in Pan et al. (2014) is given by

$$QD_{i\theta} = [\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^\top[-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^{-1}[\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]. \tag{5.3}$$

Despite of the Q-function-based Cook's statistic be the sum of the Q-function-based Cook's statistics for the fixed effects $\boldsymbol{\beta}$ and the variance components $\boldsymbol{\phi}$, $QD_{i\theta} = QD_{i\beta} + QD_{i\phi}$, in general for the variance components it can not be separated, where $\boldsymbol{\eta} = (\phi_2, \phi_3, \phi_4)^\top$ are the variance components of the random effect and ϕ_1 the variance component of the random error. Then, Pan et al. (2014) propose to use $-\operatorname{E}[\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]$ in (5.3) instead of $-\ddot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$. Pan et al. (2014) proved a theorem implying that, when replacing the log-likelihood with the Q-function in the EM algorithm, the one-step approximation to the maximum likelihood $\hat{\boldsymbol{\theta}}_{[i]}$ maintains the same accuracy, of order $O_p(r^{-2})$. The modified Q-function-based Cook's statistic is of the form

$$QD^1_{i\theta} = [\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^\top\{-\operatorname{E}[\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]\}^{-1}[\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})].$$

$$\dot{Q}_{[i]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \begin{pmatrix} \dot{Q}_{i\beta} \\ \dot{Q}_{i\phi_1} \\ \dot{Q}_{i\eta_1} \\ \dot{Q}_{i\eta_2} \\ \dot{Q}_{i\eta_3} \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{\phi_1}\mathbf{X}^\top\mathbf{r}_i^* \\ -\dfrac{1}{2\phi_1^2}\operatorname{tr}(\Omega^*) - \dfrac{1}{2\phi_1^2}\mathbf{r}_i^{*\top}\mathbf{r}_i^* + \dfrac{n}{2\phi_1} \\ \dfrac{n}{2\eta_1} - \dfrac{1}{2\eta_1^2}\operatorname{tr}\left[\mathbf{R}^{-1}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ \dfrac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\dfrac{\partial\mathbf{R}}{\partial\eta_2}\right) + \dfrac{1}{2\eta_1}\operatorname{tr}\left[\dfrac{\partial\mathbf{R}^{-1}}{\partial\eta_2}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ \dfrac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\dfrac{\partial\mathbf{R}}{\partial\eta_3}\right) + \dfrac{1}{2\eta_1}\operatorname{tr}\left[\dfrac{\partial\mathbf{R}^{-1}}{\partial\eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \end{pmatrix},$$

for $i = 1, \ldots, r$, and $\mathrm{E}[-\ddot{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)]$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$ for the spatial linear mixed model has the elements

$$
\begin{aligned}
\mathrm{E}[-\ddot{Q}_{\beta\beta}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{r}{\hat{\phi}_1}\mathbf{X}^\top\mathbf{X},\\
\mathrm{E}[-\ddot{Q}_{\beta\phi_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \mathbf{0},\\
\mathrm{E}[-\ddot{Q}_{\beta\eta_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \mathbf{0},\\
\mathrm{E}[-\ddot{Q}_{\beta\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \mathbf{0},\\
\mathrm{E}[-\ddot{Q}_{\beta\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \mathbf{0},\\
\mathrm{E}[-\ddot{Q}_{\phi_1\phi_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{2r}{\hat{\phi}_1^3}\,\mathrm{tr}(\hat{\Omega}) + \frac{nr}{2\hat{\phi}_1^2},\\
\mathrm{E}[-\ddot{Q}_{\phi_1\eta_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= 0,\\
\mathrm{E}[-\ddot{Q}_{\phi_1\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= 0,\\
\mathrm{E}[-\ddot{Q}_{\phi_1\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= 0,\\
\mathrm{E}[-\ddot{Q}_{\eta_1\eta_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{nr}{2\hat{\eta}_1^2} + \frac{r}{\hat{\eta}_1^3}\,\mathrm{tr}\left(\hat{\mathbf{R}}^{-1}\hat{\Omega}\right) + \frac{r}{\hat{\eta}_1}\,\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right),\\
\mathrm{E}[-\ddot{Q}_{\eta_2\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_2}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_2}\right) + \frac{r}{2}\,\mathrm{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial^2\hat{\mathbf{R}}}{\partial\eta_2\eta_2}\right)\\
&\quad + \frac{r}{2\hat{\eta}_1}\,\mathrm{tr}\left[\frac{\partial^2\hat{\mathbf{R}}^{-1}}{\partial\eta_2\eta_2}\left(\hat{\Omega} + \hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right],\\
\mathrm{E}[-\ddot{Q}_{\eta_1\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{r}{2\hat{\eta}_1^2}\,\mathrm{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_2}\left(\hat{\Omega} + \hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right],\\
\mathrm{E}[-\ddot{Q}_{\eta_1\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{r}{2\hat{\eta}_1^2}\,\mathrm{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_3}\left(\hat{\Omega} + \hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right],\\
\mathrm{E}[-\ddot{Q}_{\eta_2\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_3}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_2}\right) + \frac{r}{2}\,\mathrm{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial^2\hat{\mathbf{R}}}{\partial\eta_2\eta_3}\right)\\
&\quad + \frac{r}{2\hat{\eta}_1}\,\mathrm{tr}\left[\frac{\partial^2\hat{\mathbf{R}}^{-1}}{\partial\eta_2\eta_3}\left(\hat{\Omega} + \hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right],\\
\mathrm{E}[-\ddot{Q}_{\eta_3\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{r}{2}\,\mathrm{tr}\left(\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_3}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_3}\right) + \frac{r}{2}\,\mathrm{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial^2\hat{\mathbf{R}}}{\partial\eta_3\eta_3}\right)\\
&\quad + \frac{r}{2\hat{\eta}_1}\,\mathrm{tr}\left[\frac{\partial^2\hat{\mathbf{R}}^{-1}}{\partial\eta_3\eta_3}\left(\hat{\Omega} + \hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right].
\end{aligned}
$$

More details are given in Appendix B.1.

Because the matrix $\mathrm{E}[\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]$ is always block-diagonal with respect to the parameters $\boldsymbol{\beta}$, η and ϕ_1, the modified Q-function-based Cook's statistic $QD^1_{i\theta}$, for $i = 1, \ldots, r$, can thus be decomposed into three components

$$QD^1_{i\theta} = QD^1_{i\beta} + QD^1_{i\eta} + QD^1_{i\phi_1},$$

where $QD^1_{i\beta}$, $QD^1_{i\eta}$ and $QD^1_{i\phi_1}$ are the modified Q-function-based Cook's statistics corresponding to the fixed effects $\boldsymbol{\beta}$, the between subject covariance components $\boldsymbol{\eta}$ and the within-subject covariance components ϕ_1, respectively.

5.4.2 Diagnostics at observation level

For the model in study, we have two levels of responses, namely, subjects and repeated measures/observations. Intuitively, an influential subject may or may not contain influential observations.

Observation-deletion

$$\begin{aligned} D^1_{[ij]\theta} &= [\dot{\mathcal{L}}_{[ij]}(\hat{\boldsymbol{\theta}})]^\top \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}[\dot{\mathcal{L}}_{[ij]}(\hat{\boldsymbol{\theta}})] \\ &= [U_{[ij]}(\hat{\boldsymbol{\theta}})]^\top \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}[U_{[ij]}(\hat{\boldsymbol{\theta}})], \end{aligned}$$

for $i = 1, \ldots, r$ and $j = 1, \ldots, n$, where $[ij]$ is to designate the j-th observation of the i-th subject is deleted from the dataset, $U_{[ij]}(\hat{\boldsymbol{\theta}}) = \partial \mathcal{L}_{[ij]}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and

$$\begin{aligned} \mathcal{L}_{[ij]}(\hat{\boldsymbol{\theta}}) &= -\frac{(nr-1)}{2}\log(2\pi) - \frac{(r-1)}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2} log|\boldsymbol{\Sigma}_{[j]}| \\ &\quad -\frac{1}{2}\sum_{\substack{k=1 \\ k\neq i}}^{r} (\mathbf{Y}_k - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_k - \mathbf{X}\boldsymbol{\beta}) \\ &\quad -\frac{1}{2}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}_{[j]}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\boldsymbol{\beta}). \end{aligned}$$

Thus,

$$
\begin{aligned}
U_{[ij]}(\hat{\boldsymbol{\beta}}) &= \frac{\partial \mathcal{L}_{[ij]}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{X}_{[j]}^\top \boldsymbol{\Sigma}_{[j]}^{-1}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}}), \\
U(\hat{\boldsymbol{\phi}}) &= \frac{\partial \mathcal{L}_{[ij]}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} = \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \hat{\boldsymbol{\Sigma}}}{\partial \phi_l}\right) - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{[j]}^{-1}\frac{\partial \hat{\boldsymbol{\Sigma}}_{[j]}}{\partial \phi_l}\right) \\
&\quad \frac{1}{2}(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \frac{\partial \hat{\boldsymbol{\Sigma}}^{-1}}{\partial \phi_l}(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&\quad -\frac{1}{2}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})^\top \frac{\partial \hat{\boldsymbol{\Sigma}}_{[j]}^{-1}}{\partial \phi_l}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}}),
\end{aligned}
$$

for $l = 1, \ldots, 4$, or in vec notation

$$
\begin{aligned}
U(\hat{\boldsymbol{\phi}}) &= \frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \phi}\operatorname{vec}(\boldsymbol{\Sigma}^{-1}) \\
&\quad -\frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}}\operatorname{vec}(\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}^{-1}) \\
&\quad -\frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma}_{[j]})}{\partial \phi}\operatorname{vec}(\boldsymbol{\Sigma}_{[j]}^{-1}) \\
&\quad +\frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma}_{[j]})}{\partial \boldsymbol{\phi}}\operatorname{vec}(\boldsymbol{\Sigma}_{[j]}^{-1}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_{[j]}^{-1}),
\end{aligned}
$$

where $\mathbf{X}_{[j]}$ is an $(n-1) \times p$ matrix in which the j-th row of $\mathbf{X}$ is deleted, $\mathbf{Y}_{i[j]}$ is $\mathbf{Y}_i$ without the j-th observation, $\hat{\mathbf{R}}_{[j]}$ and $\hat{\boldsymbol{\Sigma}}_{[j]}$ are $(n-1) \times (n-1)$ matrices.

The modified Q-function-based Cook's statistic is of the form

$$
QD^1_{[ij]\theta} = [\dot{Q}_{[ij]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^\top \{-\operatorname{E}[\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]\}^{-1}[\dot{Q}_{[ij]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})],
$$

where $\dot{Q}_{[ij]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ has the elements

$$
\begin{aligned}
\dot{Q}_{[ij]\beta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{1}{\hat{\phi}_1}\mathbf{X}^\top\hat{\mathbf{r}}_i + \frac{1}{\hat{\phi}_1}\mathbf{X}_{[j]}^\top\hat{\mathbf{r}}i[j],\\
\dot{Q}_{[ij]\phi_1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{1}{2\hat{\phi}_1}\hat{\mathbf{r}}_i^\top\hat{\mathbf{r}}_i - \frac{\operatorname{tr}(\Omega^*)}{2\hat{\phi}_1^2} + \frac{1}{2\hat{\phi}_1^2}\hat{\mathbf{r}}_{i[j]}^\top\hat{\mathbf{r}}_{i[j]} + \frac{1}{2\hat{\phi}_1^2}\operatorname{tr}(\hat{\Omega}_{[j]}) + \frac{1}{2\hat{\phi}_1},\\
\dot{Q}_{[ij]\eta_1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{1}{2\hat{\eta}_1^2}\operatorname{tr}\left[\hat{\mathbf{R}}^{-1}(\hat{\Omega}+\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top)\right] + \frac{1}{2\hat{\eta}_1^2}\\
&\quad +\frac{1}{2\hat{\eta}_1^2}\operatorname{tr}\left[\hat{\mathbf{R}}_{[j]}^{-1}(\hat{\Omega}_{[j]}+\hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top)\right],\\
\dot{Q}_{[ij]\eta_2}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_2}\right) + \frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\hat{\eta}_2}(\hat{\Omega}+\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top)\right]\\
&\quad -\frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1}\frac{\partial\hat{\mathbf{R}}_{[j]}}{\partial\eta_2}\right) - \frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}_{[j]}^{-1}}{\partial\hat{\eta}_2}(\hat{\Omega}_{[j]}+\hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top)\right],\\
\dot{Q}_{[ij]\eta_3}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_3}\right) + \frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\hat{\eta}_3}(\hat{\Omega}+\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top)\right]\\
&\quad -\frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1}\frac{\partial\hat{\mathbf{R}}_{[j]}}{\partial\eta_2}\right) - \frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}_{[j]}^{-1}}{\partial\hat{\eta}_3}(\hat{\Omega}_{[j]}+\hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top)\right],
\end{aligned}
$$

where $\mathbf{X}_{[j]}$ is an $(n-1)\times p$ matrix in which the j-th row of $\mathbf{X}$ is deleted, $\hat{\mathbf{r}}_{i[j]} = \mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}} - \hat{\mathbf{b}}_{i[j]}$, $\mathbf{Y}_{i[j]}$ is $\mathbf{Y}_i$ without the j-th observation, $\hat{\mathbf{b}}_{i[j]} = \hat{\eta}_1\hat{\mathbf{R}}_{[j]}\hat{\boldsymbol{\Sigma}}_{[j]}^{-1}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})$, $\hat{\mathbf{R}}_{[j]}$, $\hat{\boldsymbol{\Sigma}}_{[j]}$ and $\hat{\Omega}_{[j]}$ are $(n-1)\times(n-1)$ matrices. Details are given in Appendix .

Site-deletion

Since we have the same number of observations for each site, we can remove an entire location. Following Cook's approach we have

$$
D_{i\theta} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[j]})^\top\mathbf{F}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[j]}),
$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ for the postulated model and $\hat{\boldsymbol{\theta}}_{[j]}$ is the maximum likelihood estimator of $\boldsymbol{\theta}_{[j]} = (\boldsymbol{\beta}_{[j]}^\top, \boldsymbol{\phi}_{[j]}^\top)^\top$, without the j-th individual, i.e, all the observations for a specific location for our case.

To calculate $\hat{\boldsymbol{\theta}}_{[j]}$, we may use a one-step Newton Raphson approximation at $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}_{[j]} = \hat{\boldsymbol{\theta}} + [-\ddot{\mathcal{L}}_{[j]}(\hat{\boldsymbol{\theta}})]^{-1}\dot{\mathcal{L}}_{[j]}(\hat{\boldsymbol{\theta}})$, where the dots over the functions denote derivatives with respect to $\boldsymbol{\theta}$. And to overcome some computational difficulty, using $\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})$ to replace $\ddot{\mathcal{L}}_{[j]}(\hat{\boldsymbol{\theta}})$, and so of the expectation of it. So that the approximated Cook's distance becomes

$$D^1_{[j]\theta} = [U_{[j]}(\hat{\boldsymbol{\theta}})]^\top \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1}[U_{[j]}(\hat{\boldsymbol{\theta}})],$$

for $j = 1, \ldots, n$, for all sites, where $U_{[j]}(\hat{\boldsymbol{\theta}}) = (U_{[j]}(\hat{\boldsymbol{\beta}}), U_{[j]}(\hat{\boldsymbol{\phi}}))^\top$ is the score function formula without the j-th location evaluated at $\hat{\boldsymbol{\theta}}$ (given below). We can decompose $D^1_{[j]\theta} = D^1_{[j]\beta} + D^1_{[j]\phi}$, which means that the diagnostic measure of $\boldsymbol{\theta}$ is the sum of the diagnostic measures of the fixed effects $\boldsymbol{\beta}$ and the variance components $\boldsymbol{\phi}$ in terms of $D^1_{[j]\theta}$, for $j = 1, \ldots, n$,, where $D^1_{[j]\beta} = [U_{[j]}(\hat{\boldsymbol{\beta}})]^\top \mathbf{F}^{-1}_{\beta\beta}[U_{[j]}(\hat{\boldsymbol{\beta}})]$, $D^1_{[j]\phi} = [U_{[j]}(\hat{\boldsymbol{\phi}})]^\top \mathbf{F}^{-1}_{\phi\phi}[U_{[j]}(\hat{\boldsymbol{\phi}})]$.

$$\begin{aligned} U_{[j]}(\hat{\boldsymbol{\beta}}) &= \sum_{i=1}^{r} \mathbf{X}^\top_{[j]} \hat{\boldsymbol{\Sigma}}^{-1}_{[j]} (\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}}), \\ U_{[j]}(\hat{\boldsymbol{\phi}}_l) &= -\frac{r}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}_{[j]} \frac{\partial \hat{\boldsymbol{\Sigma}}_{[j]}}{\partial \phi_l} \right) - \frac{1}{2} \sum_{i=1}^{r} (\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})^\top \frac{\partial \hat{\boldsymbol{\Sigma}}^{-1}_{[j]}}{\partial \phi_l} (\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}}), \end{aligned}$$

for $l = 1, \ldots, 4$, or in vec notation

$$\begin{aligned} U_{[j]}(\hat{\boldsymbol{\phi}}) &= -\frac{r}{2} \frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma}_{[j]})}{\partial \phi} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}_{[j]}) \\ &\quad + \frac{1}{2} \frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma}_{[j]})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}_{[j]} (\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}^{-1}_{[j]}), \end{aligned}$$

where $\mathbf{X}_{[j]}$ is an $(n-1) \times p$ matrix in which the j-th row of $\mathbf{X}$ is deleted, $\mathbf{Y}_{i[j]}$ is $\mathbf{Y}_i$ without the j-th observation, $\hat{\mathbf{R}}_{[j]}$ and $\hat{\boldsymbol{\Sigma}}_{[j]}$ are $(n-1) \times (n-1)$ matrices.

The modified Q-function-based Cook's statistic is of the form

$$QD^1_{[j]\theta} = [\dot{Q}_{[j]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^\top \{-\operatorname{E}[\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]\}^{-1} [\dot{Q}_{[j]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})],$$

for $j = 1, \ldots, n$ where,

$$\dot{Q}_{[j]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{1}{\hat{\phi}_1} \sum_{i=1}^{r} \mathbf{X}_{[j]}^\top \hat{\mathbf{r}}_{i[j]} \\ \frac{r}{2\hat{\phi}_1^2} \operatorname{tr}(\hat{\Omega}_{[j]}) + \frac{1}{2\hat{\phi}_1^2} \sum_{i=1}^{r} \hat{\mathbf{r}}_{i[j]}^\top \hat{\mathbf{r}}_{i[j]} - \frac{(n-1)r}{2\hat{\phi}_1} \\ -\frac{(n-1)r}{2\hat{\eta}_1} + \frac{1}{2\hat{\eta}_1^2} \sum_{i=1}^{r} \operatorname{tr}\left[\hat{\mathbf{R}}_{[j]}^{-1}\left(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top\right)\right] \\ -\frac{r}{2} \operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1} \frac{\partial \hat{\mathbf{R}}_{[j]}}{\partial \eta_2}\right) - \frac{1}{2\hat{\eta}_1} \sum_{i=1}^{r} \operatorname{tr}\left[\frac{\partial \hat{\mathbf{R}}_{[j]}^{-1}}{\partial \eta_2}\left(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top\right)\right] \\ -\frac{r}{2} \operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1} \frac{\partial \hat{\mathbf{R}}_{[j]}}{\partial \eta_3}\right) - \frac{1}{2\hat{\eta}_1} \sum_{i=1}^{r} \operatorname{tr}\left[\frac{\partial \hat{\mathbf{R}}_{[j]}^{-1}}{\partial \eta_3}\left(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top\right)\right] \end{pmatrix}.$$

The Appendix C.1 shows more details.

5.4.3 Cutoff value for influential cases

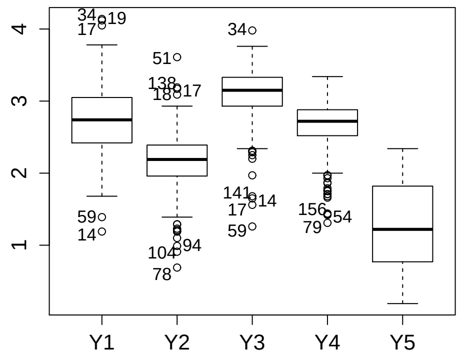
Cook and Weisberg (1982) indicated that the Cook's distance can be compared with a χ-squared distribution with an appropriate degree of freedom for calibration, and also for the Cook's distance approximated. Zhu et al. (2007) suggested to use $(p+q)/r$ for models with missing data as a rough cutoff value for calibrating the Q-function-based Cook's statistics QD_i, where $(p+q)$ is the dimension of the parameter vector $\boldsymbol{\theta}$ and r is the number of repetitions. We also suggest to use $2 * QD_i$ as a cutoff value.

5.5 Applications

5.5.1 Productivity data from the year 1998 to 2002

The data set were first analyzed by De Bastiani et al. (2016) and consist of 253 observations in each year from 1998 to 2002. The data consist of soybean productivity data and four chemical contents considered as explanatory variables were collected in a grid of $7.20 \times 7.20m$ an experimential area with $1.33ha$. The chemical contents of soil are phosphorus (P)[$mg.dm^{-3}$], potassium (K)[$cmolc.dm^{-3}$], calcium (Ca)[$cmolc.dm^{-3}$] and magnesium (Mg)[$cmolc.dm^{-3}$]. The observations were taken at the same site for each repetition. Figure 5.1 show the usual boxplots for the productivity, where only the observations taken in the harvest year 2002 do not have outliers.

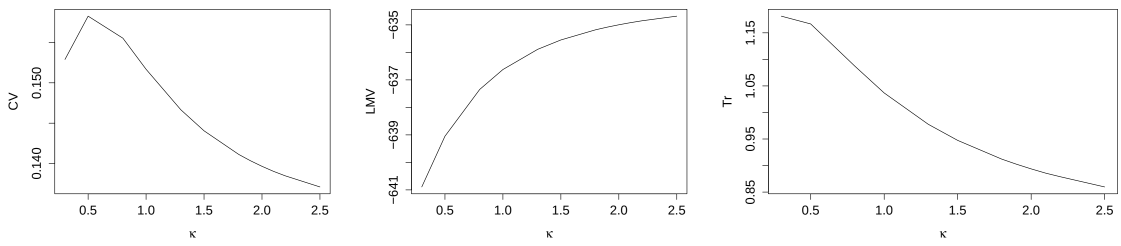
Figure 5.1: Boxplot for soybean productivity in the years 1998 (Y_1), 1999 (Y_2), 2000 (Y_3), 2001 (Y_4) and 2002 (Y_5).



Source: From the author.

In other to choose the variance covariance structure that best describe the spatial dependence of the soybean productivity, we used cross validation (CV), the log-likelihood maximum value (LMV) and the trace of the asymptotic covariance matrix of an estimated mean (Tr) criteria. These criteria lead us to select the Gaussian covariance function (Figure 5.2).

Figure 5.2: Criteria to choose the variance covariance structure (a) Cross validation (CV), (b) The log-likelihood maximum value (LMV) and (c) the trace of the asymptotic covariance matrix of an estimated mean (Tr).



(a) (b) (c)

Source: From the author.

Table 5.1 shows the parameters estimates considering the Gaussian covariance function, and the respective asymptotic standard errors (se) in parenthesis. We used the Z-test, which has normal asymptotic distribution, to test the hypothesis of the form H_0: $\beta_k = 0$ versus H_1: $\beta_k \neq 0$, where β_k is any of the parameters of the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$. According to Table 5.1, $\hat{\beta}_0$ is significant. Considering the same idea to test the ϕ's, we find that they are significants. The macronutrients P and K are needed in relatively large quantities in the soil compared to others to prevent plant deficiencies, but after a critical level is reached there is no additional yield increment.

Table 5.1: Parameters estimates for Coodetec data considering the Gaussian covariance function, the asymptotic standard errors (in parenthesis) and the p-values.

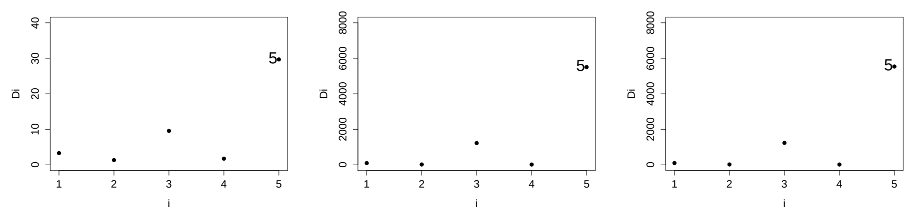
Intercept	P	K	Ca	Mg	nugget	sill	f(range)
$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
2.4013**	-0.0017	0.3270	0.0118	-0.0592	0.1927**	0.0579**	0.0407**
(0.1020)	(0.0109)	(0.1746)	(0.0392)	(0.0515)	(0.0178)	(0.0219)	(0.0098)

** significant for 5%

5.5.2 Global influence

First we consider case-deletion which allows to evaluate the individual contribution of each repetition in the process of estimation. Each repetition correspond to a year. Figure 5.3 shows Cook's distance for one-step approximation. It shows that the subject #5, that correspond to the year 2002/2003 has more influence for both $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$.

Figure 5.3: Cook's distance using one-step approximation for elimination the i-th year, a)D^1_β, b) D^1_ϕ and c)D^1_i.

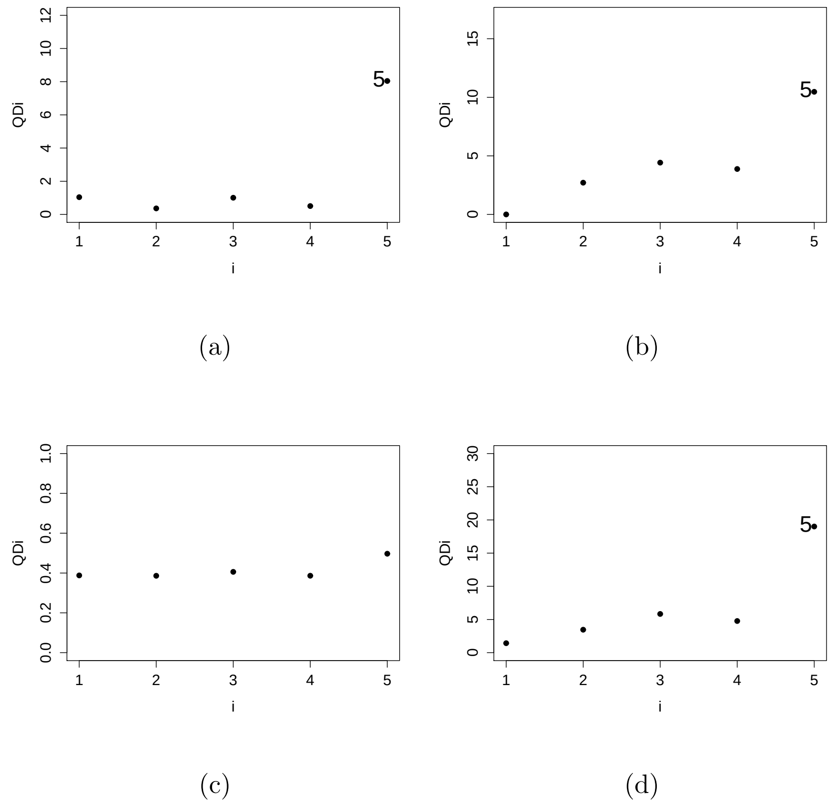


(a) (b) (c)

Source: From the author.

Figure 5.4 shows Cook's distance based on Q-function for case-deletion and also highlights the year #5, nevertheless this subject is not pointed as influential for the η's parameters, according Figure 5.4(c).

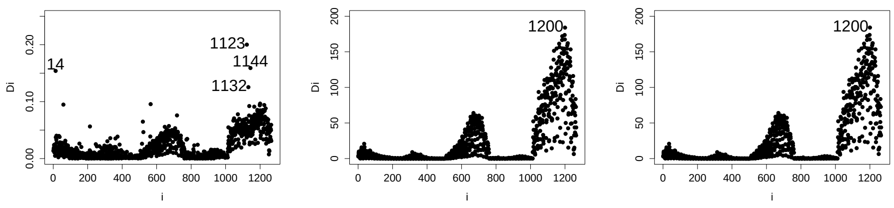
Figure 5.4: Q-function based distance using one-step approximation for elimination the i-th year, a) QD^1_{β}, b) $QD^1_{\phi_1}$, c) QD^1_{η} and d) QD^1_i.



Source: From the author.

Figure 5.5 and 5.6 shows Cook's distance based on likelihood and Q-function, respectively, for observational level when removing each observation from the data set. Both figures highlight observations from the year 2002/2003, the latest one. The observation #14 is from the the year 1998/1999 and correspond to a value 1.19 *ha*, which is the minimum value in this year. The observation #1200 has value 0.98 *ha*, below the mean, and below of the values of this location for the other year that were 2.05, 2.48, 3.62 and 2.65 *ha*.

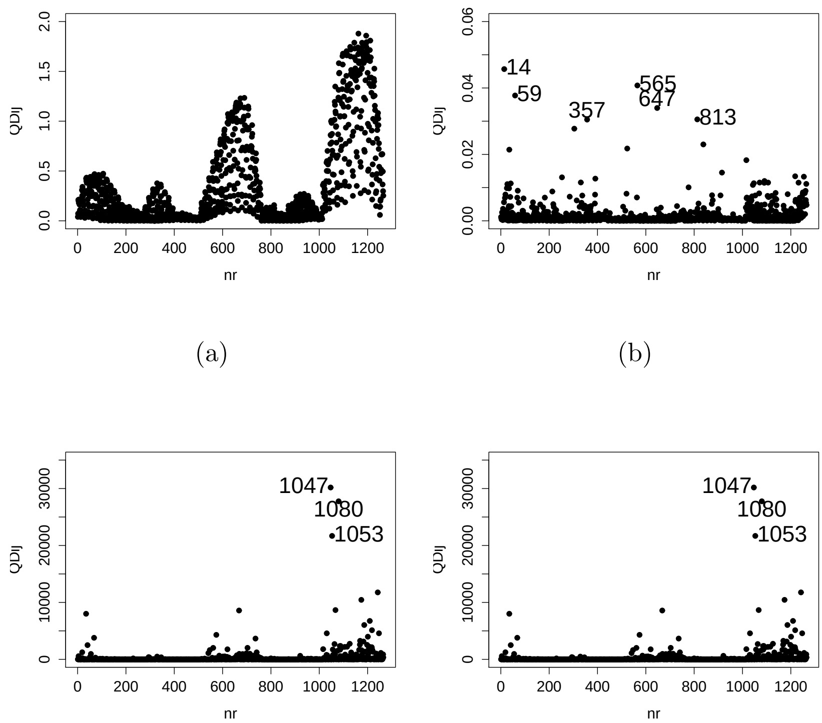
Figure 5.5: Cook's distance using one-step approximation for eliminating the j-th observation from the i-th year, a)$D^1_{ij\beta}$, b) $D^1_{ij\phi_1}$ and d)D^1_{ij}.



(a) (b) (c)

Source: From the author.

Figure 5.6: Q-function based distance using one-step approximation for eliminating the j-th observation from the i-th year, a)QD^1_{β}, b) $QD^1_{\phi_1}$, c) QD^1_{η} and d)QD^1_{i}.



(a) (b)

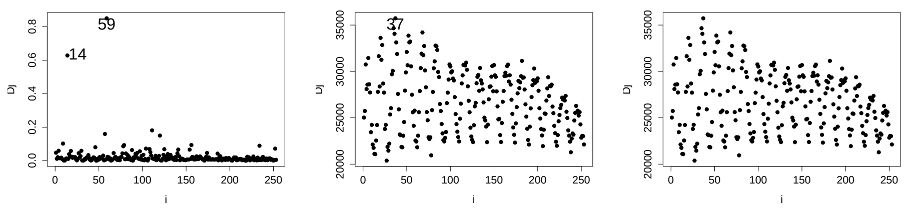
(c) (d)

Source: From the author.

Figure 5.7 shows Cook's distance for removing a location. We can see that the individuals #14 and #59 as influential for the β parameters. The individual #14 has productivity values 1.19, 2.06, 1.65, 1.71 and 1.27 from years 1998/1999 to 2002/2003, respectively. And individual #59 has productivity values 1.39, 1.52, 1.26, 1.68 and 0.26. Figure 5.6

shows Cook's distance based on Q-function for observation level in the case that a location is eliminated. Figure 5.8 shows Cook's distance based on the Q-function for the location deletion, where the location #59 was detected as influential for the β's and ϕ_1, and locations #35, #41 and #68 as influential on the η's. Location #35 correspond to 2.00, 2.40, 2.93, 3.01 and 2.26 ha from years 1998/1999 to 2002/2003, #41 correspond to the productivity values 1.97„ 1.89, 3.26, 2.92, 2.33 ha and #68 corresponds to 2.08, 2.41, 3.01, 2.85, 0.89 ha.

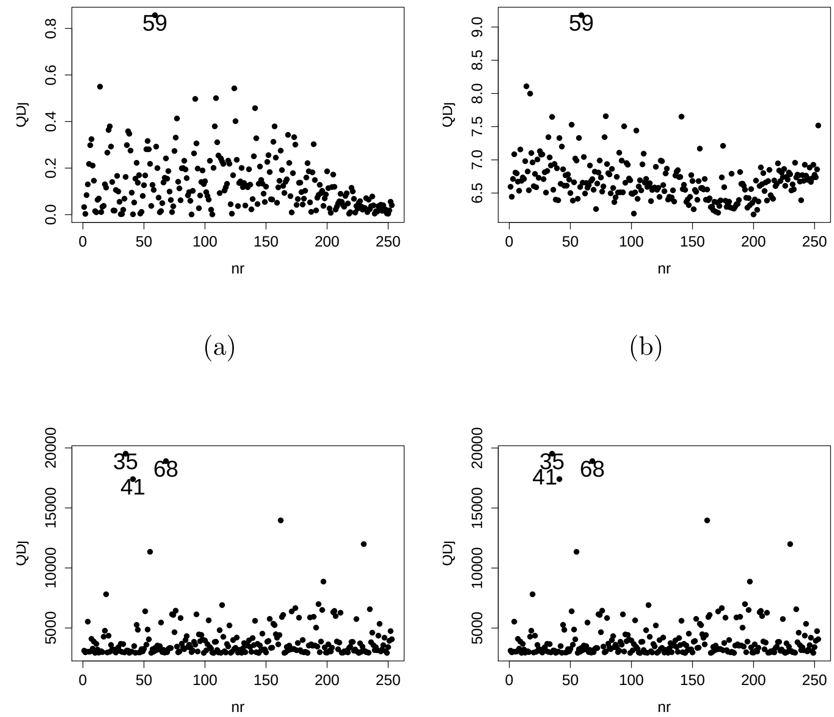
Figure 5.7: Cook's distance for the elimitation the j-th location, a) $D_{j\beta}$, b) $D_{j\phi}$ and b) D_j.



(a) (b) (c)

Source: From the author.

Figure 5.8: Cook's distance based on Q-function for the elimitation the j-th location, a) $QD_{j\beta}$, b) $QD_{j\phi}$ and b) QD_j.



(a) (b)

(c) (d)

Source: From the author.

5.6 Conclusions

In this Chapter we investigated diagnostics techniques in the same model presented in previous chapters, i.e., in the Gaussian spatial linear models with repetitions. Now, we stress the global diagnostics in this models. We presented the complete loglikelihood and the Q-function. A brief description of the maximum likelihood estimation and asymptotic standard errors calculations were shown.

We discussed some concepts of global influence based on the likelihood and on the Q-function. We carried out an application to a soybean productivity data set. This diagnostics tool permit to evaluate in more details in which parameter the observations, sites or subjects are influential. Another application on environmental science is underway. In the new application we remove the temporal effect and analyze the data considering independent repetitions, which allow us to use global diagnostics techniques.

The exposition of this approach (with the Q function) on the study of global diagnostics in Gaussian spatial models in the chapter is new and we believe that it has the potential for a wide impact in applied statistics.

Appendix C

C.1 The Q-function and derivatives for Gaussian spatial linear model

a) Linear mixed model

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}_i + \tau_i \text{ for } i = 1, \ldots, r$$

where $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}) \perp \tau_i \sim N(\mathbf{0}, \mathbf{D}_i)$.

b) Joint distribution

$$\begin{pmatrix} \mathbf{b}_i \\ \mathbf{Y}_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{X}_i\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \mathbf{G} & \mathbf{G}\mathbf{Z}_i^\top \\ \mathbf{Z}_i\mathbf{G} & \boldsymbol{\Sigma}_i \end{pmatrix}\right],$$

em que $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{D}_i$, $i = 1, \ldots, r$,

$$\begin{aligned} \mathrm{Cov}(\mathbf{b}_i, \mathbf{Y}_i) &= \mathrm{Cov}(\mathbf{b}_i, \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}_i + \epsilon_i) \\ &= \mathrm{Var}(\mathbf{b}_i)\mathbf{Z}_i^\top \\ &= \mathbf{G}\mathbf{Z}_i^\top \end{aligned}$$

c) Conditional distribution

According to Osorio (2006) the conditional distribution is given by

$$\mathbf{b}_i|\mathbf{Y}_i \sim N\left(\mathbf{G}\mathbf{Z}_i^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}), \Omega_i\right), \; i = 1\ldots, r,$$

where $\Omega_i = \mathbf{G} - \mathbf{G}\mathbf{Z}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i\mathbf{G}$.

d) Complete loglikelihood

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{i=1}^{r} \mathcal{L}_{ic}(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ are the unknown parameters and

$$\begin{aligned}\mathcal{L}_{ic}(\boldsymbol{\theta}) \;\; = \;\; & -\frac{1}{2}\log|\mathbf{D}_i| - \frac{1}{2}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\right)^\top \mathbf{D}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\right) \\ & -\frac{1}{2}\log|\mathbf{G}| - \frac{1}{2}\mathbf{b}_i^\top \mathbf{G}^{-1}\mathbf{b}_i\end{aligned}$$

e) Q-function

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \sum_{i=1}^{r} \mathrm{E}(\mathcal{L}_{ic}(\boldsymbol{\theta})|\mathbf{Y}, \boldsymbol{\theta}^*) = \sum_{i=1}^{r} Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*),$$

$$\begin{aligned}Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*) \;\; = \;\; & -\frac{1}{2}\log|\mathbf{D}_i| - \frac{1}{2}\,\mathrm{tr}(\mathbf{D}_i^{-1}\mathbf{Z}_i\Omega_i^*\mathbf{Z}_i^\top) - \frac{1}{2}\mathbf{r}_i^{*\top}\mathbf{D}_i^{-1}\mathbf{r}_i^* \\ & -\frac{1}{2}\log|\mathbf{G}| - \frac{1}{2}\,\mathrm{tr}(\mathbf{G}^{-1}\Omega_i^*) - \frac{1}{2}\mathbf{b}_i^{*\top}\mathbf{G}^{-1}\mathbf{b}_i^*, \\ = \;\; & -\frac{1}{2}\log|\mathbf{D}_i| - \frac{1}{2}\,\mathrm{tr}(\mathbf{D}_i^{-1}\mathbf{Z}_i\Omega_i^*\mathbf{Z}_i^\top) - \frac{1}{2}\mathbf{r}_i^{*\top}\mathbf{D}_i^{-1}\mathbf{r}_i^* \\ & -\frac{1}{2}\log|\mathbf{G}| - \frac{1}{2}\,\mathrm{tr}[\mathbf{G}^{-1}(\Omega_i^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top})],\end{aligned}$$

where $\boldsymbol{\theta}^*$ is the vector of parameters estimate of $\boldsymbol{\theta}$ = from previous iteration of the algorithm.

$$\begin{aligned}
\mathbf{r}_i^* &= \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i^* \\
\mathbf{b}_i^* &= \mathrm{E}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}^*) = \mathbf{G}^*\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{*-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}^*) \\
\Omega_i^* &= \mathrm{Cov}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}^*) = \mathbf{G}^* - \mathbf{G}^*\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{*-1}\mathbf{Z}_i\mathbf{G}^*
\end{aligned}$$

f) Q-function for the spatial linear mixed model

For the spatial linear model we consider $\mathbf{X}_i = \mathbf{X}$, $\mathbf{Z}_i = \mathbf{I}$, $\phi_1 = \phi_1$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^\top = (\phi_2, \phi_3, \phi_4)^\top$, so $\mathbf{R} = \mathbf{R}(\eta_2, \eta_3)$, $\mathbf{D}_i = \phi_1\mathbf{I}$, $\mathbf{G} = \mathbf{G}(\boldsymbol{\eta}) = \eta_1\mathbf{R}$, and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \phi_1\mathbf{I} + \eta_1\mathbf{R}$, so $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi_1, \boldsymbol{\eta}^\top)^\top$ which results on the particular case of the Q-function for the spatial linear mixed model given by

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \sum_{i=1}^{r} Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*),$$

$$\begin{aligned}
Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = & -\frac{n}{2}\log\phi_1 - \frac{1}{2\phi_1}\,\mathrm{tr}(\Omega^*) - \frac{1}{2\phi_1}\mathbf{r}_i^{*\top}\mathbf{r}_i^* \\
& -\frac{n}{2}\log\eta_1 - \frac{1}{2}\log|\mathbf{R}| - \frac{1}{2\eta_1}\,\mathrm{tr}[\mathbf{R}^{-1}(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top})],
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{r}_i^* &= \mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta} - \mathbf{b}_i^* \\
\mathbf{b}_i^* &= \mathrm{E}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}^*) = \eta_1^*\mathbf{R}^*\boldsymbol{\Sigma}^{*-1}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}^*) \\
\Omega_i^* &= \Omega^* = \mathrm{Cov}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}^*) = \eta_1^*\mathbf{R}^* - \eta_1^{*2}\mathbf{R}^*\boldsymbol{\Sigma}^{*-1}\mathbf{R}^*,
\end{aligned}$$

so on

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = & -\frac{nr}{2}\log\phi_1 - \frac{r}{2\phi_1}\,\mathrm{tr}(\Omega^*) - \frac{1}{2\phi_1}\sum_{i=1}^{r}\mathbf{r}_i^{*\top}\mathbf{r}_i^* \\
& -\frac{nr}{2}\log\eta_1 - \frac{r}{2}\log|\mathbf{R}| - \frac{1}{2\eta_1}\sum_{i=1}^{r}\mathrm{tr}[\mathbf{R}^{-1}(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top})],
\end{aligned} \tag{C.1}$$

g) First-order derivative of Q-function for the spatial linear mixed model

Based on (C.1) the first-order derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi_1, \eta_1, \eta_2, \eta_3)^\top$ is given by

$$\dot{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \sum_{i=1}^{r} \dot{Q}_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*),$$

with elements

$$\dot{Q}_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \begin{pmatrix} \dot{Q}_{i\beta} \\ \dot{Q}_{i\phi_1} \\ \dot{Q}_{i\eta_1} \\ \dot{Q}_{i\eta_2} \\ \dot{Q}_{i\eta_3} \end{pmatrix} = \begin{pmatrix} \frac{1}{\phi_1}\mathbf{X}^\top \mathbf{r}_i^* \\ \frac{1}{2\phi_1^2}\operatorname{tr}(\Omega^*) + \frac{1}{2\phi_1^2}\mathbf{r}_i^{*\top}\mathbf{r}_i^* - \frac{n}{2\phi_1} \\ -\frac{n}{2\eta_1} + \frac{1}{2\eta_1^2}\operatorname{tr}\left[\mathbf{R}^{-1}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ -\frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_2}\right) - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_2}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ -\frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_3}\right) - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \end{pmatrix},$$

where $\frac{\partial \mathbf{R}^{-1}}{\partial \eta_2} = \frac{\partial \mathbf{R}^{-1}}{\partial \phi_3} = -\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi_3}\mathbf{R}^{-1}$, $\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3} = \frac{\partial \mathbf{R}^{-1}}{\partial \phi_4} = -\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi_4}\mathbf{R}^{-1}$. Details in Appendix A.4 and B.1.

$$\dot{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \begin{pmatrix} \dot{Q}_{\beta} \\ \dot{Q}_{\phi_1} \\ \dot{Q}_{\eta_1} \\ \dot{Q}_{\eta_2} \\ \dot{Q}_{\eta_3} \end{pmatrix} = \begin{pmatrix} \frac{1}{\phi_1}\sum_{i=1}^{r}\mathbf{X}^\top \mathbf{r}_i^* \\ \frac{r}{2\phi_1^2}\operatorname{tr}(\Omega^*) + \frac{1}{2\phi_1^2}\sum_{i=1}^{r}\mathbf{r}_i^{*\top}\mathbf{r}_i^* - \frac{nr}{2\phi_1} \\ -\frac{nr}{2\eta_1} + \frac{1}{2\eta_1^2}\sum_{i=1}^{r}\operatorname{tr}\left[\mathbf{R}^{-1}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ -\frac{r}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_2}\right) - \frac{1}{2\eta_1}\sum_{i=1}^{r}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_2}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ -\frac{r}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_3}\right) - \frac{1}{2\eta_1}\sum_{i=1}^{r}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \end{pmatrix} \tag{C.2}$$

Letting $\boldsymbol{\theta} = \boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$ in which $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$, we have

$$\dot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = [\dot{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} = \mathbf{0},$$

which leads to the following

$$
\begin{cases}
\sum_{i=1}^{r} \mathbf{X}^\top \hat{\mathbf{r}}_i = \mathbf{0} \\
\frac{r}{2\hat{\phi}_1} \operatorname{tr}(\hat{\Omega}) + \frac{1}{2\hat{\phi}_1} \sum_{i=1}^{r} \hat{\mathbf{r}}_i^\top \hat{\mathbf{r}}_i - \frac{nr}{2} = 0 \\
-\frac{nr}{2} + \frac{1}{2\hat{\eta}_1} \sum_{i=1}^{r} \operatorname{tr}\left[\hat{\mathbf{R}}^{-1}\left(\hat{\Omega} + \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i\right)\right] = 0 \\
-\frac{r}{2} \operatorname{tr}\left(\hat{\mathbf{R}}^{-1} \frac{\partial \hat{\mathbf{R}}}{\partial \eta_2}\right) - \frac{1}{2\hat{\eta}_1} \sum_{i=1}^{r} \operatorname{tr}\left[\frac{\partial \hat{\mathbf{R}}^{-1}}{\partial \eta_2}\left(\hat{\Omega} + \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^\top\right)\right] = 0 \\
-\frac{r}{2} \operatorname{tr}\left(\hat{\mathbf{R}}^{-1} \frac{\partial \hat{\mathbf{R}}}{\partial \eta_3}\right) - \frac{1}{2\hat{\eta}_1} \sum_{i=1}^{r} \operatorname{tr}\left[\frac{\partial \hat{\mathbf{R}}^{-1}}{\partial \eta_3}\left(\hat{\Omega} + \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^\top\right)\right] = 0
\end{cases}
\tag{C.3}
$$

h) Second-order derivative of Q-function for the spatial linear mixed model

$$
\ddot{Q}_{i\theta\theta} = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},
$$

with elements

$$
\begin{aligned}
\ddot{Q}_{i\beta\beta} &= -\frac{1}{\phi_1} \mathbf{X}^\top \mathbf{X} \\
\ddot{Q}_{i\beta\phi_1} &= -\frac{1}{\phi_1^2} \mathbf{X}^\top \mathbf{r}_i^* \\
\ddot{Q}_{i\beta\eta_1} &= \mathbf{0} \\
\ddot{Q}_{i\beta\eta_2} &= \mathbf{0} \\
\ddot{Q}_{i\beta\eta_3} &= \mathbf{0} \\
\ddot{Q}_{i\phi_1\phi_1} &= -\frac{1}{\phi_1^3} \operatorname{tr}(\Omega^*) - \frac{1}{\phi_1^3} \mathbf{r}_i^{*\top} \mathbf{r}_i^* + \frac{n}{2\phi_1^2} \\
\ddot{Q}_{i\phi_1\eta_1} &= 0 \\
\ddot{Q}_{i\phi_1\eta_2} &= 0 \\
\ddot{Q}_{i\phi_1\eta_3} &= 0 \\
\ddot{Q}_{i\eta_1\eta_1} &= \frac{n}{2\eta_1^2} - \frac{1}{\eta_1^3} \operatorname{tr}\left[\mathbf{R}^{-1}\left(\Omega^* + \mathbf{b}_i^* \mathbf{b}_i^{*\top}\right)\right] \\
\ddot{Q}_{i\eta_1\eta_2} &= \frac{1}{2\eta_1^2} \operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_2}\left(\Omega^* + \mathbf{b}_i^* \mathbf{b}_i^{*\top}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
\ddot{Q}_{i\eta_1\eta_3} =& \frac{1}{2\eta_1^2}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\
\ddot{Q}_{i\eta_2\eta_2} =& -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \mathbf{R}^{-1}}{\partial \eta_2}\frac{\partial \mathbf{R}}{\partial \eta_2}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial^2 \mathbf{R}}{\partial \eta_2\eta_2}\right) - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial^2 \mathbf{R}^{-1}}{\partial \eta_2\eta_2}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\
\ddot{Q}_{i\eta_2\eta_3} =& -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3}\frac{\partial \mathbf{R}}{\partial \eta_2}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial^2 \mathbf{R}}{\partial \eta_2\eta_3}\right) - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial^2 \mathbf{R}^{-1}}{\partial \eta_2\eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\
\ddot{Q}_{i\eta_3\eta_3} =& -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3}\frac{\partial \mathbf{R}}{\partial \eta_3}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial^2 \mathbf{R}}{\partial \eta_3\eta_3}\right) - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial^2 \mathbf{R}^{-1}}{\partial \eta_3\eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right],
\end{aligned}
$$

where $\frac{\partial^2 \mathbf{R}^{-1}}{\partial \eta_j \partial \eta_k} = -2\frac{\partial \mathbf{R}^{-1}}{\partial \eta_k}\frac{\partial \mathbf{R}}{\partial \eta_j}\mathbf{R}^{-1} - \mathbf{R}^{-1}\frac{\partial^2 \mathbf{R}}{\partial \eta_j \partial \eta_k}\mathbf{R}$. More details are given in Appendix A.3 and B.1.

i) Expectation of $[-\ddot{Q}]$ evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$ for the spatial linear mixed model

By noting that

$$
\begin{aligned}
\mathrm{E}[\mathbf{r}_i^*]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= 0 \\
\mathrm{E}[\mathbf{b}_i^*\mathbf{b}_i^{*\top}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}} \\
\mathrm{E}[\mathbf{r}_i^*\mathbf{r}_i^{*\top}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \hat{\Omega} + \hat{\phi}_1\mathbf{I},
\end{aligned}
$$

and by the fact that

$$
\mathrm{E}[-Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)] = -\sum_{i=1}^{r}\mathrm{E}[Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^*)],
$$

the expectation of $[-\ddot{Q}]$ evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$ for the spatial linear mixed model has the elements

$$
\begin{aligned}
\mathrm{E}[-\ddot{Q}_{\beta\beta}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =& \frac{r}{\hat{\phi}_1}\mathbf{X}^\top\mathbf{X} \\
\mathrm{E}[-\ddot{Q}_{\beta\phi_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =& \mathbf{0} \\
\mathrm{E}[-\ddot{Q}_{\beta\eta_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =& \mathbf{0} \\
\mathrm{E}[-\ddot{Q}_{\beta\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =& \mathbf{0} \\
\mathrm{E}[-\ddot{Q}_{\beta\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =& \mathbf{0} \\
\mathrm{E}[-\ddot{Q}_{\phi_1\phi_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =& \frac{2r}{\hat{\phi}_1^3}\operatorname{tr}(\hat{\Omega}) + \frac{nr}{2\hat{\phi}_1^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[-\ddot{Q}_{\phi_1\eta_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&0\\
\mathrm{E}[-\ddot{Q}_{\phi_1\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&0\\
\mathrm{E}[-\ddot{Q}_{\phi_1\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&0\\
\mathrm{E}[-\ddot{Q}_{\eta_1\eta_1}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&-\frac{nr}{2\hat{\eta}_1^2}+\frac{r}{\hat{\eta}_1^3}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\hat{\Omega}\right)+\frac{r}{\hat{\eta}_1}\operatorname{tr}\left(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\\
\mathrm{E}[-\ddot{Q}_{\eta_2\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&\frac{r}{2}\operatorname{tr}\left(\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_2}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_2}\right)+\frac{r}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial^2\hat{\mathbf{R}}}{\partial\eta_2\eta_2}\right)\\
&+\frac{r}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial^2\hat{\mathbf{R}}^{-1}}{\partial\eta_2\eta_2}\left(\hat{\Omega}+\hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right]\\
\mathrm{E}[-\ddot{Q}_{\eta_1\eta_2}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&-\frac{r}{2\hat{\eta}_1^2}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_2}\left(\hat{\Omega}+\hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right]\\
\mathrm{E}[-\ddot{Q}_{\eta_1\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&-\frac{r}{2\hat{\eta}_1^2}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_3}\left(\hat{\Omega}+\hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right]\\
\mathrm{E}[-\ddot{Q}_{\eta_2\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&\frac{r}{2}\operatorname{tr}\left(\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_3}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_2}\right)+\frac{r}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial^2\hat{\mathbf{R}}}{\partial\eta_2\eta_3}\right)\\
&+\frac{r}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial^2\hat{\mathbf{R}}^{-1}}{\partial\eta_2\eta_3}\left(\hat{\Omega}+\hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right]\\
\mathrm{E}[-\ddot{Q}_{\eta_3\eta_3}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} =&\frac{r}{2}\operatorname{tr}\left(\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\eta_3}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_3}\right)+\frac{r}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial^2\hat{\mathbf{R}}}{\partial\eta_3\eta_3}\right)\\
&+\frac{r}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial^2\hat{\mathbf{R}}^{-1}}{\partial\eta_3\eta_3}\left(\hat{\Omega}+\hat{\eta}_1^2\hat{\mathbf{R}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{R}}\right)\right]
\end{aligned}
$$

j) Q-function and first derivatives for global influence

For the case deletion global influence study we have

$$
\begin{aligned}
Q_{[i]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) \;=\;& -\frac{n(r-1)}{2}\log\phi_1-\frac{(r-1)}{2\phi_1}\operatorname{tr}(\Omega^*)-\frac{1}{2\phi_1}\sum_{\substack{k=1\\k\neq i}}^{r}\mathbf{r}_k^{*\top}\mathbf{r}_k^*\\
&-\frac{n(r-1)}{2}\log\eta_1-\frac{(r-1)}{2}\log|\mathbf{R}|-\frac{1}{2\eta_1}\sum_{\substack{k=1\\k\neq i}}^{r}\operatorname{tr}[\mathbf{R}^{-1}(\Omega^*+\mathbf{b}_k^*\mathbf{b}_k^{*\top})].
\end{aligned}
\tag{C.4}
$$

Similarly to (C.2), the first-order derivative of $\dot{Q}_{[i]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ given in (C.5) and by noting (C.3) we obtain

$$\dot{Q}_{[i]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \begin{pmatrix} -\frac{1}{\phi_1}\mathbf{X}^\top \mathbf{r}_i^* \\ -\frac{1}{2\phi_1^2}\operatorname{tr}(\Omega^*) - \frac{1}{2\phi_1^2}\mathbf{r}_i^{*\top}\mathbf{r}_i^* + \frac{n}{2\phi_1} \\ \frac{n}{2\eta_1} - \frac{1}{2\eta_1^2}\operatorname{tr}\left[\mathbf{R}^{-1}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ \frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_2}\right) + \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_2}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \\ \frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_3}\right) + \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial \mathbf{R}^{-1}}{\partial \eta_3}\left(\Omega^* + \mathbf{b}_i^*\mathbf{b}_i^{*\top}\right)\right] \end{pmatrix}.$$

For the observation-level

$$\begin{aligned} Q_{[ij]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) \quad = \quad & -\frac{(nr-1)}{2}\log_1 - \frac{(r-1)}{2\phi_1}\operatorname{tr}(\Omega^*) - \frac{1}{2\phi_1}\left[\sum_{\substack{k=1\\k\neq i}}^{r}\mathbf{r}_k^{*\top}\mathbf{r}_k^*\right] \\ & -\frac{1}{2\phi_1}\mathbf{r}_{i[j]}^{*\top}\mathbf{r}_{i[j]}^* - \frac{(nr-1)}{2}\log\eta_1 - \frac{(r-1)}{2}\log|\mathbf{R}| - \frac{1}{2}\log|\mathbf{R}_{[j]}| \quad \text{(C.5)} \\ & -\frac{1}{2\eta_1}\sum_{\substack{k=1\\k\neq i}}^{r}\operatorname{tr}[\mathbf{R}^{-1}(\Omega^* + \mathbf{b}_k^{*\top}\mathbf{b}_k^*)] - \frac{1}{2\eta_1}\operatorname{tr}[\mathbf{R}_{[j]}^{-1}(\Omega_{[j]}^* + \mathbf{b}_{i[j]}^{*\top}\mathbf{b}_{i[j]}^*)], \end{aligned}$$

$\dot{Q}_{[ij]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ has elements

$$\begin{aligned} \dot{Q}_{[ij]\beta} \quad &= \quad \frac{1}{\phi_1}\sum_{\substack{k=1\\k\neq i}}^{r}\mathbf{X}^\top\mathbf{r}_i^* + \frac{1}{\phi_1}\mathbf{X}_{[j]}^\top\mathbf{r}_{i[j]}^* \\ \dot{Q}_{[ij]\phi_1} \quad &= \quad -\frac{(nr-1)}{2\phi_1} + \frac{(r-1)}{2\phi_1^2}\operatorname{tr}(\Omega^*) + \frac{1}{2\phi_1^2}\operatorname{tr}(\Omega_{[j]}^*) + \frac{1}{2\phi_1^2}\sum_{\substack{k=1\\k\neq i}}^{r}\mathbf{r}_{i[j]}^{*\top}\mathbf{r}_{i[j]}^* + \frac{1}{2\phi_1}\mathbf{r}_i^{*\top}\mathbf{r}_i^* \\ \dot{Q}_{[ij]\eta_1} \quad &= \quad -\frac{(nr-1)}{2\eta_1} + \frac{1}{2\eta_1^2}\sum_{\substack{k=1\\k\neq i}}^{r}\operatorname{tr}\left[\mathbf{R}^{-1}(\Omega^* + \mathbf{b}_k^{*\top}\mathbf{b}_k^*)\right] + \frac{1}{2\eta_1^2}\operatorname{tr}\left[\mathbf{R}_{[j]}^{-1}(\Omega_{[j]}^* + \mathbf{b}_{i[j]}^{*\top}\mathbf{b}_{i[j]}^*)\right] \\ \dot{Q}_{[ij]\eta_2} \quad &= \quad -\frac{(r-1)}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\eta_2}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}_{[j]}^{-1}\frac{\partial\mathbf{R}_{[j]}}{\partial\eta_2}\right) \\ & \quad -\frac{1}{2\eta_1}\sum_{\substack{k=1\\k\neq i}}^{r}\operatorname{tr}\left[\frac{\partial\mathbf{R}^{-1}}{\partial\eta_2}(\Omega^* + \mathbf{b}_k^{*\top}\mathbf{b}_k^*)\right] - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial\mathbf{R}_{[j]}^{-1}}{\partial\eta_2}(\Omega^* + \mathbf{b}_{i[j]}^{*\top}\mathbf{b}_{i[j]}^*)\right] \\ \dot{Q}_{[ij]\eta_3} \quad &= \quad -\frac{(r-1)}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\eta_3}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}_{[j]}^{-1}\frac{\partial\mathbf{R}_{[j]}}{\partial\eta_3}\right) \\ & \quad -\frac{1}{2\eta_1}\sum_{\substack{k=1\\k\neq i}}^{r}\operatorname{tr}\left[\frac{\partial\mathbf{R}^{-1}}{\partial\eta_3}(\Omega^* + \mathbf{b}_k^{*\top}\mathbf{b}_k^*)\right] - \frac{1}{2\eta_1}\operatorname{tr}\left[\frac{\partial\mathbf{R}_{[j]}^{-1}}{\partial\eta_3}(\Omega^* + \mathbf{b}_{i[j]}^{*\top}\mathbf{b}_{i[j]}^*)\right]. \end{aligned}$$

Evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$ and noting (C.3), we obtain $\dot{Q}_{[ij]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ with elements

$$\begin{aligned}
\dot{Q}_{[ij]\beta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{1}{\hat{\phi}_1}\mathbf{X}^\top\hat{\mathbf{r}}_i + \frac{1}{\hat{\phi}_1}\mathbf{X}_{[j]}^\top\hat{\mathbf{r}}_{i[j]} \\
\dot{Q}_{[ij]\phi_1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{1}{2\hat{\phi}_1}\hat{\mathbf{r}}_i^\top\hat{\mathbf{r}}_i - \frac{\operatorname{tr}(\Omega^*)}{2\hat{\phi}_1^2} + \frac{1}{2\hat{\phi}_1^2}\hat{\mathbf{r}}_{i[j]}^\top\hat{\mathbf{r}}_{i[j]} + \frac{1}{2\hat{\phi}_1^2}\operatorname{tr}(\hat{\Omega}_{[j]}) + \frac{1}{2\hat{\phi}_1} \\
\dot{Q}_{[ij]\eta_1}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= -\frac{1}{2\hat{\eta}_1^2}\operatorname{tr}\left[\hat{\mathbf{R}}^{-1}(\hat{\Omega} + \hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top)\right] + \frac{1}{2\hat{\eta}_1^2}\operatorname{tr}\left[\hat{\mathbf{R}}_{[j]}^{-1}(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top)\right] + \frac{1}{2\hat{\eta}_1^2} \\
\dot{Q}_{[ij]\eta_2}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_2}\right) - \frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1}\frac{\partial\hat{\mathbf{R}}_{[j]}}{\partial\eta_2}\right) + \frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\hat{\eta}_2}(\hat{\Omega} + \hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top)\right] \\
&\quad -\frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}_{[j]}^{-1}}{\partial\hat{\eta}_2}(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top)\right] \\
\dot{Q}_{[ij]\eta_3}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}} &= \frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}^{-1}\frac{\partial\hat{\mathbf{R}}}{\partial\eta_3}\right) - \frac{1}{2}\operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1}\frac{\partial\hat{\mathbf{R}}_{[j]}}{\partial\eta_2}\right) + \frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}^{-1}}{\partial\hat{\eta}_3}(\hat{\Omega} + \hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top)\right] \\
&\quad -\frac{1}{2\hat{\eta}_1}\operatorname{tr}\left[\frac{\partial\hat{\mathbf{R}}_{[j]}^{-1}}{\partial\hat{\eta}_3}(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^\top)\right]
\end{aligned}$$

where $\mathbf{X}_{[j]}$ is an $(n-1)\times p$ matrix in which the j-th row of $\mathbf{X}$ is deleted, $\hat{\mathbf{r}}_{i[j]} = \mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}} - \hat{\mathbf{b}}_{i[j]}$, $\mathbf{Y}_{i[j]}$ is $\mathbf{Y}_i$ without the j-th observation, $\hat{\mathbf{b}}_{i[j]} = \hat{\eta}_1\hat{\mathbf{R}}_{[j]}\hat{\boldsymbol{\Sigma}}_{[j]}^{-1}(\mathbf{Y}_{i[j]} - \mathbf{X}_{[j]}\hat{\boldsymbol{\beta}})$, $\hat{\mathbf{R}}_{[j]}$, $\hat{\boldsymbol{\Sigma}}_{[j]}$ and $\hat{\Omega}_{[j]}$ are $(n-1)\times(n-1)$ matrices.

For the observation level, in the case the we remove a location, i.e. we eliminate the j-th observation for all i, we have that $Q_{[j]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ is of form

$$\begin{aligned}
Q_{[j]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = &-\frac{(n-1)r}{2}\log\phi_1 - \frac{r}{2\phi_1}\operatorname{tr}(\Omega_{[j]}^*) - \frac{1}{2\phi_1}\sum_{i=1}^r \mathbf{r}_{i[j]}^{*\top}\mathbf{r}_{i[j]}^* \\
&-\frac{(n-1)r}{2}\log\eta_1 - \frac{r}{2}\log|\mathbf{R}_{[j]}| - \frac{1}{2\eta_1}\sum_{i=1}^r \operatorname{tr}[\mathbf{R}_{[j]}^{-1}(\Omega_{[j]}^* + \mathbf{b}_{i[j]}^*\mathbf{b}_{i[j]}^{*\top})].
\end{aligned}$$

And

$$\dot{Q}_{[j]}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \begin{pmatrix}
\frac{1}{\phi_1}\sum_{i=1}^r \mathbf{X}_{[j]}^\top\mathbf{r}_{i[j]}^* \\
\frac{r}{2\phi_1^2}\operatorname{tr}(\Omega_{[j]}^*) + \frac{1}{2\phi_1^2}\sum_{i=1}^r \mathbf{r}_{i[j]}^{*\top}\mathbf{r}_{i[j]}^* - \frac{(n-1)r}{2\phi_1} \\
-\frac{(n-1)r}{2\eta_1} + \frac{1}{2\eta_1^2}\sum_{i=1}^r \operatorname{tr}\left[\mathbf{R}_{[j]}^{-1}\left(\Omega_{[j]}^* + \mathbf{b}_{i[j]}^*\mathbf{b}_{i[j]}^{*\top}\right)\right] \\
-\frac{r}{2}\operatorname{tr}\left(\mathbf{R}_{[j]}^{-1}\frac{\partial\mathbf{R}_{[j]}}{\partial\eta_2}\right) - \frac{1}{2\eta_1}\sum_{i=1}^r \operatorname{tr}\left[\frac{\partial\mathbf{R}_{[j]}^{-1}}{\partial\eta_2}\left(\Omega_{[j]}^* + \mathbf{b}_{i[j]}^*\mathbf{b}_{i[j]}^{*\top}\right)\right] \\
-\frac{r}{2}\operatorname{tr}\left(\mathbf{R}_{[j]}^{-1}\frac{\partial\mathbf{R}_{[j]}}{\partial\eta_3}\right) - \frac{1}{2\eta_1}\sum_{i=1}^r \operatorname{tr}\left[\frac{\partial\mathbf{R}_{[j]}^{-1}}{\partial\eta_3}\left(\Omega_{[j]}^* + \mathbf{b}_{i[j]}^*\mathbf{b}_{i[j]}^{*\top}\right)\right]
\end{pmatrix},$$

thus

$$\dot{Q}_{[j]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \dfrac{1}{\hat{\phi}_1}\displaystyle\sum_{i=1}^{r} \mathbf{X}_{[j]}^{\top}\hat{\mathbf{r}}_{i[j]} \\ \dfrac{r}{2\hat{\phi}_1^2}\operatorname{tr}(\hat{\Omega}_{[j]}) + \dfrac{1}{2\hat{\phi}_1^2}\displaystyle\sum_{i=1}^{r}\hat{\mathbf{r}}_{i[j]}^{\top}\hat{\mathbf{r}}_{i[j]} - \dfrac{(n-1)r}{2\hat{\phi}_1} \\ -\dfrac{(n-1)r}{2\hat{\eta}_1} + \dfrac{1}{2\hat{\eta}_1^2}\displaystyle\sum_{i=1}^{r}\operatorname{tr}\left[\hat{\mathbf{R}}_{[j]}^{-1}\left(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^{\top}\right)\right] \\ -\dfrac{r}{2}\operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1}\dfrac{\partial\hat{\mathbf{R}}_{[j]}}{\partial\eta_2}\right) - \dfrac{1}{2\hat{\eta}_1}\displaystyle\sum_{i=1}^{r}\operatorname{tr}\left[\dfrac{\partial\hat{\mathbf{R}}_{[j]}^{-1}}{\partial\eta_2}\left(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^{\top}\right)\right] \\ -\dfrac{r}{2}\operatorname{tr}\left(\hat{\mathbf{R}}_{[j]}^{-1}\dfrac{\partial\hat{\mathbf{R}}_{[j]}}{\partial\eta_3}\right) - \dfrac{1}{2\hat{\eta}_1}\displaystyle\sum_{i=1}^{r}\operatorname{tr}\left[\dfrac{\partial\hat{\mathbf{R}}_{[j]}^{-1}}{\partial\eta_3}\left(\hat{\Omega}_{[j]} + \hat{\mathbf{b}}_{i[j]}\hat{\mathbf{b}}_{i[j]}^{\top}\right)\right] \end{pmatrix},$$

Chapter 6

Gaussian Markov random field within the generalized additive models for location scale and shape

6.1 Abstract

Este capítulo descreve a modelagem e ajuste dos componentes espaciais dos campos aleatórios Gaussianos no enfoque dos modelos aditivos generalizados de locação escala e forma (GAMLSS). Com este enfoque é possível a modelar qualquer ou todos os parâmetros da distribuição da variável resposta considerando variáveis explanatórias e efeitos espaciais. A distribuição da variável resposta pode ou não pertencer a família de distribuições exponenciais. Um novo pacote no *software* `R` foi desenvolvido para colaborar com a proposta. Utilizou-se os campos aleatórios Markovianos Gaussianos para modelar o efeito espacial nos dados de aluguel da cidade de Munich, e para explorar algumas funcionalidades e características dos dados. O potencial sobre fazer análise de dados espaciais usando GAMLSS é discutido. Argumenta-se que a flexibilidade da distribuição paramétrica, a habilidade de modelar todos os parâmetros da distribuição e ferramentas de diagnósticos do GAMLSS fornecem um ambiente adequado para a modelagem espacial.

6.2 Introduction

Since the introduction of the Generalized Additive Models for Location, Scale and Shape (GAMLSS) by Rigby and Stasinopoulos (2005), the models have been used in a variety of different fields such as actuarial science, Heller et al. (2007), biology, biosciences, energy economics, Voudouris et al. (2011), genomics, Khondoker et al. (2007), finance, fisheries, food consumption, growth curves estimation, Borghi et al. (2006) and WHO (2006, 2007), marine research, medicine, meteorology, rainfall, vaccines, film studies, Voudouris et al. (2012), etc.

Discrete spatial variation, where the variables are defined on discrete domains, such as regions, regular grids or lattices, can be modelled by Markov random fields (MRF). MRF can be applied in different areas such as spatial statistics, image analysis, structural time-series analysis, analysis of longitudinal and survival data, spatio-temporal statistics, graphical models and semiparametric models. For applications in spatial econometrics (Anselin and Florax, 1995), in spatial and space-time epidemiology Besag et al. (1991) and Schmid and Held (2001), respectively. Sang and Gelfand (2009) adopted multivariate Markov random field models with temporal dependence to cater for the whole space-time characterization. A comparison between Markov approximations and other methods for large spatial data sets is given by Bolin and Lindgren (2013).

Kunsch (1979) present many important results for Gaussian Markov random fields (GMRF). Extensive theoretical and practical details of GMRF are provided by Rue and Held (2005). In statistics, Besag and Kooperberg (1995) considered the Gaussian intrinsic autoregressive model (IAR), a very important specific case of GMRF models. Breslow and Clayton (1993), Lee and Nelder (2001) and Fahrmeir et al. (2013) incorporated IAR models within generalized linear mixed models (GLMMs). Wood (2006) presents IAR models within a generalized additive model (GAM) framework. There are few papers that use GAMLSS in a spatial framework. Mayr et al. (2012) presented GAMLSS for high-dimensional data based on boosting. Rigby et al. (2013) commented on the paper "Beyond mean regression", Kneib (2013), and presented a simplified analysis of Munich rent data with very few covariates, modelling the μ parameter with a spatial effect using an IAR model term. In this Chapter we describe in detail the theoretical basis of the GAMLSS implementation of GMRF, develop a package in `R` (De Bastiani and Stasinopoulos, 2015; R Core Team, 2015) to achieve this and explore the potential of such modelling using the Munich rent data.

Section 6.3 discusses GAMLSS models and the modelling and fitting of GMRF spatial components within GAMLSS models. In section 6.4 we present the full Munich rent data set, the strategy to choose a model and the results for the Munich rent data set. Section 6.5 investigates the adequacy of the chosen model using residual diagnostic worm plots. The implementation in `R` is described in the Appendix and the `R` code used in the analysis is available from the authors at `www.gamlss.org`. Section 6.6 presents relevant conclusions.

6.3 Methodology

Section 6.3.1 defines the GAMLSS framework, while Section 6.3.2 describes its estimation procedure. Section 6.3.3 describes how the GMRF models can be incorporated within the GAMLSS framework.

6.3.1 The GAMLSS framework

GAMLSS provides a very general and flexible system for modelling a response variable. The distribution of the response variable is selected by the user from a very wide range of distributions available in the `gamlss` package in `R`, Rigby and Stasinopoulos (2005), including highly skewed and kurtotic continuous and discrete distributions. The `gamlss` package includes distributions with up to four parameters, denoted by μ, σ, ν and τ, which usually represent the location (e.g. mean), scale (e.g. standard deviation), and skewness and kurtosis shape parameters, respectively. All the parameters of the response variable distribution can be modelled using parametric and/or nonparametric smooth functions of explanatory variables, thus allowing modelling of the location, scale and shape parameters. Specifically, a GAMLSS model assumes that, for $i = 1, 2, \ldots, n$, independent observations Y_i have probability (density) function $f_Y(y_i|\boldsymbol{\theta}^i)$ conditional on $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i})^\top = (\mu_i, \sigma_i, \nu_i, \tau_i)^\top$ a vector of four distribution parameters, each of which can be a function of the explanatory variables. Rigby and Stasinopoulos (2005) define an original formulation of a GAMLSS model as follows.

For $k = 1, 2, 3, 4$, let $g_k(.)$ be a known monotonic link function relating the distribution parameter $\boldsymbol{\theta}_k = (\theta_{k1}, \ldots, \theta_{kn})^\top$ to predictor $\boldsymbol{\eta}_k = (\eta_{k1}, \ldots, \eta_{kn})^\top$. Then we set

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \tag{6.1}$$

where $\mathbf{X}_k$ is a known design matrix, $\boldsymbol{\beta}_k = (\beta_{1k}, \ldots, \beta_{J'_k k})^\top$ is a parameter vector of length J'_k, h_{jk} is a smooth nonparametric function of variable X_{jk} and the $\mathbf{x}_{jk}$'s are vectors of length n, for $k = 1, 2, 3, 4$ and $j = 1, \ldots, J_k$.

Model (6.1) can be written in a random effects form and random effects can also be included in the model for the $n \times 1$ vectors $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$:

$$\begin{aligned} g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1}\boldsymbol{\gamma}_{j1} \\ g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2}\boldsymbol{\gamma}_{j2} \\ g_3(\boldsymbol{\nu}) &= \boldsymbol{\eta}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3}\boldsymbol{\gamma}_{j3} \\ g_4(\boldsymbol{\tau}) &= \boldsymbol{\eta}_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4}\boldsymbol{\gamma}_{j4}, \end{aligned} \tag{6.2}$$

where here the random effects parameters $\boldsymbol{\gamma}_{jk}$ are assumed to have independent (prior) normal distributions with $\boldsymbol{\gamma}_{jk} \sim N_{q_{jk}}(\mathbf{0}, \lambda_{jk}^{-1}\mathbf{G}_{jk}^{-1})$ and $\mathbf{G}_{jk}^{-1}$ is the (generalized) inverse of a $q_{jk} \times q_{jk}$ symmetric matrix $\mathbf{G}_{jk}$, where if $\mathbf{G}_{jk}$ is singular then $\boldsymbol{\gamma}_{jk}$ has an improper prior density function proportional to $\exp(-\frac{1}{2}\lambda_{jk}\boldsymbol{\gamma}_{jk}^\top\mathbf{G}_{jk}\boldsymbol{\gamma}_{jk})$.

Different formulations of the $\mathbf{Z}$'s and the $\mathbf{G}$'s result in different types of additive terms, for example, random effects terms, smoothing terms, time series terms or spatial terms as presented in Section 6.3.3. The advantage of modelling spatial data within GAMLSS is that different distributions beside the exponential family can be fitted and also it is possible, if needed, to model spatially any or all the parameters of the distribution.

6.3.2 Estimation of the model

The log likelihood function for the GAMLSS model (6.2) under the assumption that observations of the response variable are independent is given by

$$\ell = \sum_{i=1}^{n} \log f_Y(y_i|\mu_i, \sigma_i, \nu_i, \tau_i),$$

where $f_Y(.)$ represents the probability (density) function of the response variable. The penalised log-likelihood function for model (6.2) is given by

$$\ell_p = \ell - \frac{1}{2} \sum_{k=1}^{4} \sum_{j=1}^{J_k} \lambda_{jk} \boldsymbol{\gamma}_{jk}^{\top} \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk}. \tag{6.3}$$

We will need estimates for the 'betas', $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)^{\top}$, the parameters of the linear part of the model, the 'gammas', $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{11}, \ldots, \boldsymbol{\gamma}_{J_1 1}, \boldsymbol{\gamma}_{12}, \ldots, \boldsymbol{\gamma}_{J_4 4})^{\top}$, the random effects parameters, and the 'lambdas' $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{J_1 1}, \lambda_{12}, \ldots, \lambda_{J_4 4})^{\top}$, the hyper-parameters of the model.

Within the GAMLSS framework the linear parameters $\boldsymbol{\beta}$ and the random effects parameters $\boldsymbol{\gamma}$ are estimated (for fixed values of the smoothing hyper-parameters $\boldsymbol{\lambda}$) by maximizing the penalized likelihood function ℓ_p given by (6.3). There are two basic algorithms to achieve this, the `RS` and the `CG` algorithms. Both use an iteratively reweighted (penalised) least squares algorithm. Appendix C of Rigby and Stasinopoulos (2005) shows that both algorithms lead, for given $\boldsymbol{\lambda}$ hyper-parameters, to the maximum penalised log likelihood estimates for the betas and the gammas, i.e. $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$. Appendix A.1 of Rigby and Stasinopoulos (2005) shows that these estimates are also posterior mode (or MAP) estimates. The hyper-parameters $\boldsymbol{\lambda}$ can be estimated locally, see (Rigby and Stasinopoulos, 2013), or globally, see (Rigby and Stasinopoulos, 2005). The local methods are in general a lot faster and easier to implement than the global ones. 'Local' means that the method of estimation of the hyper-parameters applies each time within the `RS` or `CG` GAMLSS algorithms and 'global' means the method is applied outside the `RS` or `CG` GAMLSS algorithms. In addition, for either 'local' or 'global' estimation, there are (at least) three different criteria for estimating the smoothing hyper-parameters:

1. minimising the generalised cross validation (GCV), Wood (2006),
2. minimising the generalized Akaike information criteria (GAIC), Akaike (1983),
3. maximum likelihood (ML).

The default method in the GAMLSS software implementation is 'local ML' in which the (smoothing) hyper-parameters (and therefore their corresponding effective degrees of freedom) are estimated automatically using a local maximum likelihood (ML) procedure see (Rigby and Stasinopoulos, 2013). (This 'local ML' procedure is a penalised quasi-likelihood (PQL) method, Breslow and Clayton (1993)).

6.3.3 Gaussian Markov Random Fields

A Markov random field (MRF) is a set of random variables having a Markov property based on conditional independence assumptions and described by an undirected graph, $\mathcal{G}$, where each vertex represents an areal unit and each edge connects two areal units and represents a neighbouring relationship, Rue and Held (2005). Areal data are sometimes called lattice data, and often the lattice is a 2-dimensional grid in the plane, either finite or infinite.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph (Edwards, 2000; Whittaker, 2009) that consists of vertices $\mathcal{V} = (1, 2, \ldots, q)$, and a set of edges $\mathcal{E}$, where a typical edge is (m, t), $m, t \in \mathcal{V}$. Undirected is in the sense that (m, t) and (t, m) refer to the same edge. Following Rue and Held (2005), a random vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ is called a Gaussian MRF (i.e. GMRF) with respect to the graph $\mathcal{G}$, with mean $\boldsymbol{\mu}$ and precision matrix $\lambda\mathbf{G}$, if and only if its density has the form

$$\pi(\boldsymbol{\gamma}) \propto \exp\left[-\frac{1}{2}\lambda(\boldsymbol{\gamma} - \boldsymbol{\mu})^\top \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\mu})\right]$$

and

$$G_{mt} \neq 0 \iff (m, t) \in \mathcal{E} \text{ for } m \neq t,$$

where G_{mt} is the element of matrix $\mathbf{G}$ for row m and column t.

Hence the nonzero pattern of $\mathbf{G}$ determines $\mathcal{G}$. We can read off from $\mathbf{G}$ whether γ_m and

γ_t are conditionally independent, because a well known theorem is this field (Theorem 3.2 of Rue and Held (2005)) says that γ_m and γ_t are conditionally independent, given γ_r for all r not equal to m or t, if and only if $G_{mt} = 0$. It also means in practice that the precision matrix is often sparse. When $\mathbf{G}$ is non-singular, another way to represent a GMRF, by its conditional mean and precision matrix, was given in Besag (1974), and known as the conditional autoregressive model (CAR).

When $\mathbf{G}$ is singular the GMRF model can be represented by the intrinsic autoregressive model (IAR), Besag and Kooperberg (1995). The IAR model has been used for spatially structured random effects in generalized linear models Banerjee et al. (2014). Rue and Held (2005) and Wood (2006) present the IAR model within a generalized additive model (GAM) framework. In this context we extend to the GAMLSS framework.

Assume that a response variable and explanatory variables are recorded at observations which belong spatially to one of a set of areas (or regions). Zero, one or more than one observation may be recorded in each region. To incorporate IAR models within the GAMLSS model (6.2), set $\mathbf{Z}$ to be an index matrix defining which observation belongs to which area, and let $\boldsymbol{\gamma}$ be the vector of q spatial random effects and assume $\boldsymbol{\gamma} \sim N_q(0, \lambda^{-1}\mathbf{G}^{-1})$, where $\mathbf{G}^{-1}$ is the (generalized) inverse of a $q \times q$ matrix, $\mathbf{G}$. In the following IAR model, based on Besag and Higdon (1999), the matrix $\mathbf{G}$ contains the information about the neighbours (adjacent regions), with elements given by $G_{mm} = n_m$ where n_m is the total number of adjacent regions to region m and $G_{mt} = -1$ if region m and t are adjacent, and zero otherwise, for $m = 1, \ldots, q$ and $t = 1, \ldots, q$. This model has the attractive property that conditional on λ and γ_t for all $t \neq m$, then $\gamma_m \sim N(\sum \gamma_t n_m^{-1}, (\lambda n_m)^{-1})$ where the summation is over all regions which are neighbours of region m.

The nonzero pattern of the matrix $\mathbf{G}$ determines the graph $\mathcal{G}$. A non-zero value in matrix $\mathbf{G}$ implies a connection between the two corresponding regions in the graph $\mathcal{G}$ (they are connected neighbours). The zero value in matrix $\mathbf{G}$ implies no connection between the two regions in the graph $\mathcal{G}$ and hence that the corresponding spatial random effects γ_m and γ_t for the two regions are conditionally independent (given the other spatial random effects γ_r for all r not equal to m or t).

The `R` implementation of the above IAR model as a predictor term for any parameter of the distribution of the response variable in a GAMLSS model is achieved by the `R` package `gamlss.spatial` which is described in the Appendix.

6.4 Application to the Munich rent data

Here we use the package `gamlss.spatial` to provide a detailed spatial analysis of a data set on rents for flats in the City of Munich.

6.4.1 The data

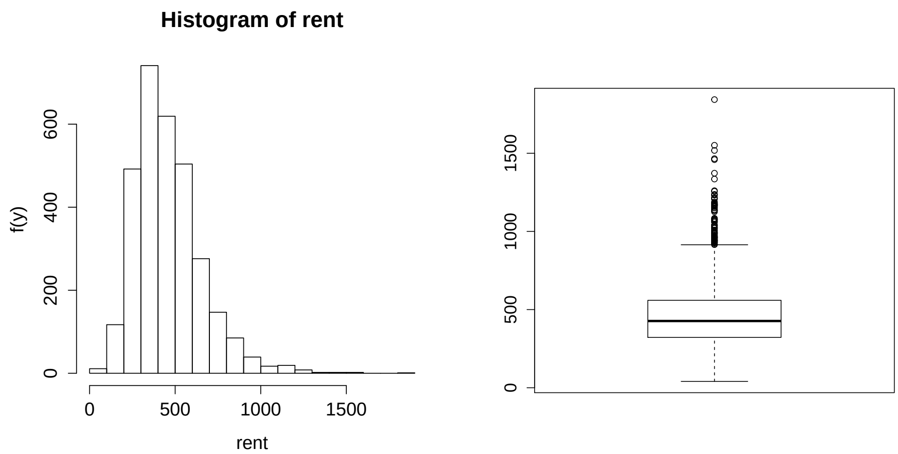
The response variable is the rent, (i.e. the monthly rental price, which remains after having subtracted all running costs and incidentals) of properties in the city of Munich, Kneib (2013). We used the Munich rent data in the year 1999, available from data frame `rent99` in the `gamlss.data` package in `R`. The data frame `rent99` has 3082 observations on the following 9 variables:

- `rent`: rent per month (in Euro),
- `rentsqm`: rent per month per square meter (in Euro),
- `area`: living area in square meters,
- `yearc`: year of construction,
- `location`: quality of location: a factor indicating whether the location is average location, (1), good location, (2), or top location, (3),
- `bath`: quality of bathroom: a factor indicating whether the bathroom facilities are standard, (0), or premium, (1),
- `kitchen`: quality of kitchen: a factor indicating whether the kitchen is standard, (0), or premium, (1),
- `cheating`: central heating: a factor indicating a property with central heating, (1), or without central heating, (0),
- `district`: district in Munich (this provides the spatial explanatory variable).

In the data frame `rent99` the variables `location`, `bath`, `kitchen` and `cheating` are declared as factors with reference levels 1, 0, 0 and 0 respectively. The reference level for `cheating` was changed to 1 in the analysis, because most properties have central heating.

The distribution of the monthly rent is asymmetric and skewed towards the right as is shown in Figure 6.1.

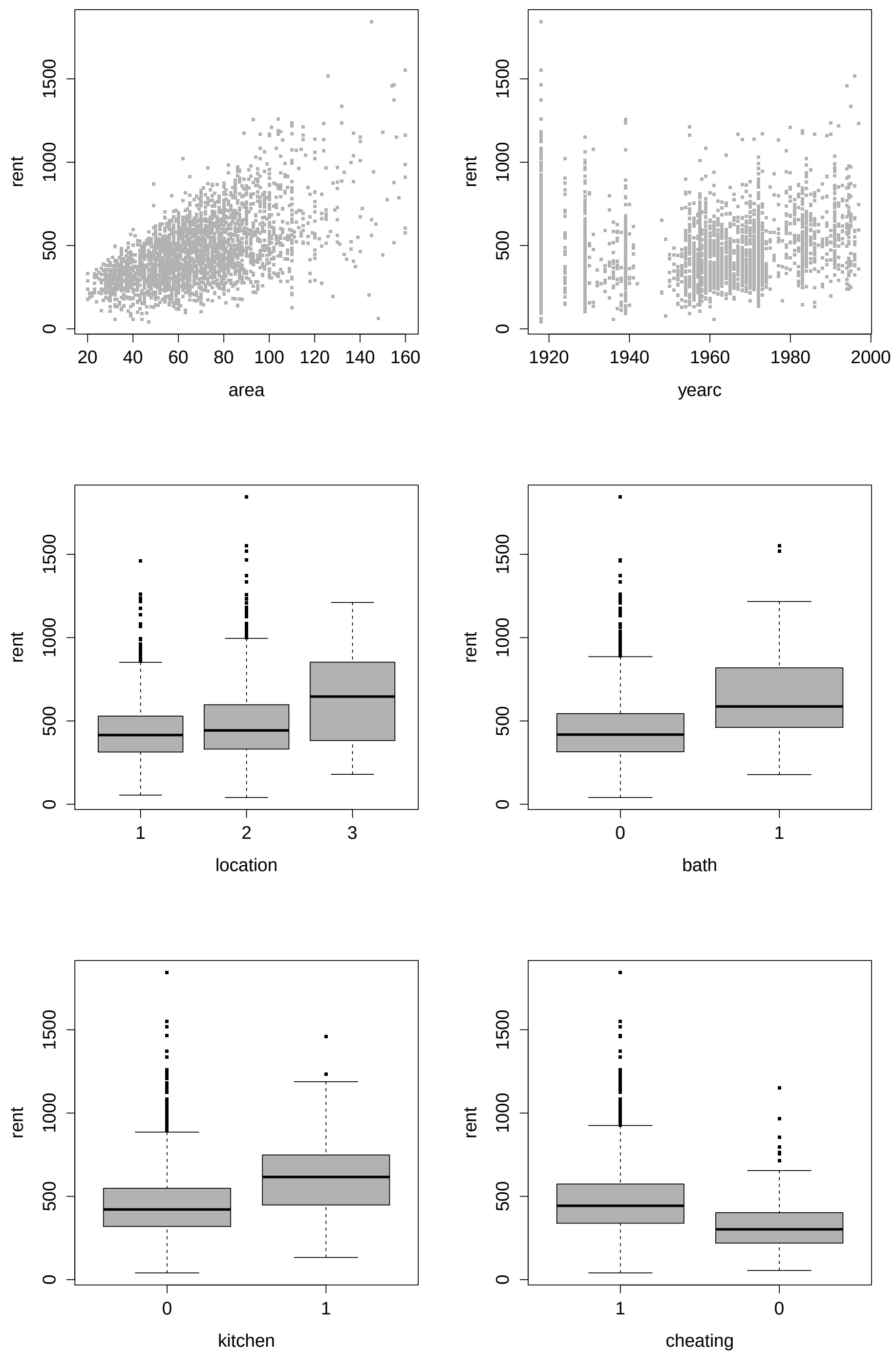
Figure 6.1: Histogram and box and whisker plots for rent data from the year 1999 in Munich.



Source: From the author.

Figure 6.2 shows plots of the `rent` against each of the above explanatory variables. Although these are bivariate exploratory plots and take no account of the interplay between the explanatory variables, they give an indication of the complexity of this data. The first two explanatory variables, `area` and `yearc`, are continuous. The plot of `rent` against `area` suggests a positive relationship between median rent and area, with an increased variation for larger `area`. The assumption of homogeneity in the variance of the `rent99` variable appears to be violated here. There is also some indication of positive skewness in the distribution of the `rent` variable. The peculiarity of the plot of `rent` against `yearc` is due to the method of data collection. The plot suggests that for houses up to 1960 the median rent price is roughly constant, but for flats constructed after that year there is an increasing trend in the median rent price. The remaining box and whisker plots display how the rent price varies according to the explanatory factors. The median rent price increases as the location changes from average to good and then to top location. The median rent price also increases if the flat has a premium bathroom, a premium kitchen or central heating. There are no surprises in the plots here, but again the problem of skewness is prominent with generally (but not always) longer upper than lower tails.

Figure 6.2: Plot of the `rent99` against explanatory variables `area`, `yearc`, `location`, `bath`, `kitchen` and `cheating`.



Source: From the author.

In summary, any statistical model used for the analysis of the above data should be able to deal with the complexity of the relationship between `rent` and the explanatory variables. The dependence of the median of the response variable `rent` on floor space (`area`) and year of construction (`yearc`) is non-linear and non-parametric smoothing functions may be needed. Median rent may also depend on interactions between the explanatory variables. There is clear indication of nonhomogeneity of the variance of rent. The variance of the response variable `rent` may depend on its mean and/or explanatory variables. There is clear indication of skewness in the distribution which may also depend on explanatory variables. The median rent (and the variance and skewness of rent) may also depend on the spatial explanatory variable (`district`), which is a key part of the analysis.

6.4.2 Model selection strategy

This section describes the model selection strategies adopted in this paper. Let $\mathcal{M} = \{\mathcal{D}, \mathcal{L}, \mathcal{T}, \boldsymbol{\lambda}\}$ represent a GAMLSS model as defined in Section 6.3.1. The components of $\mathcal{M}$ are defined as follows:

$\mathcal{D}$ specifies the distribution of the response variable,

$\mathcal{L}$ specifies the set of link functions for the distribution parameters μ, σ, ν and τ,

$\mathcal{T}$ specifies the terms appearing in the predictors for μ, σ, ν and τ,

$\boldsymbol{\lambda}$ specifies the smoothing hyperparameters which determine the amount of smoothing of continuous explanatory variables (`area` and `yearc`) and of the spatial effect (`district`).

In the search for an appropriate GAMLSS model for any new data set, all the above four components have to be specified as objectively as possible. The GAMLSS framework requires that the empirical researchers have a good understanding of the properties of the distributions from the list of available distributions in the GAMLSS framework.

The selection of the appropriate distribution $\mathcal{D}$ is done in two stages, the fitting stage and the diagnostic stage. The fitting stage involves the comparison of different fitted models using a generalised Akaike information criterion (GAIC). The diagnostic stage involves the normalized quantile residuals, Dunn and Smyth (1996), or 'z-scores', which provide information about the adequacy of the model and can be used in connection with diagnostic plots like worm plots, van Buuren and Fredriks (2001), or other test statistics eg. Z-statistcs and Q-statistcs, Royston and Wright (2000). The selection of the link function $\mathcal{L}$ is usually determined by the range of parameters. For a given distribution for the response variable, the selection of the terms $\mathcal{T}$ for the parameters of the distributions is done using a stepwise GAIC procedure.

Preliminary analysis, using distributions defined on the positive real line, indicated that the Box-Cox Cole and Green distribution (Cole and Green, 1992), $BCCGo(\mu, \sigma, \nu)$, seems an appropriate distribution for the rent data to use for model selection. The *BCCGo* distribution has a default log link for the median μ. If we use an identity link for μ it implies an additive model for μ and so, for example, changing from an unpopular to a popular district results in a fixed change in median rent, irrespective of how large an area the property has and irrespective of its year. It is more likely that the change in median rent is not a fixed amount but a fixed percentage, implying that a multiplicative model is more appropriate, i.e, a log link for μ. It was found that the log link for μ provided a better fit to the data then the identity link.

Because the fitting time of the spatial GMRF term for `district` in the model is longer that the rest of the terms, first we used a selection procedure for all explanatory variables (apart from `district`) for all distribution parameters (μ, σ and ν) using a generalised Akaike information criterion, GAIC, with penalty equal 4. Then, given the selected model, we tried adding the GMRF term IAR, with penalty equal 2. The reason for the choice of $k = 4$ in GAIC for the selection of terms (excluding the spatial effect) is that several terms have a single parameter and a 5% significance level for a generalized likelihood ratio test for a single parameter being different from zero is based on an (asymptotic) Chi-squared distribution with critical value $\chi^2_{1,0.05} = 3.84 \approx 4$. The spatial term involves many effective parameters being jointly tested and so a lower critical value per effective parameter is appropriate. When choosing whether to select a spatial term, we decided to use the standard AIC with $k = 2$.

The procedure to select the explanatory variables using the *BCCGo* distribution is first to fit an initial starting model and then:

(1) use a forward GAIC selection procedure to select an appropriate model for μ, with σ and ν as constants,

(2) use a forward selection procedure to select an appropriate model for σ, given the model for μ obtained in (1) and for ν fitted as a constant,

(3) use a forward selection procedure to select an appropriate model for ν, given the models for μ and σ obtained in (1) and (2), respectively,

(4) use a backward elimination procedure to select an appropriate model for σ, given the models for μ and ν obtained in (1) and (3), respectively;

(5) use a backward elimination procedure to select an appropriate model for μ, given the models for σ and ν obtained in (3) and (4), respectively.

The above procedure is executed in `gamlss` using the function `stepGAICAll.A`. The resulting chosen model may contain different explanatory variables for μ, σ and ν. Then from this model we

(i) add the `district` as a spatial effect for μ using the IAR spatial model,

(ii) add the `district` as a spatial effect for μ and σ using the IAR spatial model,

(iii) add the `district` as a spatial effect for μ, σ and ν using the IAR spatial model.

The (smoothing) hyper-parameters $\boldsymbol{\lambda}$ can be fixed or estimated from the data. The standard way of fixing the (smoothing) hyper-parameters is by fixing their effective degrees of freedom (edf) for smoothing.

The local maximum likelihood estimation method for each λ is the method used in our analysis. Hence, the model terms were selected using the GAIC, while the smoothing parameters (and hence their corresponding edf) were chosen using local maximum likelihood.

6.4.3 Results

In Section 6.4.2 we explained the model selection strategy. The final chosen fitted model `m2final` is given by

$$
\begin{aligned}
Y \sim& BCCGo(\hat{\mu}, \hat{\sigma}, \hat{\nu}),\\
\log(\hat{\mu}) =& 6.06 + h_{11}(yearc) + h_{21}(area) + s(district)\\
&+ 0.079(\text{if location=2, good}) + 0.211(\text{if location=3, top})\\
&- (0.255 - 0.0038 nyearc)(\text{if cheating=0, no central heating})\\
&+ (0.146 - 0.0034 nyearc + 0.0023 narea)(\text{if kitchen=1, premium})\\
&+ 0.067(\text{if bath=1, premium}),\\
\log(\hat{\sigma}) =& 11.811 + h_{12}(yearc) + 0.0016(area)\\
&+ 0.231(\text{if cheating=0, no central heating}),\\
\hat{\nu} =& -12.377 + h_{13}(yearc) + h_{23}(area) + 2.381(\text{if kitchen=1, premium}),
\end{aligned}
\tag{6.4}
$$

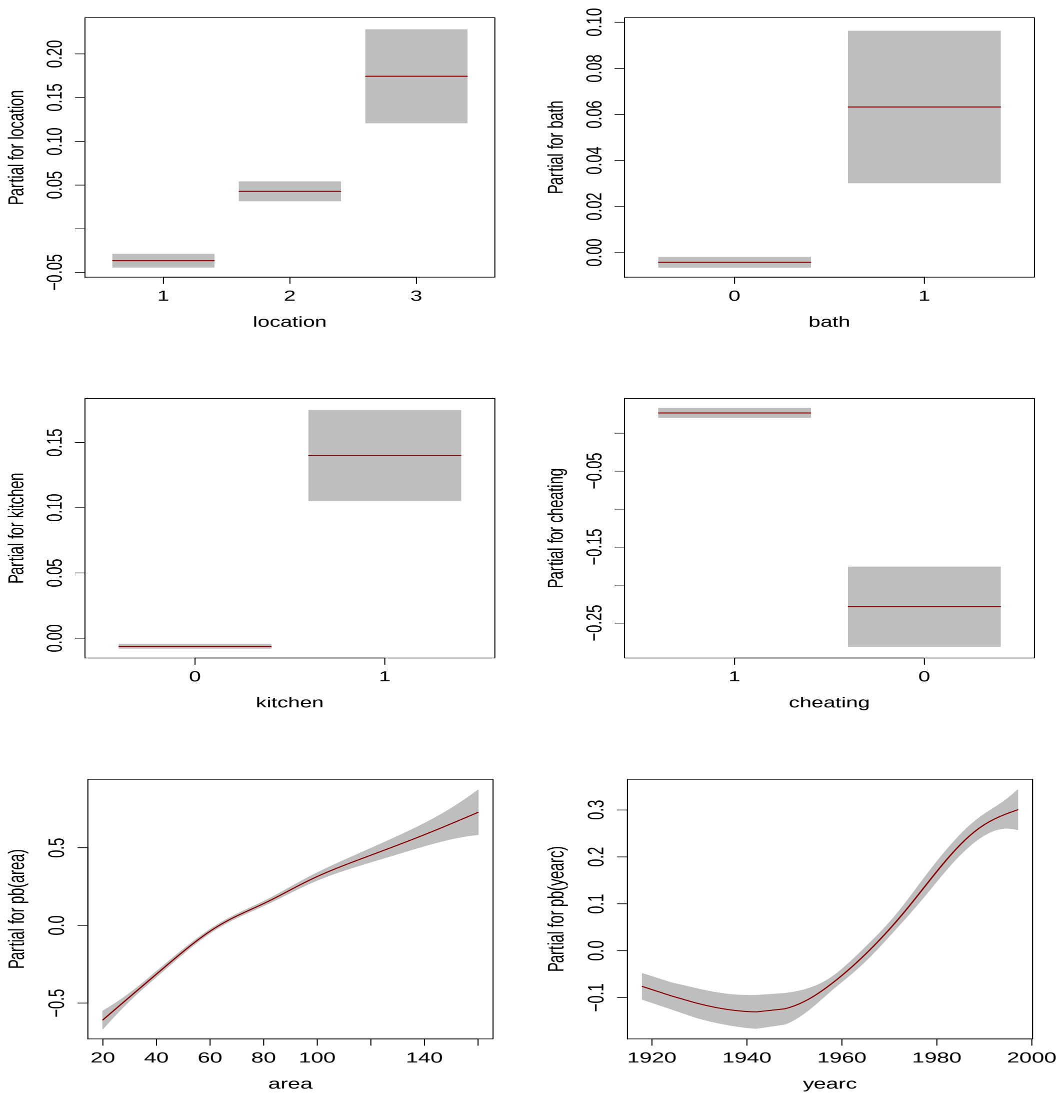
where the h functions are smooth non-parametric functions and s is an IAR spatial smoothing function. The distribution $BCCGo(\mu, \sigma, \nu)$ has a multiplicative model for the median μ, (resulting from the log link for μ), and *nyearc* and *narea* are respectively *yearc* and *area* centred at their means (i.e. subtract their means, 67.37 and 1956.31, respectively). The median μ model includes a spatial term in `district` (using the GMRF model IAR), and provides an improvement (i.e. reduction) in AIC. We also fitted the model with additional spatial effects for σ and ν but the improvement was too small so we opted for the simpler model, in this case, the spatial effect just for μ.

Figures 6.3, 6.4 and 6.5 display the fitted parametric terms and smooth functions in $\log(\hat{\mu})$ in the final chosen model (6.4). Their effects are additive for $\log(\hat{\mu})$ and hence multiplicative for the fitted median rent $\hat{\mu}$. The fitted median rent generally increases with area and year of construction (from Figure 6.3). A good location results in a 8.2% [calculated by $(e^{0.079} - 1) \times 100$] increase in fitted median rent (relative to an average location), while a premium location results in a 23.5% increase, and a premium bathroom results in a 6.9% increase.

The effect on median rent of no central heating depends on the year of construction. No central heating results in a 22.5% decrease in median rent for the average year of construction, and a higher % decrease for older properties. The effect of a premium kitchen on median rent depends on both year of construction and area of the property, resulting in a 15.6% increase in median rent for a property with average year of construction and average area, and a higher % increase for older or larger properties.

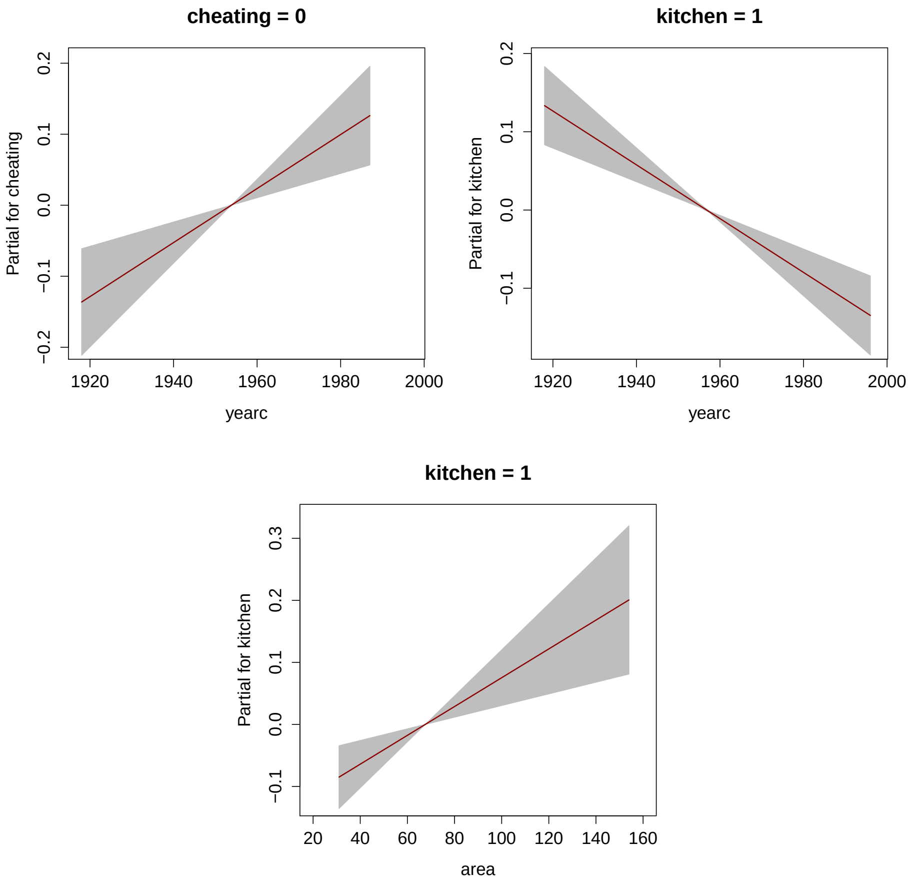
Figure 6.5 shows the `district` effect on $\log(\hat{\mu})$ where we can see that the rent prices are higher in the centre and southeast regions than in the north and west regions of the Munich city. Relative to the baseline district a region with the best district has a 10.5% [i.e $(e^{0.10} - 1) \times 100$] higher fitted median rent, while a region with worst district has a 9.5% [i.e $(1 - e^{-0.10}) \times 100\%$] lower fitted median rent (assuming all other explanatory variables including location type are fixed).

Figure 6.3: Term plots for μ.



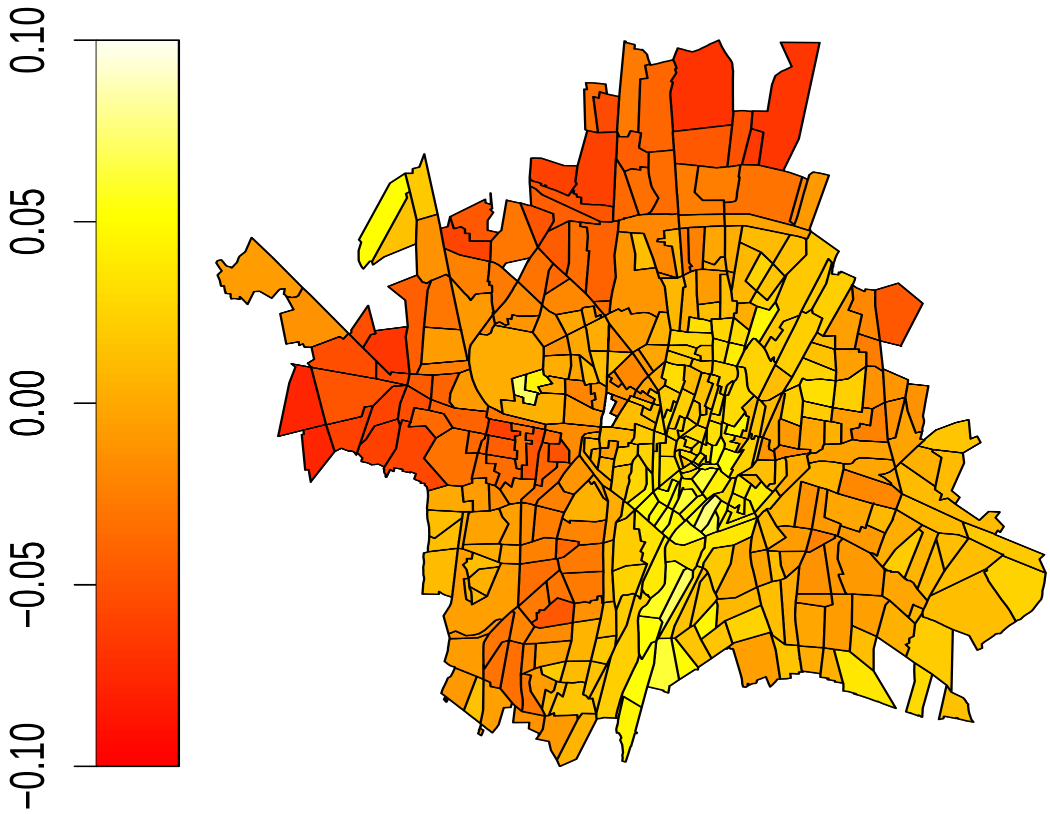
Source: From the author.

Figure 6.4: Term plots of the interactions for μ.



Source: From the author.

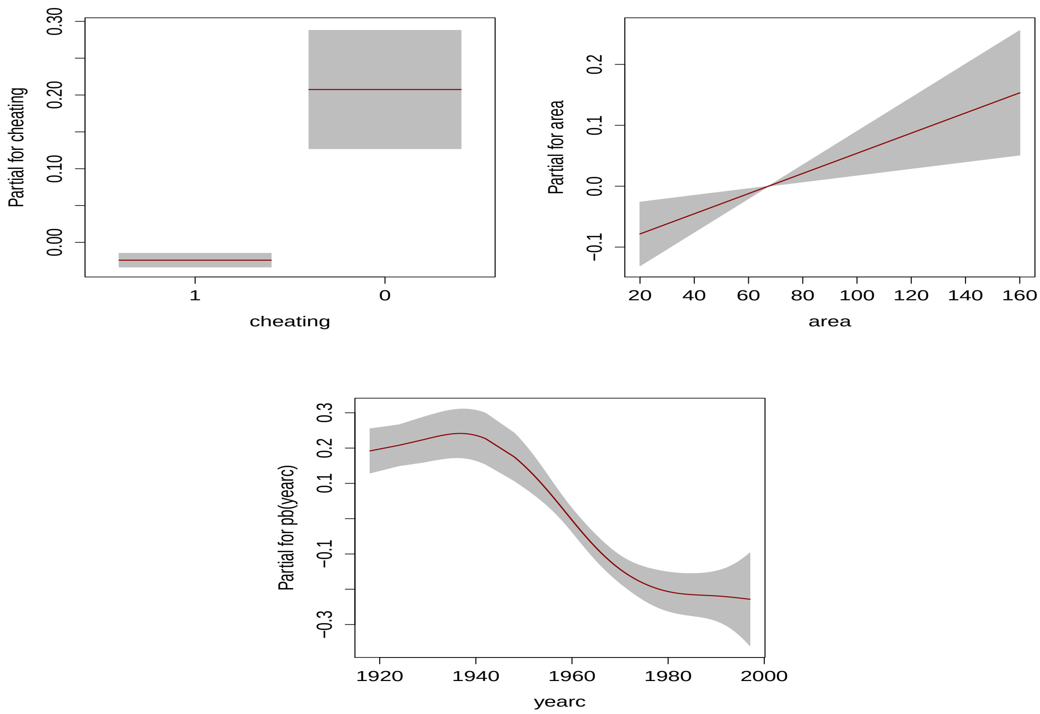
Figure 6.5: The fitted spatial effect for μ for the chosen model with spatial effect.



Source: From the author.

Figure 6.6 shows the fitted parametric terms and smooth function in $\log(\hat{\sigma})$, in the final chosen model (6.4). Figure 6.6 shows that the fitted $\hat{\sigma}$ (the approximate coefficient of variation of rent) increases with area but decreases with year of construction. No central heating results in a 26.1% increase in $\hat{\sigma}$. [It should be noted that if the total effective degrees of freedom used in the model for μ is high relative to the sample size, then this can result in negative bias in $\hat{\sigma}$. This was not the case in the fitted model (6.4)].

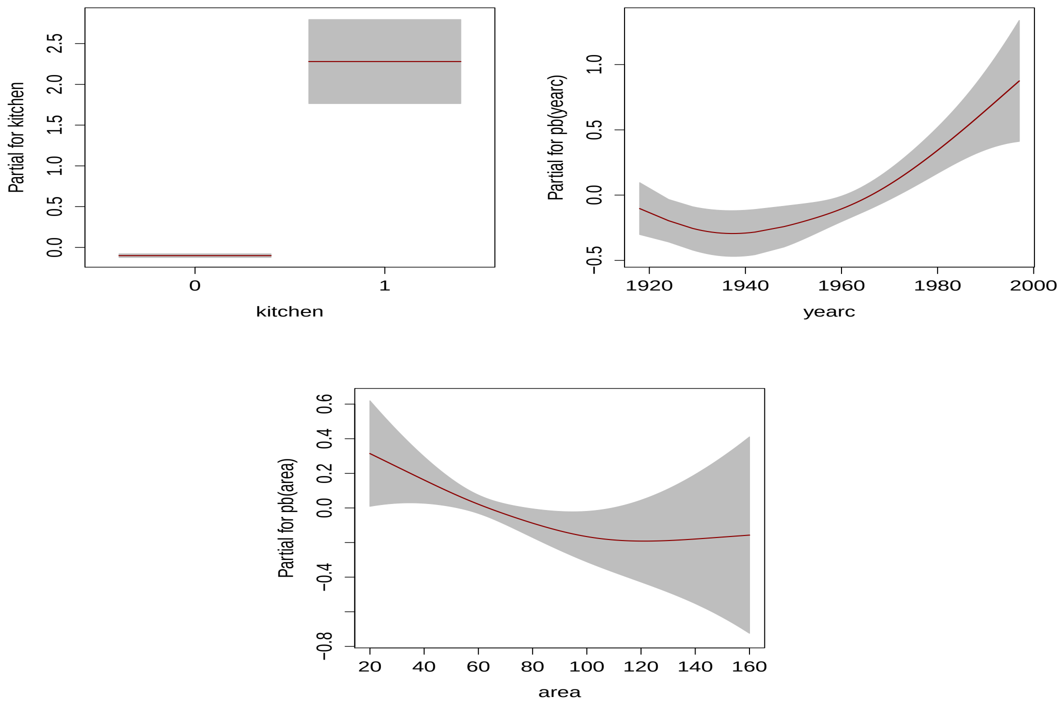
Figure 6.6: Term plots for σ.



Source: From the author.

Figure 6.7 shows that the fitted $\hat{\nu}$ (the skewness parameter) in (6.4) decreases with area but increases with year of construction. Note that decreasing $\hat{\nu}$ increases the positive skewness of the fitted distribution for rent. A premium kitchen results in an increase of 2.4 in $\hat{\nu}$. Hence larger older properties with a standard kitchen have a more positively skew fitted distribution for rent.

Figure 6.7: Term plots for ν.

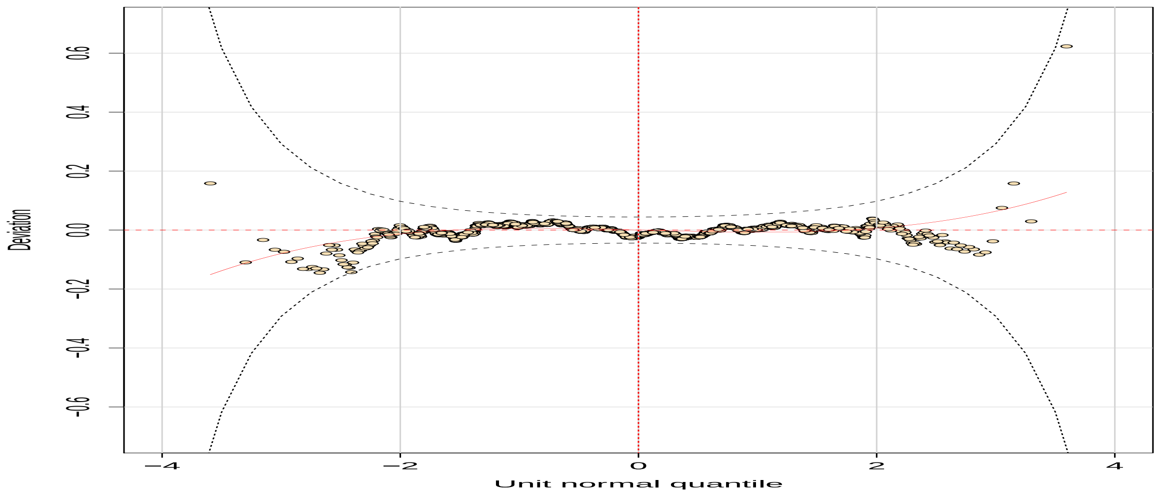


Source: From the author.

6.5 Residual diagnostics

We check the adequacy of the fitted model using (normalised quantile) residuals, Dunn and Smyth (1996). If the model is correct then the true residuals have a standard normal distribution. Figure 6.8 displays a worm plot for the residuals of the chosen fitted model. The worm plot is a detrended normal QQ plot of the residuals which indicates a reasonable fit to the data, since over 95% of the points lie within the elliptical (dashed) 95% pointwise interval bands.

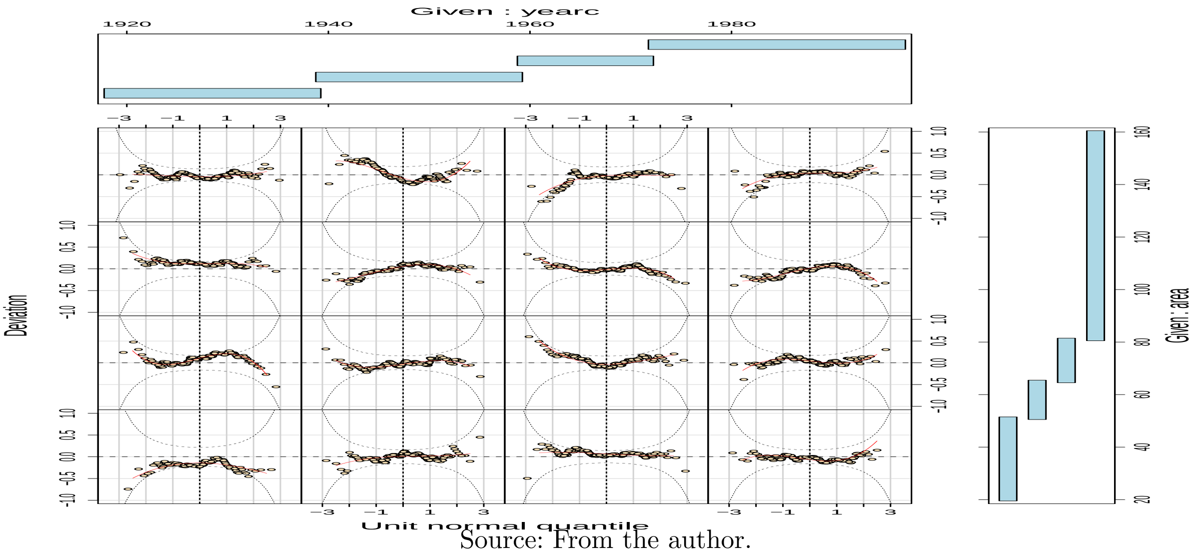
Figure 6.8: Worm plot of the residuals for the chosen final model `m2final`.



Source: From the author.

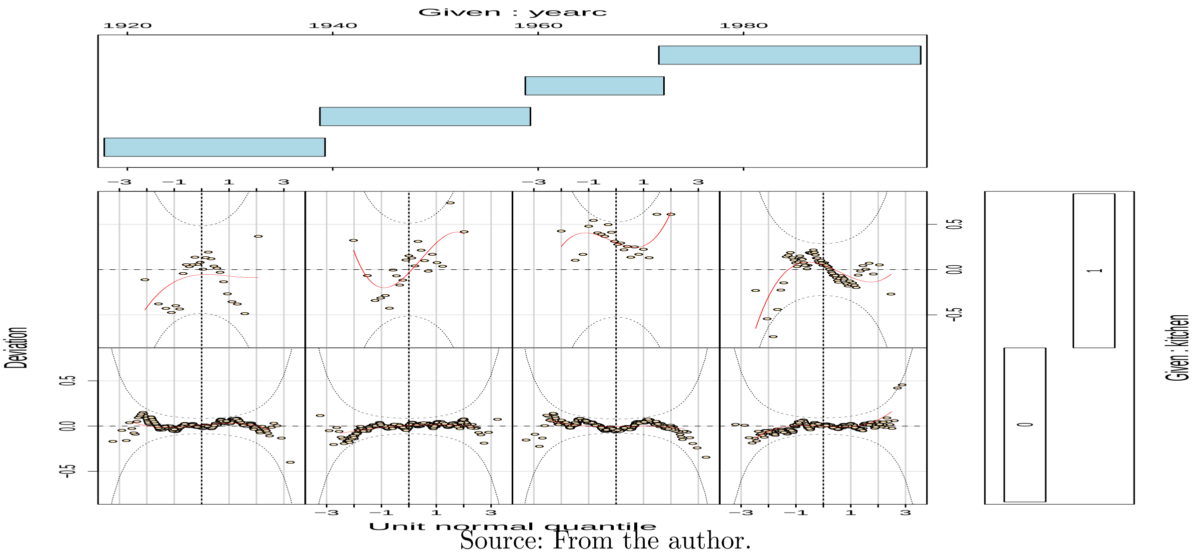
In order to investigate the adequacy of the chosen model for different combinations of the two continuous explanatory variables `yearc` and `area`, we cut each explanatory variable into four non-overlaping intervals with equal numbers of observations giving 16 joint intervals and obtain a worm plot (i.e detrented QQ plot) for cases in each of the 16 joint intervals. This is a way of highlighting failures of the model within different joint ranges of the two explanatory variables. Figure 6.9 shows the result, (obtained by a single `worm` command in the `gamlss` package), where above the plot the four intervals for `yearc` are displayed and to the right of the plot the four intervals of `area` are displayed. The worm plots generally indicate a reasonable fit to the data in the 16 joint intervals. Similarly, Figure 6.10 displays the worm plots for combinations of the two explanatory variables `yearc` and `kitchen`, which also indicate a reasonable fit to the data. For the sake of brevity, the worm plots for individual explanatory variables and for other combinations of two explanatory variables were omitted here, but they also indicated a reasonable fit to the data.

Figure 6.9: Worm plot of the residuals split by the `yearc` and `area` variables for the final model.



Source: From the author.

Figure 6.10: Worm plot of the residuals split by the `yearc` and `kitchen` variables for the final model.



Source: From the author.

6.6 Conclusions

We have shown that the GAMLSS framework provides a platform to fit, compare and check spatial models for the parameters of the distribution of a response variable which may be non exponential family. This includes continuous response variable distributions which are highly positively or negatively skewed and/or have high or low kurtosis (i.e. leptokurtic or platykurtic), discrete count distributions that are overdispersed (eg negative binomial) or have excess zeros (eg zero inflated negative binomial), or mixed continuous-discrete distributions (eg zero inflated gamma and inflated beta). The spatial analysis shown in this paper can be applied to other data sets that have geographical information specifying the neighbours of each region.

We would like to finish by emphasizing that looking at a single statistical model in isolation is not a good practice. Any chosen model should be able to stand up to scrutiny and that involves being able to compare it with alternative models and checking its assumptions.

Appendix D

D.1 The R implementation of GMRF spatial model within GAMLSS

Here we explain the implementation of the important GMRF submodel, the IAR model described in Section 6.3.3, within GAMLSS. The IAR model is implemented in the package `gamlss.spatial` thought the function `gmrf()`. A new package was needed because of several dependencies of the function `gmrf()` on existing but not standard `R` packages. Also `gmrf()` is the first function of a series of additive term spatial functions that are in progress. The function `gmrf()` fits an IAR term within the predictor of any distribution parameter in a GAMLSS model. There are two methods implemented for estimating the (smoothing) hyperparameter λ. The two different methods should produce identical results and can be seen as PQL methods, (Breslow and Clayton, 1993). The method is selected by the argument `method` of the function `gmrf()`. There are two possible values for the method:

i) `method = "Q"` which estimates the spatial IAR (smoothing) hyper-parameter λ by minimizing the Q-function, (Rigby and Stasinopoulos, 2013), which is a way to minimize the local marginal likelihood function,

ii) `method = "A"` which estimates the spatial IAR (smoothing) hyper-parameter λ using the "alternating" method to minimize the local marginal likelihood, see (Rigby and Stasinopoulos, 2013).

To perform the analysis, we need the matrix $\mathbf{G}$, which has the information about the relationships between the areas, showing if they are neighbouring areas or not. If two polygons areas have at least a single point in common, then they are treated as neighbours. The function `gmrf()` accepts three different ways to pass the geographical information:

i) `polys`, the polygon: a 2-column matrix of coordinates defining the boundary for each area,

ii) `neighbour`, a list of the polygon area neighbours of each area, or

iii) `precision`, the matrix **G**.

If the `polys` information is given, then the function `gmrf()` will automatically compute the matrix **G** to do the analysis. The same happens if the `neighbour` information is given. The fastest way to estimate the spatial IAR model in (6.2), as described in sections 6.3.2 and 6.3.3, is to give the `precision` (i.e. matrix **G**) since no extra calulations are needed.

Extra utility functions available within the package to obtain the matrix **G** before performing the analysis (to speed up the fitting) are:

`polys2nb()` which creates `neighbour` list from the polygon `polys` information and

`nb2prec()` which creates the matrix **G** from the `neighbour` information.

When fitting several models for model selection this saves time. To plot the fitted values of a fitted `gmrf` object the function `draw.polys()` is available. For more details about the `gamlss.spatial` package, see the help file in `http://cran.r-project.org/`.

Chapter 7

Discussion

7.1 Resumo

Neste capítulo são apresentadas as conclusões gerais da tese e temas para trabalhos futuros.

7.2 Conclusion

In this work we presented inference and diagnostics in spatial model with different frameworks and techniques. In the geoestatistics framework, Chapter 2 extended the Gaussian spatial linear model relaxing the assumption of normality of observations and local influence methodology, for one single realization of the process. Chapters 3 presented inference on Gaussian spatial linear models with repetitions, likelihood ratio test and the Bartllet corrected version, and we amended and inference approach to estimate the smooth parameter from the Matérn family class of models. Chapters 4 and 5 considered new approach on local influence diagnostics and global diagnostics on Gaussian spatial linear models with repetitions, respectively. In the Markov random fields framework, Chapter 6 presented the generalized additive models for location, scale and shape and showed the flexibility of these models. The exposition of spatial models in GAMLSS in the paper is new and we believe that it has the potential for a wide impact in applied statistics. We believe this is well illustrated by the real data set analysis in the paper.

7.3 Future research

Some ideas of ongoing and/or future research are for instance:

- To obtain the modified likelihood ratio statistic and the corrected version by using Bartlett corrected factor;

- Non-normal alternative distributions for the Gaussian spatial linear models with repetitions such as the t multivariate distribution, and the framework of spatio-temporal models for this type of data set;

- The study of identifiability in the ellyptical spatial linear models with repetitions considering heteroskedasticity, and

- the spatio-temporal approach in geostatistics and within the GAMLSS framework.

References

Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions*. Dover Publications.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute*, 50:277–290.

Anselin, L. and Florax, R. (1995). *New directions in spatial econometrics*. Springer-Verlag, New York.

Assumpção, R. A., B., Uribe-Opazo, M., and M., G. (2014). Analysis of local influence in geostatistics using student's t-distribution. *Journal of Applied Statistics*, 41(11):2323–2341.

Atkinson, A. and Riani, M. a. C. A. (2004). *Exploring multivariate data with the forward search*. Springer, New York.

Bailey, T. and Krzanowski, W. (2000). Extensions to spatial factor methods with an illustration in geochemistry. *Mathematical Geology*, 32(6):657–682.

Banerjee, M. (1998). Cook's distance in linear longitudinal models. *Communications in statistics: Theory and Methods*, 27:2973–2983.

Banerjee, S., Carlin, B., and Gelfand, A. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC.

Barndorff-Nielsen, O. E. (1983). On a formula to the distribution of the maximum likelihood estimator. *Biometrika*, 70:343–365.

Barndorff-Nielsen, O. E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society B*, 56:125–140.

Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after bartlett adjustment of the likelihood ratio statistic. *Biometrika*, 75:374–378.

Barnett, V. (2004). *Environmental statistics: methods and applications*. Wiley, Chichester.

Bartlett, M. (1937). Properties of sufficiency and statistical test. *Proceeding of the Royal Society A*, 160:268–282.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36:192–236.

Besag, J. and Higdon, P. (1999). Bayesian analysis of agriculture field experiments (with discussion). *Journal of the Royal Statistical Society, B*, 61:691–746.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43:1–59.

Beyerlein, A., Fahrmeir, L., Mansmann, U., and Toschke, A. (2008). Alternative regression models to assess increase in childhood bmi. *BMC medical research methodology*, 8(1):59.

Billor, N. and Loynes, R. (1993). Local influence: a new approach. *Communication in Statistics Theory and Methods*, 22:1595–1611.

Bolin, D. and Lindgren, F. (2013). A comparison between markov approximations and other methods for large spatial data sets. *Computational Statistics & Data Analysis*, 61:7–21.

Borghi, E., de Onis, M., Garza, C., Van den Broeck, J., Frongillo, E. A., Grummer-Strawn, L., Van Buuren, S., Pan, H., Molinari, L., Martorell, R., Onyango, A. W., and Martines, J. C. (2006). Construction of the world health organization child growth standards: selection of methods for attained growth curves. *Statistics in Medicine*, 25:247–265.

Borssoi, J. (2014). *Modelos mistos lineares elípticos com erros de medição*. PhD thesis, Universidade de São Paulo.

Borssoi, J., De Bastiani, F., Uribe-Opazo, M., and Galea, M. (2011). Local influence of explanatory variables in gaussian spatial linear models. *Chilean Journal of Statistics*, 2(2):29–38.

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

Cadigan, N. and Farell, P. (2002). Generalized local in uence with applications to fish stock cohort analysis. *Applied Statistics*, 51:469–483.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(368-385).

Cerioli, A. and Riani, R. (1999). The ordering of spatial data and the detection of multiple outliers. *Journal of Computational and Graphical Statistics*, 8:239–258.

Chang, C.-H., Huang, H.-C., and Ing, C.-K. (2014). Asymptotic theory of generalized information criterion for geostatistical regression model selection. *The Annals of Statistics*, 42(6):2441–2468.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity analysis in linear regression.* John Wiley and Sons, New York.

Chen, F. and Zhu, H. (2009). Perturbation selection and local influence analysis for nonlinearstructural equation model. *Psychometrika*, 74:493–516.

Chen, H.-S., Simpson, D. G., and Ying, Z. (2000). Infill asymptotic s for a stochastic process model with measurement error. *Statistica Sinica*, 10:141—156.

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

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

Christensen, R., Pearson, L., and Johnson, W. (1992b). Case-deletion diagnostics for mixed models. *Technometrics*, 34:38–45.

Christensen, W. and Amemiya, Y. (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association*, 97-(457):302–317.

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in Medicine.*, 11:1305–1319.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics (American Statistical Association)*, 19(1):15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Serie B*, 48:133–169.

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

Cordeiro, G. and Cribari-Neto, C. (2014). *An Introduction to Bartlett Correction and Bias Reduction.* Springer-Verlag Berlin Heidelberg, 1st edition.

Cordeiro, G. M. (2004). Corrected likelihood ratio tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation*, 74:609–620.

Cordeiro, G. M., Cysneiros, A. H. M. A., and Cysneiros, F. J. A. (2006). Bartlett adjustments for overdispersed generalized linear models. *Communications in Statistics. Theory and Methods*, 35(5):937–952.

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, 49:1–39.

Cressie, N. (1993). *Statistics for Spatial Data.* John Wiley, New York.

Cribari-Neto, F. and Cordeiro, G. (1996). On bartlett and bartlett-type corrections. *Econometric Reviews*, 15(4):339–367.

Cribari-Neto, F. and Zarkos, S. (2003). Computing using ox. *Econometric and StatisticalComputational Economics*, 21:277–295.

Currie, I. D., Durban, M., and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):259–280.

Cysneiros, A. H. M. A. and Ferrari, S. L. P. (2006). An improved likelihood ratio test for varying dispersion in exponential family nonlinear models. *Statistics & Probability Letters*, 76:255–265.

Cysneiros, F. J. A., Cordeiro, G. M., and Cysneiros, A. H. M. A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computacional and Simulation*, 80:451–461.

Davison, A. C. and Tsai, C. L. (1992). Regression model diagnostics. *International Statistical Review*, 60(337–353).

De Bastiani, F., Cysneiros, A. H. M. A., Galea, M., and Uribe-Opazo, M. A. (2016). Local influence in gaussian spatial linear models with repetitions.

De Bastiani, F., Cysneiros, A. H. M. A., Uribe-Opazo, M. A., and Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *TEST*, 24:322–340.

De Bastiani, F. and Stasinopoulos, M. (2015). *gamlss.spatial: Package to fit spatial data in gamlss. R package version 1.4.*

De Gruttola, V., Ware, J. H., and Louis, T. A. (1987). Influence analysis of generalized least squares estimators. *Theory and Method*, 82(399):911–917.

Demidenko, E. (2004). *Mixed Models: Theory and Applications.* John Wiley & Sons, New Jersey.

Demidenko, E. and Stukel, T. A. (2005). Influence analysis for linear mixed-effects models. *Statistics in Medicine*, 24:893–909.

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

Diggle, P. (2003). *Statistical analysis of spatial point patterns.* Edward Arnold. 2nd edition.

Diggle, P. J. and Ribeiro Jr., P. J. (2007). *Model-based Geostatistics.* Springer.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Doornik, J. (2006). *An Object-oriented Matrix Programming Language - Ox4.* London: Timberlake Consultants.

Dunn, P. K. and Smyth, G. K. (1996). Randomised quantile residuals. *Journal of Computational and Graphical Statistics*, 5:236–244.

Edwards, D. (2000). *Introduction to Graphical Modelling.* Springer, New York.

Eilers, P. (2003). A perfect smoother. *Analytical Chemistry*, 75(14):3631–3636.

Eilers, P. H. and Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties (with comments and rejoinder). *Statistical Science*, 11:89–121.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, methods and applications.* Springer, Berlin.

Fang, K. and Anderson, T. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions.* Allerton Press, New York.

Fang, K., Kotz, S., and Ng, K. (1990). *Symmetric multivariate and related distributions.* Chapman & Hall, London.

Fang, K. and Zhang, Y. (1990). *Generalized Multivariate Analysis.* Springer-Verlag and Science Press, Berli.

Fenske, N., Fahrmeir, L., Rzehak, P., and Höhle, M. (2008). Detection of risk factors for obesity in early childhood with quantile regression methods for longitudinal data.

Ferrari, S. and Uribe-Opazo, M. (2001). Corrected likelihood ratio tests in a class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics*, 15:49–67.

Ferrari, S. L. P., Cysneiros, A. H. M. A., and Cribari-Neto, F. (2004). An improved test for heteroskedasticity using adjusted modified profile likelihood inference. *Journal of Statistical Planning and Inferences*, 57:353–361.

Ferrari, S. L. P., Lucambio, F., and Cribari-Neto, F. (2005). Improve profile likelihood inference. *Journal of Statistical Planning and Inference*, 134(2):373–391.

Filzmoser, P., Ruiz-Gazen, A., and Thomas-Agnan, C. (2014). Identification of local multivariate outliers. *Statistical Papers*, 55:29–47.

Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling*. New York, Springer.

Galea, M., Paula, G. A., and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, 46:71–79.

Galea, M., Paula, G. A., and Cysneiros, F. J. A. (2005). On diagnostic in symmetrical nonlinear models. *Statistics & Probability Letters*, 73:459–467.

Galea, M., Paula, G. A., and Uribe-Opazo, M. A. (2003). On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers*, 44:23–45.

Garcia-Papani, F., A.Uribe-Opazo, M., Leiva, V., and Aykroyd, R. G. (2016). Birnbaum-saunders spatial modelling and diagnostics applied to agricultural engineering data. *Stochastic Environmental Research and Risk Assessment*, (online).

Genton, M. G. (1998). Highly robust variogram estimation. *Mathematical Geology*, 30(2).

Genton, M. G. and Zhang, H. (2012). Identifiability problems in some non-gaussian spatial random fields. *Chilean Journal of Statistics*, 2012(3):171–179.

Giménez, P. and Galea, M. (2013). Influence measures on corrected score estimators in functional heteroscedastic measurement error models. *Journal of Multivariate Analysis*, pages 1–15.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matern cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

Goodal, C. and Mardia, K. V. (1994). Challenges in multivariate spatio-temporal modeling. In *XVIIth International Biometric Conference*, pages 1–17, Ontario, Canada.

Gradshteyn, I. and Ryzhik, I. (2000). *Tables of Integrals, Series and Products.* Academic Press, New York.

Gupta, A. K. and Varga, T. (1993). *Elliptically contoured models in statistics.* Kluwer Academic Publishers, Massachusetts.

Handcock, M. and Wallis, J. (1994). An approach to statistical spatial- temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association 89*, 89:368–390.

Haskard, K. A., Cullis, B. R., and Verbyla, A. P. (2007). Anisotropic matern correlation and spatial prediction using reml. *Journal of Agricultural, Biological, and Environmental Statistics*, 12(2):147–160.

Haslett, J. and Dillane, D. (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation in the linear mixed model. *J. R. Statist. Soc. B*, 66:131–143.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Heller, G., Stasinopoulos, D., Rigby, R., and De Jong, P. (2007). Mean and dispersion modelling for policy claims costs. *Scandinavian Actuarial Journal*, 2007(4):281–292.

Hoaglin, D. and Welsh, R. (1978). The hat matrix in regression and anova. *The American Statistician*, 32:17–22.

Ibacache-Pulgar, G. and Paula, G. (2011). Local influence for student-t partially linear models. *Computational Statistics & Data Analysis*, 55:1462–1478.

Isaaks, E. and Srisvastava, R. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

Jameson, A. (1968). Solution of the equation ax + xb = c by inversion of an m x m or n x n matrix. *SIAM J Appl Math*, 16:1020–1023.

Jones, R. (1989). Fitting a stochastic partial differential equation to aquifer data. *Stochastic Hydrology and Hydraulics Journal*, 3:85–96.

Journel, A. G. and Huijbregts, J. C. (2004). *Mining Geostatistics*. The Blackburn Press.

Kano, Y., Berkane, M., and Bentler, P. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*, 88:135–143.

Kaufman, C. G. and Shaby, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. 100(2):473–484.

Kelker, D. (1970). Distribution theory of spherical and a location-scale parameter generalization. *Sankhya A*, 32:419–430.

Khondoker, M. R., Glasbey, C. A., and Worton, B. J. (2007). A comparison of parametric and nonparametric methods for normalising cdna microarray data. *Biometrical Journal*, 49(6):815–823.

Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, 13:275–303.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139.

Kunsch, H. R. (1979). Gaussian markov random fields. *Journal of the Faculty of Science. University of Tokyo, Section IA. Mathematics*, 26:53–73.

Lange, K. and Sinsheimer, J. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2:175–198.

Lange, K. L., Little, R. J., and Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84:881–896.

Lawley, D. N. (1956). A general method of approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43:295–303.

Lee, Y. and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, 88:987–1006.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed model. *Biometrics*, 54:570–582.

Little, R. J. and Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers*, 41:211–224.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146.

Mardia, K. V. and Watkins, A. J. (1989). On multimodality of the likelihood in the spatial linear model. *Biometrika*, 76(2):289–295.

Martin, R. (1992). Leverage, influence and residuals in regression models when observations are correlated. *Communications in Statistics - Theory and Methods*, 21:1183–1212.

Matérn, B. (1960). *Spatial variation.*, volume 49. Meddelanden fren Statens Skogsée-forskningsinstitut. 2nd Edition (1986), Lecture Notes in Statistics, No. 36, Springer, New York.

Matheron, G. (1962). Traité de géostatistique appliquée. Technical report.

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

Matheron, G. (1965). *Les Variables Régionaliseés et leur Estimation.* Masson, Paris.

Matheron, G. (1970). *Random functionts and their application on geology.* Merian, D. F.

Mayr, A., Fenske, N., Hofner, B., Kneib, T., and Schmid, M. (2012). Generalized additive models for location, scale and shape for highdimensional data - a flexible approach based on boosting,. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61:403–427.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models.* CRC Press.

McCullagh, P. and Tibishirani, R. (1990). A simple method for the adjustment of profile likelihood. *Jounal of the Royal Statistical Society B*, 52:325–344.

Melo, T. F. N., Ferrari, S. L. P., and Cribari-Neto, F. (2009). Improved testing inference in mixed linear models. *Computational Statistics & Data Analysis*, 53(7):2573–2582.

Mencía, J. and Sentana, E. (2005). Estimation and testing of dynamic models with generalized hyperbolic innovations. *CEPR Discussion Papers 5177.*

Militino, A., Palacius, M., and Ugarte, M. (2006). Outliers detection in multivariate spatial linear models. *Journal of Statistical Planning and Inference*, 136:125–146.

Minasny, B. and McBratney, A. B. (2005). The matérn function as a general model for soil variograms. *Geoderma*, 128(3-4):192–207.

Minozzo, M. and Fruttini, D. (2004). Loglinear spatial factor analysis: an application to diabetes mellitus complications. *Environmetrics*, 15(5):423–434.

Mitchell, A. (1989). The information matrix, skewness tensor and α-connections for the general multivariate elliptic distribution. *Annals of the Institute of Statistical Mathematics*, 41:289–304.

Nel, D. G. (1980). On matrix differentiation in statistics. *South African Statistical Journal*, 14:137–193.

Nychka, D., Furrer, R., and Sain, S. (2014). *fields: Tools for spatial data.* R package version 7.1.

Osorio, F., Paula, G., and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis*, 51(9):4354–4368.

Osorio, F. A. (2006). *Diagnóstico de influência em modelos elípticos com efeitos mistos.* PhD thesis, Universidade de São Paulo.

Pace, L. and Savan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective.* World Scientific, Singapore.

Pan, J., Fei, Y., and Foster, P. (2014). Case-deletion diagnostics for linear mixed models. *Technometrics*, 56(3):269–281.

Paula, G. A. (1993). Assessing local influence in restricted regression models. *Comput Stat Data Anal*, 16:63–79.

Pebesma, E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30:683–691.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in r. *R News*, 5(2).

Perrin, O. and Meiring, W. (1999). Identifiability for non-stationary spatial structure. *Journal of Applied Probability*, 36(4):1244–1250.

Poon, W. and Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society, B*, 61:51–61.

Preisser, J. S. and Qaqish, B. F. (1996). Deletin diagnostics for generalised estimating equations. *Biometrika*, 83:551–562.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

Reich, B. J. and Fuentes, M. (2007). A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264.

Ribeiro Jr, P. J. and Diggle, P. J. (2001). geor: a package for geostatistical analysis. *R-NEWS*, 1(2):15–18.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape, (with discussion). *Applied Statistics*, 54:507–554.

Rigby, R. A. and Stasinopoulos, D. M. (2013). Automatic smoothing parameter selection in gamlss with an application to centile estimation. *Statistical methods in medical research*, 23(4):318–332.

Rigby, R. A., Stasinopoulos, D. M., and Voudouris, V. (2013). Methods for the ordering and comparison of theoretical distributions for parametric models in the presence of heavy tails. *Statistical Modelling*, 13(4):335–348.

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley. 237pp.

Ross, W. (1987). The geometry of case deletion and the assessment of influence in non-linear regression. *The Canadian Journal of Statistics*, 15:91–103.

Royston, P. and Wright, E. M. (2000). Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine*, 19:2943–2962.

Rue, H. and Held, L. (2005). *Gaussian markov random fields: theory and applications*. Chapman & Hall, USA.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16:407–426.

Savalli, C., Paula, G. A., and Cysneiros, F. J. A. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, 6:59–76.

Schabenberger, O. and Gotway, C. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, London.

Schmid, V. and Held, L. (2001). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics*, 60:1034–1042.

Sherman, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. John Wiley & Sons.

Smith, R. L. (2001). Environmental statistics. notes at the conference board of the mathematical sciences (cbms) course at university of washington.

St. Laurent, R. and Cook, R. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, 87:985–990.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.

Stern, S. E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Jounal of the Royal Statistical Society B*, 59:653–665.

Tsai, C. L. (1986). Discussion of assessment of local influence by R. D. Cook. *J. Roy. Statist. Soc. Ser. B.*, 48:165.

Uribe-Opazo, M. A., Borssoi, J. A., and Galea, M. (2012). Influence diagnostics in gaussian spatial linear models. *Journal of Applied Statistics*, 39(3):615–630.

van Buuren, S. and Fredriks, M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, 20:1259–1277.

Vasconcellos, K. L. P. and Zea Fernandez, L. M. (2009). Influence analysis with homogeneous linear restrictions. *Computational Statistics & Data Analysis*, 53(11):3787–3794.

Verschuren, O., Bloemen, M., Kruitwagen, C., and Takken, T. (2010). Reference values for anaerobic performance and agility in ambulatory children and adolescents with cerebral palsy. *Developmental Medicine and Child Neurology*, 52(10):222–228.

Voudouris, V., Gilchrist, R., Rigby, R. A., Sedgwick, J., and Stasinopoulos, D. M. (2012). Modelling skewness and kurtosis with the bcpe density in gamlss. *Journal of Applied Statistics*, 39(6):1279–1293.

Voudouris, V., Stasinopoulos, D. M., Rigby, R. A., and Di Maio, C. (2011). The aceges laboratory for energy policy: Exploring the production of crude oil. *Energy Policy*, 39(9):5480–5489.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley-Interscience, Hoboken NJ.

Wang, N. and Heckman, N. (2009). Identifiability in linear mixed models. Technical report, Departament of Statistics, University of British Columbia.

Warnes, J. (1986). Sensitivity analysis for universal kriging. *Mathematical Geology*, 18:653–676.

Warnes, J. J. and Ripley, B. D. (1987). Problems with likelihood estimation of covariance functions of spatial gaussian processes. *Biometrika*, 74:640–642.

Waternaux, C., Laird, N. M., and Ware, J. H. (1989). Methods for analysis of longitudinal data: blood-lead concentrations and cognitive development. *Journal of the American Statistical Association*, 84(405):33–41.

Webster, R. and Oliver, M. (2007). *Geostatistics for Environmental Scientists*. John Wiley, New York, 2nd edition.

Wei, B., Hu, Y., and Fung, W. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25:25–37.

Wei, B.-C. (1998). *Exponential Family Nonlinear Models*. Springer.

Whittaker, J. (2009). *Graphical Models in Applied Multivariate Statistics*. Wiley, New York.

WHO, M. G. R. S. G. (2006). *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development*. Geneva: World Health Organization.

WHO, M. G. R. S. G. (2007). *WHO Child Growth Standards: Head circumference-for-age, arm circumference-for-age, triceps circumference-for-age and subscapular skinford-for-age: Methods and development*. Geneva: World Health Organization.

Wood, S. N. (2006). *Generalized Additive Models. An introduction with R*. Chapman and Hall.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a gaussian process. *Journal of Multivariate Analysis*, 36:280–296.

Yorinori, J. T., Paiva, W. M., Frederick, R. D., Costamilan, L. M., Bertagnolli, P. F., Hartman, G. E., V., G. C., and J., N. J. (2005). Epidemics of soybean rust (phakopsora pachyrhizi) in brazil and paraguay from 2001 to 2003. *Plant Disease*, 89:675–677.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, 71(354):400–405.

Zewotir, T. and Galpin, J. S. (2005). Influence diagnostics for linear mixed models. *Journal of Data Science*, 3:153–177.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58(1):129–136.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92(4):921–936.

Zhu, H. T., Ibrahim, J. G., Lee, S., and Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *Annals of Statistics*, 35:2565–2588.

Zhu, H. T. and Lee, S. Y. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society*, B63:111–126.

Zhu, J., Eickhoff, J. C., and Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics*, 61(3):674–683.

Zimmerman, D. (1989). Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models. *Journal of Statistical Computation and Simulation*, 32:1–15.

Zimmerman, D. and Zimmerman, M. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, 33(1):77–91.

Zucker, D. M., Lieberman, O., and Manor, O. (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society B*, 62:827–838.