

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM ESTATÍSTICA

**MODELOS ESPACIAIS COM DISTRIBUIÇÃO
NORMAL INDEPENDENTE**

DOUGLAS MATEUS DA SILVA

Orientadora: Lourdes Coral Contreras Montenegro

Co-orientadora: Thais Paiva Galletti

BELO HORIZONTE - MG
FEVEREIRO - 2017

Dedicatória

*À minha mãe Elisabete,
ao meu pai Geraldo,
à minha irmã Andressa e
à minha avó Maria (in memoriam).*

Agradecimentos

- Em primeiro lugar, agradeço à minha família pelo apoio e incentivo constantes.
- À minha orientadora, Professora Lourdes Coral Contreras Montenegro, pelo apoio e orientação em todo esse trabalho, bem como na minha jornada acadêmica.
- À minha co-orientadora, Professora Thais Paiva Galletti, pelo apoio e auxílio em todo esse trabalho.
- Ao Professor Marcos Oliveira Prates, pelas sugestões e auxílio em fases importantes deste trabalho.
- Aos professores do Departamento de Estatística pelos ensinamentos concedidos e disposição para ajudar.
- Aos meus amigos que sempre me incentivaram a continuar e seguir em frente, principalmente nos momentos difíceis.
- À CAPES pelo apoio financeiro.
- E agradeço a Deus pela oportunidade de realizar esse sonho.

Resumo

Modelos espaciais lineares, conhecidos como modelos geoestatísticos, são utilizados no estudo de fenômenos espacialmente correlacionados. Embora seja comum admitir que o processo espacial seja Gaussiano, a presença de observações atípicas torna imprópria essa análise, sendo necessário recorrer a distribuições simétricas de caudas mais pesadas. Neste trabalho, estudou-se o modelo espacial linear com distribuição normal/independente, que é uma alternativa ao modelo espacial Gaussiano. Nesta classe de distribuições estão presentes distribuições mais robustas, tais como a t-Student e slash. Foram desenvolvidos métodos de estimação por maximização da função de verossimilhança e algoritmo EM e realizadas análises, através de simulações, para avaliar o desempenho dos algoritmos. A construção de mapas temáticos foi realizada através da técnica de krigagem adaptada para o modelo proposto. Utilizando o método de influência local proposto por Cook (1986), apresentou-se um estudo de diagnóstico para o modelo espacial linear com distribuição normal/independente e ilustramos a metodologia proposta com um conjunto de dados de precipitação diária na Suíça referentes ao acidente de Chernobyl de 1986.

Palavras chaves: modelo geoestatístico, distribuição normal/independente, influência local.

Abstract

Linear spacial models, known as geostatistics models, are used in the study of phenomena that are spatially correlated. Although it is usual to admit that the spacial process follows a Gaussian distribution, the presence of atypical observations can make this analysis inappropriate. Thus, it might be necessary to consider symmetrical distributions with heavier tails. In this work, we studied the linear spatial model with normal/independent distribution, which is an alternative to the Gaussian spacial model. In this class, more robust distributions are included, such as t-Student and slash. We developed estimation methods by maximizing the likelihood function and EM algorithm. We also performed simulations under different scenarios to evaluate the performance of the algorithms. The construction of thematic maps was performed using the kriging technique adapted to the proposed model. Using the local influence method proposed by Cook (1986), we show a diagnostic study of the linear spacial model with normal/independent distribution. We conclude with an illustration of the proposed methodology with a data set of daily precipitation in Switzerland after the Chernobyl disaster occurred in 1986.

Keywords: geostatistical model, normal/independent distribution, local influence.

Sumário

Lista de Figuras	vii
Lista de Tabelas	xi
1 Introdução	1
1.1 Distribuição normal/independente	2
1.2 Processo geoestatístico	6
1.2.1 Condições de estacionariedade	7
1.2.2 Covariograma e variograma	9
1.3 Modelo geoestatístico	14
1.3.1 Predição	15
1.3.2 Seleção de modelos	16
1.4 Influência local	17
2 Modelo espacial linear com distribuição normal/independente	20
2.1 Descrição do modelo	20
2.2 Função de log-verossimilhança	20
2.3 Estimação dos parâmetros via máxima verossimilhança	21
2.4 Estimação dos parâmetros via algoritmo EM	24
2.5 Matriz de informação observada	26
2.6 Matriz de informação esperada	27
2.7 Influência local no modelo espacial linear com distribuição normal/independente	28
2.8 Predição	28
2.8.1 Distribuição normal	29
2.8.2 Distribuição t-Student	29
2.8.3 Distribuição slash	30
2.8.4 Comparação	30
2.9 Considerações computacionais	31
2.9.1 Método Numérico BFGS	31
2.9.2 Simulação de processos espaciais	33
2.9.3 Software e pacotes utilizados	34
3 Estudo de simulações	35
3.1 Desempenho dos algoritmos	36

3.2	Desempenho das estimativas realizadas por máxima verossimilhança	39
3.2.1	Variação no parâmetro da média	39
3.2.2	Variação nos parâmetros de escala	44
3.3	Predição	60
3.4	Influência local	61
3.4.1	Função de covariância exponencial	61
3.4.2	Função de covariância Gaussiana	64
4	Aplicação a dados experimentais	66
4.1	Dados de precipitação pluviométrica	66
4.1.1	Análise sem covariável	69
4.1.2	Análise com covariável	73
4.1.3	Análise sem a observação 14	77
4.1.4	Consideração geral	78
5	Considerações finais	80
	Referências Bibliográficas	81
	Apêndice A: Estimação dos parâmetros por máxima verossimilhança	84
	Apêndice B: Reparametrização da matriz escala Σ	86
B.1	Aplicada a função de log-verossimilhança	86
B.2	Aplicada a função Q	87
	Apêndice C: Valor esperado da distribuição condicional de U	89
C.1	Distribuição t-Student	89
C.2	Distribuição slash	89
	Apêndice D: Maximização da função Q	91
	Apêndice E: Derivadas da matriz Σ	92
	Apêndice F: Derivadas da matriz Δ	93

Lista de Figuras

1.1	Gráfico da função densidade da distribuição t-Student univariada com $\mu = 0$ e $\sigma = 1$ para alguns valores de ν	5
1.2	Gráfico da função densidade da distribuição slash univariada com $\mu = 0$ e $\sigma = 1$ para alguns valores de ν	6
1.3	Gráfico de covariograma e semivariograma isotrópicos em função da distância $ h $	11
1.4	Parâmetros do semivariograma.	12
1.5	Representação do gráfico de influência local.	19
3.1	Malha regular de tamanho 100 em uma região quadrada com identificação dos locais amostrados.	35
3.2	Gráfico de tempo de execução (em segundos) para estimação do vetor de parâmetros θ por MV e algoritmo EM de 500 simulações considerando tamanho de amostra 25 e função de covariância exponencial.	37
3.3	Gráfico de ARB (%) das estimativas de (a) $\hat{\mu}$, (b) $\hat{\phi}_1$, (c) $\hat{\phi}_2$ e (d) $\hat{\phi}_3$ realizadas por MV e algoritmo EM de 500 simulações considerando tamanho de amostra 25 e função de covariância exponencial.	38
3.4	Gráfico de $\hat{\mu}$ vs. μ considerando as distribuições em estudos e função de covariância exponencial.	40
3.5	Intervalo de 95% de confiança empírico para μ considerando as distribuições em estudos e função de covariância exponencial.	40
3.6	Gráfico de $\hat{\beta}_0$ vs. β_0 considerando as distribuições em estudos e função de covariância exponencial.	41
3.7	Gráficos de (a) $\hat{\beta}_1$ vs. β_1 e (b) $\hat{\beta}_2$ vs. β_2 considerando a função de covariância exponencial e distribuição normal, t-Student com 5 g.l. e slash com $\nu = 5$, respectivamente.	42
3.8	Intervalo de 95% de confiança empírico para (a) β_1 e (b) β_2 considerando as distribuições em estudos e função de covariância exponencial.	43
3.9	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de ϕ_1 considerando a função de covariância exponencial e distribuição normal.	44

3.10	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de ϕ_2 considerando a função de covariância exponencial e distribuição normal.	45
3.11	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de ϕ_3 considerando a função de covariância exponencial e distribuição normal.	45
3.12	Gráficos de (a) $\hat{\phi}_2$ vs. $\hat{\phi}_1$, (b) $\hat{\phi}_3$ vs. $\hat{\phi}_1$ e (c) $\hat{\phi}_3$ vs. $\hat{\phi}_2$ de 5000 simulações de um modelo ajustado com distribuição normal, função de covariância exponencial, tamanho de amostra 25 e vetor de parâmetros $\theta = (2, 1, 1, 2)$	46
3.13	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de ϕ_1 considerando a função de covariância Gaussiana e distribuição normal.	46
3.14	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de ϕ_2 considerando a função de covariância Gaussiana e distribuição normal.	47
3.15	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de ϕ_3 considerando a função de covariância Gaussiana e distribuição normal.	47
3.16	Intervalo de 95% de confiança empírico para ϕ considerando as funções de covariância em estudo, distribuição normal e tamanho de amostra 100.	48
3.17	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_1 considerando distribuição normal e as funções de covariância em estudo.	49
3.18	Gráfico de $\hat{\sigma}_h$ vs. σ_h ara alguns valores de ϕ_2 considerando distribuição normal e as funções de covariância em estudo.	50
3.19	Gráfico de $\hat{\sigma}_h$ vs. σ_h ara alguns valores de ϕ_3 considerando distribuição normal e as funções de covariância em estudo.	51
3.20	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância exponencial e distribuição t-Student com 5 g.l.	52
3.21	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância Gaussiana e distribuição t-Student com 5 g.l.	52
3.22	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_1 considerando distribuição t-Student com 5 g.l. e as funções de covariância em estudo.	53

3.23	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_2 considerando distribuição t-Student com 5 g.l. e as funções de covariância em estudo.	54
3.24	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_3 considerando distribuição t-Student com 5 g.l. e as funções de covariância em estudo.	55
3.25	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância exponencial e distribuição slash com $\nu = 5$	56
3.26	Análise do desempenho da estimação do vetor de parâmetros ϕ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância Gaussiana e distribuição slash com $\nu = 5$	56
3.27	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_1 considerando distribuição slash com $\nu = 5$ e as funções de covariância em estudo.	57
3.28	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_2 considerando distribuição slash com $\nu = 5$ e as funções de covariância em estudo.	58
3.29	Gráfico de $\hat{\sigma}_h$ vs. σ_h para alguns valores de ϕ_3 considerando distribuição slash com $\nu = 5$ e as funções de covariância em estudo.	59
3.30	Análise de predição para dados simulados considerando distribuição normal, função de covariância exponencial e modelo sem covariável.	60
3.31	Gráficos de (a) boxplot e (b) postplot do conjunto de dados simulado considerando a distribuição normal e função de covariância exponencial.	62
3.32	Gráficos de diagnóstico do conjunto de dados perturbado considerando a função de covariância exponencial e perturbação aditiva.	63
3.33	Gráficos de (a) boxplot e (b) postplot do conjunto de dados simulado considerando a distribuição normal e função de covariância Gaussiana.	65
3.34	Gráficos de diagnóstico do conjunto de dados perturbado considerando a função de covariância Gaussiana e perturbação aditiva.	65
4.1	Gráficos de (a) boxplot, (b) postplot, (c) coordenada X1 (leste-oeste) e (d) coordenada X2 (norte-sul) para os dados de precipitação.	67
4.2	Variograma direcional para as quatro direções principais dos dados de precipitação.	68
4.3	Gráfico de envelope simulado para o variograma omnidirecional dos dados de precipitação.	68
4.4	Gráficos de diagnóstico para os dados de precipitação considerando o modelo de média constante.	71

4.5	Mapas temáticos de predição para os dados de precipitação considerando modelos de média constante ajustados com (a) distribuição normal e função de covariância Gaussiana, (b) distribuição t-Student e função de covariância Gaussiana e (c) distribuição slash e função de covariância exponencial.	72
4.6	Gráficos de diagnóstico para os dados de precipitação considerando o modelo com covariável.	75
4.7	Mapas temáticos de predição para os dados de precipitação considerando modelos com covariável ajustados com distribuição (a) normal, (b) t-Student e (c) distribuição slash e função de covariância Gaussiana.	76
4.8	Mapas temáticos de predição considerando distribuição (a) normal, (b) t-Student com 1 g.l. e (c) slash com $\nu = 1$ para os modelos sem covariável (primeira coluna) e com covariável (segunda coluna) retirando a observação 14.	79

Lista de Tabelas

2.1	Valores de $-2V_{g(\delta)}$ para as distribuições em estudo.	22
3.1	Resultado da estimação do vetor de parâmetros θ e erros padrão (entre parênteses) por MV e algoritmo EM para o estudo de simulação considerando as distribuições normal, t-Student com 3, 5 e 10 g.l. e slash com $\nu = 5$	37
3.2	Estimativa do vetor de parâmetros ϕ considerando as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$ e parâmetros verdadeiros $\phi = (1, 1, 2)$	41
3.3	Estimativa do vetor de parâmetros ϕ considerando estimação com covariáveis para a estrutura da média, distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$ e parâmetros verdadeiros $\phi = (1, 1, 2)$	
3.4	Estimativas de $\theta = (\mu, \phi_1, \phi_2, \phi_3)$ e seus erros padrão dos modelos sem e com perturbação considerando a distribuição normal, t-Student com 5 g.l., slash com $\nu = 5$ e função de covariância exponencial.	62
3.5	Estimativas de $\theta = (\mu, \phi_1, \phi_2, \phi_3)$ dos modelos sem e com perturbação considerando a distribuição normal, t-Student com 5 g.l., slash com $\nu = 5$ e função de covariância Gaussiana.	64
4.1	Estatísticas descritivas para os dados de precipitação.	67
4.2	Estimativas e erros padrão entre parênteses dos parâmetros dos modelos ajustados sem covariável.	69
4.3	Estatísticas de validação de modelo para os modelos ajustados sem covariável.	70
4.4	Estimativas e erros padrão entre parênteses dos parâmetros dos modelos ajustados com covariável.	73
4.5	Estatísticas de validação de modelo para os modelos ajustados com covariável.	74
4.6	Estimativas e erros padrão entre parênteses dos parâmetros dos modelos ajustados sem a observação 14.	77
4.7	Estatísticas de validação de modelo para os modelos ajustados sem a observação 14.	78

Capítulo 1

Introdução

No estudo de diversos fenômenos naturais, como o teor de um elemento presente no solo ou a precipitação anual em determinada região, é necessário que se considere a associação espacial das variáveis de interesse. A hipótese de independência entre as variáveis se torna inadequada e as ferramentas produzidas a partir dessa suposição podem fornecer resultados impróprios. A partir disso, a geoestatística utiliza procedimentos para tratar dados espacialmente distribuídos e correlacionados, considerando a localização da variável, que pode ser expressa por um sistema de coordenadas, além de seu valor.

Segundo a primeira lei da geografia, "Todas as coisas estão relacionadas, mas coisas mais próximas são mais parecidas do que coisas mais distantes" (Tobler, 1970). Matheron (1963) desenvolveu a teoria das variáveis regionalizadas com o objetivo de considerar em sua análise o aspecto aleatório e estruturado do fenômeno em estudo. Na geoestatística, os dados são modelados através de uma função de média e de uma função de covariância, que determina a lei comportamental da associação das variáveis. A associação pode ser de tendência espacial, autocorrelação espacial ou ambos. Muitos estudos práticos tem sido desenvolvidos nessa área. Alguns exemplos podem ser vistos em Jácomo and Flores (2009), Shing et al. (2012) e Hongyu (2012).

Embora seja comumente assumido que o processo aleatório seja Gaussiano para esse tipo de dados, tais suposições podem ser questionáveis na presença de observações atípicas. Quando a suposição de normalidade é violada, uma alternativa é fazer uma transformação na variável de modo que a variável transformada siga uma distribuição normal. De Oliveira et al. (1997) utilizaram uma família de distribuições conhecida como Box-Cox de transformações. Uma aplicação pode ser vista em Oda-Souza et al. (2008), em que foi utilizada a transformação raiz quadrada na variável de interesse para adequar à suposição de normalidade. Diggle et al. (1998) propõem o uso de modelos espaciais lineares generalizados. Assumpção (2010) propõe a distribuição t-Student como uma extensão robusta do modelo normal, e De Bastiani (2012) propõe a classe de distribuição da família de contornos elípticos para o mesmo objetivo.

No presente estudo, consideramos modelos espaciais lineares com distribuição da família normal/independente, bastante discutida por Lange and Sinsheimer (1993). Desenvolvemos a estimação dos parâmetros desse modelo através da maximização da função de verossimilhança e algoritmo EM considerando as distribuições normal, t-Student e slash. O desempenho dos al-

goritmos foi avaliado através de um estudo de simulação sob diferentes cenários. Considerando o método de influência local proposto por Cook (1986), desenvolvemos a análise de diagnóstico para o modelo proposto para avaliar a sensibilidade das estimativas dos parâmetros do modelo considerando o esquema de perturbação aditiva na variável resposta.

Na Seção 1.1, apresentamos a classe de distribuição normal/independente e suas propriedades. Discutimos os conceitos de processo geoestatístico e de estacionariedade na Seção 1.2, e definimos o modelo geoestatístico na Seção 1.3. A Seção 1.4 descreve o método de influência local utilizado nesse estudo. No Capítulo 2, apresentamos o modelo geoestatístico com distribuição normal/independente, a estimação dos parâmetros do modelo, bem como a predição e análise de influência local para esse tipo de modelo. Um estudo de simulações é desenvolvido no Capítulo 3, e uma aplicação da metodologia em um conjunto de dados é realizada no Capítulo 4. O estudo é concluído com uma discussão no Capítulo 5.

1.1 Distribuição normal/independente

O principal atrativo da classe de distribuições normal/independente (NI) é permitir estender os modelos desenvolvidos sob o erro normal considerando distribuições simétricas com caudas mais pesadas do que a normal. Assim, Lange and Sinsheimer (1993) discutiram a classe de distribuições normal/independente (também chamada de mistura de escala normal) e suas propriedades. Algumas importantes distribuições nesta classe são a t-Student e slash e essas distribuições mais robustas serão o foco do nosso estudo. A seguir introduzimos as definições da classe de distribuição NI.

Definição 1.1 *Considere uma variável aleatória positiva U e um vetor aleatório \mathbf{Z} normal k -variado independente com média $\mathbf{0}$ e matriz de covariância não singular Σ . Se $\boldsymbol{\mu}$ é um vetor constante de dimensão k , então $\mathbf{Y} = \boldsymbol{\mu} + U^{-\frac{1}{2}}\mathbf{Z}$ é dito ser NI. Equivalentemente, \mathbf{Y} tem distribuição NI se a distribuição condicional de \mathbf{Y} dado $U = u$ é normal multivariada com vetor de média $\boldsymbol{\mu}$ e matriz de covariância $u^{-1}\Sigma$. A função densidade de \mathbf{Y} é*

$$f(\mathbf{y}) = \int_0^\infty \frac{u^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{u\delta}{2}} dH(u),$$

em que $\delta = (\mathbf{y} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})$ e $H(u)$ é a distribuição de mistura.

Propriedades:

- (i) Se $E(\mathbf{Y})$ existe, então $E(\mathbf{Y}) = \boldsymbol{\mu}$.

Prova. Assumindo que exista $E(\mathbf{Y})$, temos

$$\begin{aligned}
 E(\mathbf{Y}) &= E(\boldsymbol{\mu} + U^{-\frac{1}{2}}\mathbf{Z}) \\
 &= E(\boldsymbol{\mu}) + E(U^{-\frac{1}{2}})E(\mathbf{Z}) \\
 &= \boldsymbol{\mu} + E(U^{-\frac{1}{2}})0 \\
 &= \boldsymbol{\mu}
 \end{aligned}$$

■

(ii) Se $Cov(\mathbf{Y})$ existe, então $cov(\mathbf{Y}) = E(U^{-1})\boldsymbol{\Sigma}$.

Prova. Assumindo que exista $Cov(\mathbf{Y})$, temos

$$\begin{aligned}
 Cov(\mathbf{Y}) &= E[(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))'] \\
 &= E[(\boldsymbol{\mu} + U^{-\frac{1}{2}}\mathbf{Z} - \boldsymbol{\mu})(\boldsymbol{\mu} + U^{-\frac{1}{2}}\mathbf{Z} - \boldsymbol{\mu})'] \\
 &= E[(U^{-\frac{1}{2}}\mathbf{Z})(U^{-\frac{1}{2}}\mathbf{Z})'] \\
 &= E[U^{-1}\mathbf{Z}\mathbf{Z}'] \\
 &= E[U^{-1}]E[\mathbf{Z}\mathbf{Z}'] \\
 &= E[U^{-1}]\boldsymbol{\Sigma}
 \end{aligned}$$

■

(iii) Se \mathbf{A} é uma matriz de dimensão $k^* \times k$ e \mathbf{b} é um vetor constante de dimensão k^* , então a variável aleatória $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{b}$ pertence à classe de distribuição normal/independente e $\mathbf{X} \sim NI(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}'\boldsymbol{\Sigma}\mathbf{A}, \boldsymbol{\nu})$ (Valle, 1994).

(iv) Para encontrar a distribuição marginal de um subvetor de \mathbf{Y} consideramos a seguinte partição

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

em que \mathbf{Y}_1 e \mathbf{Y}_2 são vetores de dimensão $k_1 \times 1$ e $k_2 \times 1$, respectivamente. Então $\mathbf{Y}_1 \sim NI(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\nu})$ e $\mathbf{Y}_2 \sim NI(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \boldsymbol{\nu})$. Esse resultado é facilmente obtido aplicando a propriedade (iii) e escolhendo \mathbf{A} de forma adequada. Por exemplo, considerando $\mathbf{A} = [\mathbf{I}_{k_1} \mathbf{0}]$, em que $\mathbf{0}$ é uma matriz de zeros com dimensão $k_1 \times k_2$ e \mathbf{I}_{k_1} é a matriz identidade de tamanho k_1 .

Para indicar que uma variável tem distribuição normal/independente, denotamos que $\mathbf{Y} \sim$

$NI(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ em que $\boldsymbol{\nu}$ é o vetor de parâmetros da distribuição da variável de mistura. Exemplos desta classe de distribuições são dados por:

- (i) **Distribuição normal:** Se a distribuição de U é uma variável aleatória degenerada em 1 então a densidade de \mathbf{Y} é da forma

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{k}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\delta},$$

em que δ tem a forma como na Definição 1.1.

- (ii) **Distribuição t-Student:** Considere que U tem distribuição $Gama(\nu/2, \nu/2)$, então \mathbf{Y} terá distribuição t-Student com ν graus de liberdade (g.l.) e denotaremos $Y \sim t_\nu$. A densidade de U é dada por

$$f(u) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} e^{-\frac{\nu}{2}u} u^{\frac{\nu}{2}-1},$$

em que $\Gamma(\cdot)$ é a função Gama. A função de densidade multivariada da distribuição t-Student fica dada por

$$f(\mathbf{y}) = \frac{\Gamma((\nu + n)/2)}{\Gamma(\nu/2)(\nu\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{\delta}{\nu} \right]^{-(\nu+n)/2}.$$

Quando $\nu \rightarrow \infty$, a distribuição t-Student converge para a distribuição normal. A Figura 1.1 apresenta o gráfico da função densidade para a distribuição t-Student univariada com média nula, $\sigma = 1$ e alguns valores de ν . A linha contínua representa o gráfico da distribuição normal padrão. Observamos que rapidamente, com o aumento dos graus de liberdade, a distribuição t-Student se aproxima da distribuição normal padrão.

- (iii) **Distribuição slash:** Se considerarmos que U tem distribuição $Beta(\nu, 1)$, então \mathbf{Y} terá distribuição slash. A função de densidade multivariada da distribuição Slash é da forma

$$f(\mathbf{y}) = \frac{\nu}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} G\left(\frac{n}{2} + \nu, \frac{\delta}{2}\right)$$

em que $G\left(\frac{n}{2} + \nu, \frac{\delta}{2}\right) = \int_0^1 u^{(n/2)+\nu-1} e^{-u\delta/2} du$ representa a função gama incompleta.

Adicionando as constantes normalizadoras em $G\left(\frac{n}{2} + \nu, \frac{\delta}{2}\right)$, a função de densidade da distribuição slash toma a forma

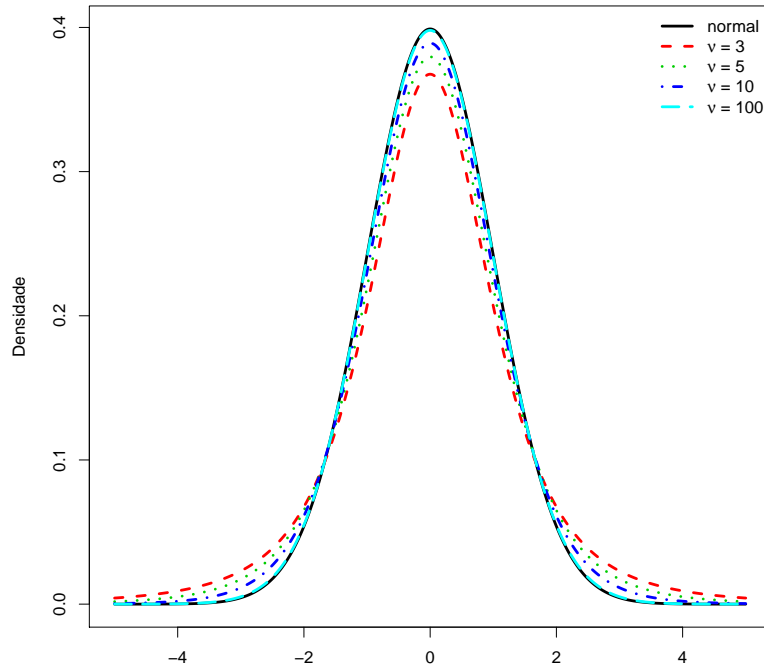


Figura 1.1: Gráfico da função densidade da distribuição t-Student univariada com $\mu = 0$ e $\sigma = 1$ para alguns valores de ν .

$$f(\mathbf{y}) = \frac{\nu}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \frac{\Gamma(n/2 + \nu)}{(\delta/2)^{n/2 + \nu}} F_S(1),$$

em que $F_S(1)$ é o valor da função de distribuição acumulada de uma distribuição Gama $(\frac{n}{2} + \nu, \frac{\delta}{2})$ no ponto 1.

Quando $\nu \rightarrow \infty$, a distribuição slash converge para a distribuição normal. A Figura 1.2 apresenta o gráfico da função densidade para a distribuição slash univariada com média nula, $\sigma = 1$ e alguns valores de ν . A linha contínua representa o gráfico da distribuição normal padrão. Observamos que rapidamente, com o aumento do valor de ν , a distribuição slash se aproxima da distribuição normal.

Existe uma relação entre a classe de distribuição normal/independente e as distribuições de contornos elípticos. Podemos reescrever a função de densidade do vetor aleatório \mathbf{Y} como

$$f(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g(\delta), \quad (1.1)$$

em que $\delta = (\mathbf{y} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ e $g(\cdot)$ é uma função contínua e diferenciável, $g : \mathbb{R} \rightarrow [0, \infty)$,

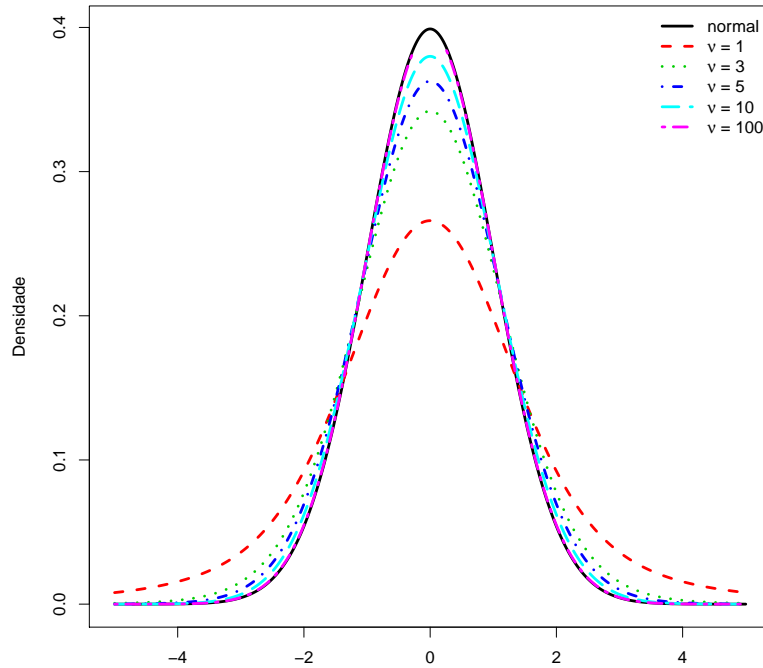


Figura 1.2: Gráfico da função densidade da distribuição slash univariada com $\mu = 0$ e $\sigma = 1$ para alguns valores de ν .

tal que

$$g(\delta) = \int_0^{\infty} \frac{u^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}} e^{-\frac{u\delta}{2}} dH(u).$$

É mais simples trabalhar com a função densidade de \mathbf{Y} em (1.1) nos cálculos que envolvem derivadas, como na estimação de parâmetros de modelo por máxima verossimilhança e na obtenção da matriz de informação observada.

1.2 Processo geoestatístico

Seja Y a variável resposta (no contexto unidimensional). $Y(\mathbf{s})$ representa a variável aleatória Y no ponto \mathbf{s} do espaço de domínio D , em que $D \subset \mathbb{R}^d$, com $d \geq 1$. Em cada ponto desse espaço, $Y(\mathbf{s})$ é uma variável aleatória com função específica de distribuição de probabilidade. Existem infinitos pontos no espaço D , uma vez que é contínuo, e conseqüentemente existem infinitas variáveis aleatórias $Y(\mathbf{s})$ com suas próprias funções de probabilidade. Esse conjunto de variáveis aleatórias é denominado como um processo estocástico.

Um processo geoestatístico $\{Y(\mathbf{s}); \mathbf{s} \in D\}$ é um processo estocástico em que \mathbf{s} representa um ponto que se desloca continuamente sobre uma região D . Essa região D é denominada domínio do processo geoestatístico. Os domínios mais usuais estão contidos em \mathbb{R} , caracterizado

normalmente por séries temporais, \mathbb{R}^2 , em regiões planas ou mapas, e \mathbb{R}^3 , quando há, além das coordenadas da localização, a profundidade, por exemplo. O ponto \mathbf{s} é denominado localização por ser usualmente referido ao local onde se observa a realização da variável aleatória $Y(\mathbf{s})$. Um processo $Y(\mathbf{s})$ é caracterizado através de sua função de distribuição de probabilidade (*f.d.p.*) conjunta, definida para um número finito de localizações,

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(y_1, \dots, y_n) = P[Y(\mathbf{s}_1) \leq y_1, \dots, Y(\mathbf{s}_n) \leq y_n],$$

em que $\mathbf{s}_1, \dots, \mathbf{s}_n$ são localizações do domínio D , y_1, \dots, y_n pertencem a \mathbb{R} e n é um número inteiro.

Para duas localizações \mathbf{s}_i e \mathbf{s}_j do domínio D , as variáveis aleatórias $Y(\mathbf{s}_i)$ e $Y(\mathbf{s}_j)$ não possuem, necessariamente, a mesma distribuição e normalmente não são independentes. É comum conhecermos apenas uma realização do processo $Y(\mathbf{s})$, o que torna impossível realizar inferência sobre o mesmo. Portanto, é necessário admitir hipóteses restritivas sobre o processo, ditas sobre sua estacionariedade. As principais são apresentadas a seguir.

1.2.1 Condições de estacionariedade

As condições de estacionariedade garantem a regularidade do processo estocástico em todo o domínio da região em estudo. Algumas admitem fortes suposições ao passo que outras tem suposições consideradas mais fracas. A seguir apresentamos a definição da condição de estacionariedade forte.

Definição 1.2 *Seja um processo geoestatístico $Y(\mathbf{s})$ com domínio $D \subset \mathbb{R}^d$ e com f.d.p. conjunta $F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(y_1, \dots, y_n)$, em que $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$, $y_1, \dots, y_n \in \mathbb{R}$ e $n \in \mathbb{N}$. O processo $Y(\mathbf{s})$ apresenta estacionariedade forte se, para qualquer conjunto finito de localizações $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, para qualquer conjunto finito de número reais $\{y_1, \dots, y_n\}$ e para qualquer vetor $\mathbf{h} \in \mathbb{R}^d$ tal que $\mathbf{s}_i + \mathbf{h} \in D$ ($i = 1, \dots, n$), então*

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(y_1, \dots, y_n) = F_{\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_n + \mathbf{h}}(y_1, \dots, y_n), \quad \forall n \in \mathbb{N}.$$

A Definição 1.2 diz que a lei de distribuição do processo $Y(\mathbf{s})$ é invariante por translação, isto é, os fatores do seu comportamento agem de forma similar em toda a área de estudo. Em outras palavras, as variáveis aleatórias do processo são identicamente distribuídas. Esta é uma hipótese forte para ser assumida em diversos fenômenos naturais. Por isso, uma hipótese menos restritiva deve ser assumida, a estacionariedade de segunda ordem.

Definição 1.3 Um processo $Y(\mathbf{s})$ com domínio $D \subset \mathbb{R}$ apresenta estacionariedade de segunda ordem se possui primeiro e segundo momentos constantes. Para dois pontos \mathbf{s}_i e \mathbf{s}_j , seja \mathbf{h} o vetor de distância de \mathbf{s}_i a \mathbf{s}_j . Então, temos que

$$(i) \ E[Y(\mathbf{s})] = \mu, \ \forall \mathbf{s} \in D;$$

$$(ii) \ Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = E[Y(\mathbf{s} + \mathbf{h}) \cdot Y(\mathbf{s})] - \mu^2 = C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{h});$$

$$(iii) \ Var[Y(\mathbf{s})] = E[(Y(\mathbf{s}) - \mu)^2] = C(0).$$

A primeira condição da Definição 1.3 garante que o processo possui média constante em todo domínio D . A segunda condição garante que a covariância entre duas observações depende apenas da distância entre elas. A função $C(\cdot)$ é definida como o covariograma do processo ou função de covariância estacionária.

Apesar de ser menos restritiva, a estacionariedade de segunda ordem supõe que a variável aleatória tenha variância finita, o que às vezes é falso. Quando a variável aleatória possui uma capacidade infinita de dispersão, precisamos de outra hipótese mais fraca, a hipótese intrínseca.

Definição 1.4 Um processo $Y(\mathbf{s})$ com domínio $D \subset \mathbb{R}$ apresenta estacionariedade intrínseca se

$$(i) \ E[Y(\mathbf{s}_i) - Y(\mathbf{s}_j)] = 0, \ \forall \mathbf{s}_i, \mathbf{s}_j \in D;$$

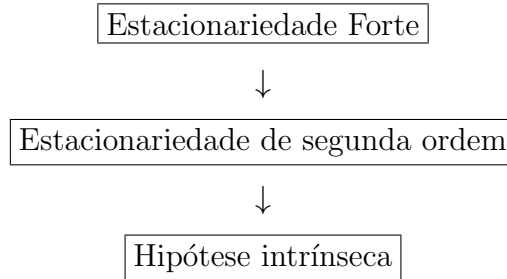
$$(ii) \ Var[Y(\mathbf{s}_i) - Y(\mathbf{s}_j)] = 2\gamma(\mathbf{s}_i - \mathbf{s}_j) = 2\gamma(\mathbf{h}), \ \forall \mathbf{s}_i, \mathbf{s}_j \in D.$$

Sob hipótese intrínseca, são os acréscimos da variável aleatória que obedecem à estacionariedade de segunda ordem e não a própria variável. As variáveis aleatórias $Y(\mathbf{s}_i) - Y(\mathbf{s}_j)$, $\forall \mathbf{s}_i, \mathbf{s}_j \in D$, são chamadas de incrementos do processo. A primeira condição da Definição 1.4 garante que os incrementos possuem média nula. Caso $Y(\mathbf{s})$ possua esperança finita, esta é apenas outra forma de garantir que o processo é estacionário na média, como visto na primeira condição da Definição 1.3. A segunda condição da Definição 1.4 garante que os incrementos tenham variância constante definida apenas pelo vetor de distância \mathbf{h} que separa as duas observações. A função $2\gamma(\mathbf{h})$ é definida como o variograma do processo, e a função $\gamma(\mathbf{h})$ como o semivariograma.

Um processo fortemente estacionário é, logicamente, estacionário de segunda ordem, uma vez que todos os momentos são invariantes por translação. O contrário pode não ser verdade. No caso de um processo Gaussiano (em que a *f.d.p* conjunta segue uma distribuição normal), a estacionariedade de segunda ordem implica em estacionariedade forte, já que os primeiro e segundo momentos definem completamente a distribuição do processo. Se o processo verifica

estacionariedade de segunda ordem, verifica a hipótese intrínseca, mas o inverso não é válido. Isso se deve ao fato de que os semivariogramas para a estacionariedade de segunda ordem são sempre limitados, ao contrário que há semivariogramas de processos que se enquadram na hipótese intrínseca e não são limitados.

Podemos esquematizar a relação entre as formas de estacionariedade, sendo que a seta significa "implica", como:



Observando que o inverso não é estritamente válido.

1.2.2 Covariograma e variograma

Seja $Y(\mathbf{s})$ um processo geoestatístico sob estacionariedade de segunda ordem. Como visto na Definição 1.3, o processo possui uma função de covariância estacionária, $C(\cdot)$, denominado covariograma. Para que a função $C(\cdot)$ seja um covariograma válido, deve satisfazer algumas propriedades, como

- (i) $C(0) = \text{Var}[Y(\mathbf{s})], \forall \mathbf{s} \in D$;
- (ii) $C(\mathbf{h}) = C(-\mathbf{h}), \forall \mathbf{h} \in \mathbb{R}^d$;
- (iii) o covariograma é positivo definido, isto é,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0, \forall n \in \mathbb{N}, \forall a_1, \dots, a_n \in \mathbb{R}, \forall \mathbf{s}_1, \dots, \mathbf{s}_j \in D.$$

A condição (iii) é necessária e suficiente para que se defina uma família de covariogramas válidos, porém não é uma condição fácil de ser verificada (Diggle and Ribeiro, 2007). Existem famílias paramétricas de covariogramas que satisfazem as propriedades citadas, como veremos posteriormente. Um covariograma é isotrópico quando depende apenas do vetor \mathbf{h} através de seu módulo, não dependendo de sua direção. Desse modo, para qualquer localização \mathbf{s} , a covariância entre $Y(\mathbf{s})$ e qualquer variável localizada sobre uma circunferência de raio $\|\mathbf{h}\|$ é constante. Caso contrário, temos anisotropia. Supondo a existência do covariograma, podemos

expressar a função de variograma como

$$\begin{aligned}
2\gamma(\mathbf{h}) &= \text{Var}[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})], \\
&= \text{Var}[Y(\mathbf{s} + \mathbf{h})] + \text{Var}[Y(\mathbf{s})] - 2\text{Cov}[Y(\mathbf{s} + \mathbf{h}), Y(\mathbf{s})], \\
&= 2C(0) - 2C(\mathbf{h}).
\end{aligned} \tag{1.2}$$

Note que o variograma pode ser definido pelo covariograma quando este existe. A função de variograma deve satisfazer algumas propriedades, como

- (i) $2\gamma(0) = 0$;
- (ii) $2\gamma(\mathbf{h}) = 2\gamma(-\mathbf{h})$, $\forall \mathbf{h} \in \mathbb{R}^d$;
- (iii) o variograma é negativo definido, isto é,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0, \quad \forall n \in \mathbb{N},$$

em que, a_1, \dots, a_n são tais que $\sum_{i=1}^n a_i = 0$ e $s_i, \dots, s_j \in D$.

- (iv) o variograma cresce mais lentamente que $\|\mathbf{h}\|^2$, isto é,

$$\lim_{\|\mathbf{h}\| \rightarrow \infty} \frac{2\gamma(\mathbf{h})}{\|\mathbf{h}\|^2} = 0.$$

Se um variograma satisfazer as propriedades (i) a (iv), é dito um variograma válido. O semivariograma é definido como metade do variograma, $\gamma(\|\mathbf{h}\|) = C(0) - C(\|\mathbf{h}\|)$, e possui as mesmas propriedades que o variograma. Desse modo, é comum utilizar o semivariograma em vez do variograma (Andriotti, 2003). Assim como o covariograma, um variograma é isotrópico quando depende apenas do vetor \mathbf{h} através de seu módulo, não dependendo de sua direção. Desse modo, o valor do variograma é o mesmo para quaisquer pares de variáveis aleatórias equidistantes. Uma forma de verificar isotropia é estudar o comportamento do variograma nas direções 0° , 45° , 90° e 135° , em que 0° é a direção Norte-Sul com sentido para o Norte.

Note que o variograma é definido para qualquer processo que obedeça à hipótese intrínseca, já o covariograma somente existe para processos sob estacionariedade de segunda ordem. Assim, se o processo possui covariograma, existe também seu variograma como visto em (1.2). Porém, a existência do variograma não implica a existência do covariograma. Esse é um dos motivos

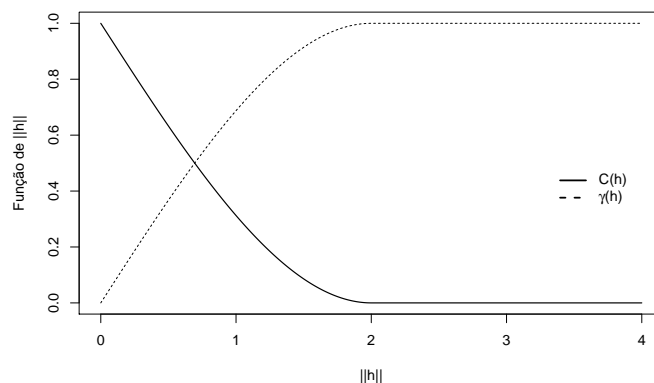


Figura 1.3: Gráfico de covariograma e semivariograma isotrópicos em função da distância $\|\mathbf{h}\|$.

pelo qual a função de variograma é preferida. É usual utilizar a função de semivariância, uma vez que possui as mesmas propriedades que o variograma.

A Figura 1.3 apresenta um exemplo de covariograma e semivariograma isotrópicos em função da distância $\|\mathbf{h}\|$. Em fenômenos naturais, espera-se que observações mais próximas sejam mais parecidas, ou seja, apresentem correlação maior do que observações distantes. Pelo gráfico do covariograma, é nítido que a correlação tende a zero quando a distância $\|\mathbf{h}\|$ aumenta. Ao contrário do gráfico de semivariograma, que começa da origem e atinge um patamar. Nesse caso, dizemos que o semivariograma é limitado. Caso não seja, não há covariograma relacionado. Apesar da propriedade (i) do semivariograma dizer que seu gráfico começa na origem, na prática pode haver descontinuidade nesse ponto. Esse efeito é chamado de efeito pepita (*nugget effect*). Se pudéssemos coletar duas amostras no mesmo local ($\|\mathbf{h}\| = 0$), os valores observados deveriam ser exatamente os mesmos. Porém, haverá uma variação entre essas amostras, seja por erro de medida, de amostragem ou outros. Assim, o gráfico do semivariograma não converge para zero quando $\|\mathbf{h}\|$ se aproxima da origem. Quando não há correlação espacial, a variação da variável aleatória se dá apenas pelo efeito pepita e é chamado de efeito pepita puro.

A Figura 1.4 apresenta um gráfico de semivariograma limitado. No gráfico de semivariograma $\gamma(\mathbf{h})$, podemos observar os seguintes parâmetros:

- Efeito pepita ou *nugget effect* (ϕ_1): representa a variabilidade não espacial.
- Contribuição ou *partial sill* (ϕ_2): representa a variabilidade espacial.
- Patamar: é o valor do semivariograma correspondente ao seu alcance a partir do qual o semivariograma se estabiliza, dado por $\phi_1 + \phi_2$. Quando o semivariograma não é limitado, não existe patamar.

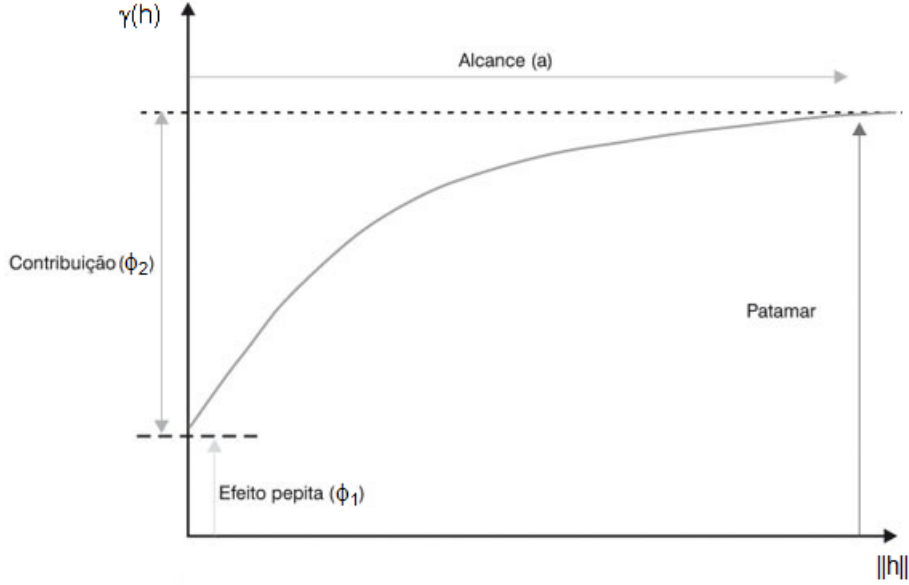


Figura 1.4: Parâmetros do semivariograma.

- Alcance ou *range* (a): distância a partir da qual as amostras passam a ser consideradas independentes, ou seja, a partir da qual a variação média entre duas observações não é mais função da distância entre elas. Quando essa distância tende a ∞ , considera-se que o semivariograma atinge um patamar apenas assintoticamente. Então, é usado um alcance chamado de alcance prático. Diggle and Ribeiro (2007) definem o alcance prático como sendo o valor de $\|h\|$ tal que o valor da correlação seja 0,05.

A seguir, apresentamos algumas famílias paramétricas de semivariogramas.

- **Modelo esférico:**

$$\gamma(h) = \begin{cases} 0 & \text{se } \|h\| = 0; \\ \phi_1 + \phi_2 \left(\frac{3}{2} \left[\frac{\|h\|}{\phi_3} \right] - \frac{1}{2} \left[\frac{\|h\|}{\phi_3} \right]^3 \right) & \text{se } 0 < \|h\| < \phi_3; \\ \phi_1 + \phi_2 & \text{se } \|h\| \geq \phi_3; \end{cases}$$

$$C(h) = \begin{cases} \phi_1 + \phi_2 & \text{se } \|h\| = 0; \\ \phi_2 \left(1 - \frac{3}{2} \left[\frac{\|h\|}{\phi_3} \right] + \frac{1}{2} \left[\frac{\|h\|}{\phi_3} \right]^3 \right) & \text{se } 0 < \|h\| < \phi_3; \\ 0 & \text{se } \|h\| \geq \phi_3; \end{cases}$$

em que $\theta = (\phi_1, \phi_2, \phi_3)$ e $\phi_3 > 0$. Esse modelo é válido em \mathbb{R}^d , $d = 1, 2$ e 3 . Apesar de atingir o patamar no valor $\|h\| = \phi_3$, Andriotti (2003) define seu alcance prático como sendo $2/3$ de ϕ_3 .

- **Modelo exponencial:**

$$\gamma(h) = \begin{cases} 0 & \text{se } \|h\| = 0; \\ \phi_1 + \phi_2 \left(1 - \left(e^{-\frac{\|h\|}{\phi_3}}\right)\right) & \text{se } \|h\| > 0; \end{cases}$$

$$C(h) = \begin{cases} \phi_1 + \phi_2 & \text{se } \|h\| = 0; \\ \phi_2 e^{-\frac{\|h\|}{\phi_3}} & \text{se } \|h\| > 0; \end{cases}$$

em que $\theta = (\phi_1, \phi_2, \phi_3)$ e $\phi_3 > 0$. Esse modelo é válido em \mathbb{R}^d , $d = 1, 2$ e 3 . O patamar é alcançado pela curva somente de forma assintótica. Seu alcance prático é igual a $3\phi_3$.

- **Modelo Gaussiano:**

$$\gamma(h) = \begin{cases} 0 & \text{se } \|h\| = 0; \\ \phi_1 + \phi_2 \left(1 - \left(e^{-\left(\frac{\|h\|}{\phi_3}\right)^2}\right)\right) & \text{se } \|h\| > 0; \end{cases}$$

$$C(h) = \begin{cases} \phi_1 + \phi_2 & \text{se } \|h\| = 0; \\ \phi_2 e^{-\left(\frac{\|h\|}{\phi_3}\right)^2} & \text{se } \|h\| > 0; \end{cases}$$

em que $\theta = (\phi_1, \phi_2, \phi_3)$ e $\phi_3 > 0$. Esse modelo é válido em \mathbb{R}^d , $d = 1, 2$ e 3 . Assim como o modelo exponencial, o modelo Gaussiano (também chamado de modelo parabólico) alcança o patamar apenas assintoticamente e seu alcance prático é igual a $\sqrt{3}\phi_3$. É o modelo que reflete mais regularidade da variável estudada e a presença do efeito pepita significa apenas erros de medida.

- **Modelo Matérn:**

$$\gamma(h) = \begin{cases} 0 & \text{se } \|h\| = 0; \\ \phi_1 + \phi_2 \left[1 - (2^{\kappa-1}\Gamma(\kappa))^{-1} \left(\frac{\|h\|}{\phi_3}\right)^\kappa K_\kappa\left(\frac{\|h\|}{\phi_3}\right)\right] & \text{se } \|h\| > 0; \end{cases}$$

$$C(h) = \begin{cases} \phi_1 + \phi_2 & \text{se } \|h\| = 0; \\ \phi_2 (2^{\kappa-1}\Gamma(\kappa))^{-1} \left(\frac{\|h\|}{\phi_3}\right)^\kappa K_\kappa\left(\frac{\|h\|}{\phi_3}\right) & \text{se } \|h\| > 0; \end{cases}$$

em que $\theta = (\phi_1, \phi_2, \phi_3, \kappa)$, $\phi_3 > 0$, $\kappa > 0$, K_κ é a função de Bessel modificada de ordem κ e Γ é a função Gama. Esse modelo foi proposto por Matérn (1960) e é válido em qualquer dimensão $d \geq 1$. Os modelos exponencial e Gaussiano são casos particulares do modelo Matérn, em que $\kappa = 0,5$ no primeiro caso e κ tende a ∞ no segundo caso. Pela dificuldade de identificabilidade de todos os parâmetros desse modelo, normalmente o valor de κ é

fixado e avaliado em vários valores.

1.3 Modelo geoestatístico

Tendo como base o processo geoestatístico, nosso interesse é modelar o processo e realizar inferências a partir de dados coletados em pontos conhecidos do espaço. A seguir introduzimos a definição do modelo geoestatístico.

Definição 1.5 *Seja um processo geoestatístico $Y(\mathbf{s})$ com domínio $D \subset \mathbb{R}^d$, $d \geq 1$. Definimos o seguinte modelo geoestatístico*

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (1.3)$$

em que tanto o componente determinístico $\mu(\mathbf{s})$ quanto o estocástico $\epsilon(\mathbf{s})$ dependem da localização espacial em que $Y(\mathbf{s})$ é observado. Assume-se que o componente de erro tem média nula, i.e., $E(\epsilon(\mathbf{s})) = 0$ e função de covariância $Cov(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j))$.

Como vimos na Seção 1.2.1, precisamos considerar hipóteses sobre o processo para podermos realizar inferência. Nesse trabalho, vamos considerar processos que obedeçam a condição de estacionariedade de segunda ordem, isto é, processos cujos primeiro e segundo momentos são constantes. Suponha que para algumas funções conhecidas de \mathbf{s}_i , $x_1(\mathbf{s}_i), x_2(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i)$, a média do processo pode ser expressa como

$$\mu(\mathbf{s}_i) = \sum_{j=1}^p \beta_j x_j(\mathbf{s}_i), \quad (1.4)$$

em que β_1, \dots, β_p são parâmetros a serem estimados. O modelo pode ser expresso na forma matricial

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (1.5)$$

em que $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'$, \mathbf{X} é a matriz de delineamento $n \times p$ com a i -ésima linha $\mathbf{x}'_i = (x_1(\mathbf{s}_i), x_2(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i))$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ e $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \dots, \epsilon(\mathbf{s}_n))'$. Supondo uma distribuição para o vetor $\boldsymbol{\epsilon}$, a matriz de covariância de \mathbf{Y} será a mesma que a matriz de covariância dos erros (correlação induzida). Pela Definição 1.1, $Cov(\mathbf{Y}) = E(U^{-1})\boldsymbol{\Sigma}$, em que a matriz de escala $\boldsymbol{\Sigma} = [(\sigma_{ij})]$, $\sigma_{ij} = Cov(\mathbf{s}_i, \mathbf{s}_j)$, pode ser expressa da seguinte forma paramétrica

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\phi}) = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}(\phi_3), \quad (1.6)$$

em que ϕ_1 , ϕ_2 e ϕ_3 são os parâmetros do semivariograma discutidos na Seção 1.2.2 e $\mathbf{R}(\phi_3) =$

$[(r_{ij})]$ representa a matriz de correlação, é uma matriz de dimensão $n \times n$ simétrica com elementos na diagonal $r_{ii} = 1$, para $i = 1, \dots, n$, com elementos $r_{ij} = (1/\phi_2)\sigma_{ij}$ para $i \neq j = 1, \dots, n$ e $\phi_2 \neq 0$, e $r_{ij} = 0$ para $i \neq j = 1, \dots, n$ e $\phi_2 = 0$.

1.3.1 Predição

Na análise geoestatística, a confecção de mapas temáticos é importante para descrição da variável resposta em todo o espaço de domínio. Para isso, os valores observados são utilizados para prever o valor da variável resposta em pontos que não foram amostrados e que são internos à região de estudo. Nesse caso, é realizado um processo de interpolação espacial. Esse processo, em geoestatística, é chamado de krigagem, desenvolvido por Matheron em homenagem à Daniel Krige pelo seu trabalho de minas de ouro na África do Sul (Andriotti, 2003).

Segundo Diggle and Ribeiro (2007), a predição espacial pode ser definida da seguinte forma: seja \mathbf{Y} o vetor de dimensão $n \times 1$ de variáveis aleatórias que são observadas e T uma variável aleatória a qual queremos prever seu valor a partir dos valores de \mathbf{Y} . Qualquer função de \mathbf{Y} , denotada como $\hat{T} = t(\mathbf{Y})$, é um preditor pontual para T . Estamos interessados no estimador que minimiza o erro quadrático médio (EQM) de \hat{T} , dado por

$$EQM(\hat{T}) = E[(\hat{T} - T)^2], \quad (1.7)$$

em que a esperança é em respeito à distribuição conjunta de T e \hat{T} ou, de forma equivalente, de T e Y . O estimador que minimiza (1.7) é mostrado pelo seguinte teorema.

Teorema 1.1 *A função $EQM(\hat{T})$ atinge o valor mínimo em $\hat{T} = E_T(T|Y)$.*

Prova. Podemos escrever (1.7) como

$$E[(\hat{T} - T)^2] = E_Y[E_T[(\hat{T} - T)^2|\mathbf{Y}]], \quad (1.8)$$

em que E_Y e E_T denotam a esperança em relação a \mathbf{Y} e a T , respectivamente. A esperança interna de (1.8) pode ser expressa na forma

$$E_T[(\hat{T} - T)^2|\mathbf{Y}] = Var_T\{(\hat{T} - T)|\mathbf{Y}\} + \{E_T[(\hat{T} - T)|\mathbf{Y}]\}^2. \quad (1.9)$$

Condicionando em \mathbf{Y} , a expressão (1.9) simplifica-se na forma

$$E_T[(\hat{T} - T)^2|\mathbf{Y}] = Var_T\{T|\mathbf{Y}\} + \{\hat{T} - E_T[T|\mathbf{Y}]\}^2. \quad (1.10)$$

Aplicando a esperança em \mathbf{Y} , a equação (1.10) é dada por

$$E[(\hat{T} - T)^2] = E_Y[Var_T[T|\mathbf{Y}]] + E_Y\{[\hat{T} - E_T[T|\mathbf{Y}]]^2\}. \quad (1.11)$$

O primeiro termo do lado direito de (1.11) não depende da escolha de \hat{T} . Então, minimizar (1.7) equivale a minimizar o segundo termo do lado direito de (1.11). Veja que este termo é não negativo e é igual a zero quando $\hat{T} = E_T(T|Y)$, o que conclui a prova. ■

O estimador pontual dado pelo Teorema (1.1) é simplesmente a esperança de T condicional a \mathbf{Y} . Note que, se conhecemos a distribuição de $T|\mathbf{Y}$, a média é apenas uma das medidas resumo que podemos utilizar. Além disso, não é invariante por transformação, isto é, se \hat{T} é o estimador que minimiza o EQM para T , em geral, $g(\hat{T})$ não é o estimador que minimiza o EQM para $g(T)$. O termo $Var_T[T|\mathbf{Y}]$ é chamado de variância de predição. Em particular, a variância de predição em um local observado estima o EQM de \hat{T} nesse ponto.

Na Seção 2.8 será discutida a forma de predição quando consideramos a classe de distribuição normal/independente no modelo geoestatístico.

1.3.2 Seleção de modelos

A partir de vários modelos de uma mesma análise, é interessante conhecer qual é o melhor dado algum critério. Entre os principais critérios de seleção de modelos, encontram-se o critério de informação de Akaike, o valor máximo da função de verossimilhança e a validação cruzada.

Critério de informação de Akaike: Akaike (1973) desenvolveu o critério de informação de Akaike, AIC , para a seleção de modelos que se baseia na minimização da divergência de Kullback-Leibler (D_{KL}). Akaike desenvolveu uma estimativa de D_{KL} a partir do valor da função de log-verossimilhança aplicada em seu ponto máximo acrescida de uma penalidade relacionada ao número de parâmetros do modelo. O AIC é definido como

$$AIC = -2l(\hat{\boldsymbol{\theta}}, \mathbf{y}) + 2p,$$

em que $l(\hat{\boldsymbol{\theta}}, \mathbf{y})$ é o valor da função de log-verossimilhança aplicada no ponto $\hat{\boldsymbol{\theta}}$ e p é o número de parâmetros do modelo. O termo de acréscimo é chamado de função de penalização e tem como objetivo corrigir um viés proveniente da comparação de modelos com número de parâmetros diferentes. Deve ser escolhido o modelo que apresentar o menor valor de AIC .

Valor máximo da função de verossimilhança: Na estimação de parâmetros via máxima verossimilhança, queremos encontrar o vetor de parâmetros $\hat{\theta}$ que faz com que a amostra obtida seja a mais verossímil, ou seja, maximiza a função de verossimilhança. Sendo o logaritmo uma função monótona crescente, o valor de θ que maximiza a função de verossimilhança é também o que maximiza a função de log-verossimilhança, que é mais fácil de ser analisada. Assim, o melhor modelo é aquele que apresenta o maior valor da função de log-verossimilhança aplicado no vetor $\hat{\theta}$.

Validação Cruzada: A validação cruzada é um método que permite avaliar a estimativa da predição. Assume-se que um determinado componente da amostra não tenha sido coletado e, então, é feita a predição de seu valor considerando os dados restantes. Andriotti (2003) define validação cruzada como o procedimento mediante o qual cada amostra é retirada do conjunto de dados e é feita uma krigagem para avaliar o valor da amostra retirada; o valor médio das diferenças será tão próximo de zero quanto melhor for a estimativa.

É necessário especificar uma função de covariância para poder realizar a krigagem. Sejam $Y(\mathbf{s}_i)$ e $\hat{Y}(\mathbf{s}_i)$ os valores observado e predito da variável resposta no local \mathbf{s}_i , respectivamente. Se o gráfico de $\hat{Y}(\mathbf{s}_i) \times Y(\mathbf{s}_i)$ apresentar resultados sobre uma reta de regressão próxima da bissetriz, há indícios de boas estimativas, ou seja, os valores preditos estão próximos dos valores reais. Também pode-se observar o coeficiente de correlação linear de Pearson, que deve apresentar valor próximo de um. Essa análise não prova que a função de covariância escolhida seja a correta, mas sim que não está grosseiramente incorreta. Em outras palavras, não sabemos se a função de covariância escolhida é a verdadeira, mas que pode ser utilizada para ter boas estimativas.

A partir dos resíduos $\epsilon_i = Y(\mathbf{s}_i) - \hat{Y}(\mathbf{s}_i)$, definimos o erro médio reduzido (*EMR*) como

$$EMR = \frac{1}{n} \sum_i^n \epsilon_i, \tag{1.12}$$

Espera-se que o valor do *EMR* próximo de zero.

1.4 Influência local

Uma das etapas de verificação de modelo é a análise de diagnóstico e Cook (1986) propõe um método baseado no estimador de máxima verossimilhança para avaliar a robustez de um modelo

submetido a pequenas perturbações.

Sejam $l(\boldsymbol{\theta})$ a função de log-verossimilhança do modelo postulado, $\boldsymbol{\theta}$ um vetor $p \times 1$ de parâmetros desconhecidos, $\boldsymbol{\theta} \in \mathbb{R}^p$ e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$ um vetor de perturbações, $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^n$. A função de log-verossimilhança do modelo perturbado é dada por $l(\boldsymbol{\theta}|\boldsymbol{\omega})$ e assumiremos que existe $\boldsymbol{\omega}_0 \in \Omega$ tal que $l(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = l(\boldsymbol{\theta})$, ou seja, o vetor $\boldsymbol{\omega}_0$ representa a ausência de perturbação.

Sejam $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ os EMV do modelo postulado e perturbado, respectivamente. Cook (1986) propõe estudar a comparação entre esses estimadores através do afastamento pela verossimilhança (*likelihood displacement*), $LD(\boldsymbol{\omega})$, dada por

$$LD(\boldsymbol{\omega}) = 2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})], \quad \boldsymbol{\omega} \in \Omega. \quad (1.13)$$

Em particular, $LD(\boldsymbol{\omega}_0) = 0$. O sentido da distância entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$, baseado em $LD(\boldsymbol{\omega})$, pode depender da concavidade da função de log-verossimilhança $l(\boldsymbol{\theta})$. Se $l(\boldsymbol{\theta})$ é suficientemente achatada, pode-se dizer que $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ estão próximos entre si, pois não haverá grande mudança de $l(\boldsymbol{\theta})$ após a perturbação. Se $l(\boldsymbol{\theta})$ for suficientemente concentrada em torno de $\hat{\boldsymbol{\theta}}$, estas estimativas podem estar distantes entre si (Cook, 1987).

A idéia de Cook (1986) é estudar o comportamento da função $LD(\boldsymbol{\omega})$ em uma vizinhança de $\boldsymbol{\omega}_0$. O gráfico de $LD(\boldsymbol{\omega})$ vs. $\boldsymbol{\omega}$ pode ser visto como a superfície geométrica formada pelos valores do vetor $(n + 1)$ -dimensional

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix},$$

em que $\boldsymbol{\omega}$ varia em Ω . $\alpha(\boldsymbol{\omega})$ é chamada de gráfico de influência local e é uma superfície em \mathbb{R}^{n+1} , sendo utilizada para avaliar a influência quando se varia $\boldsymbol{\omega}$ através de Ω . A Figura 1.5 apresenta uma ilustração gráfica da superfície $\alpha(\boldsymbol{\omega})$, mas essa visualização só é possível quando $p \leq 2$. Sendo $\hat{\boldsymbol{\theta}}$ o estimador de máxima verossimilhança do modelo postulado, então $LD(\boldsymbol{\omega})$ atinge um ponto de mínimo local em $\boldsymbol{\omega}_0$ e o método avalia a curvatura normal $\alpha(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$.

O método consiste em considerar o plano tangente (T_0) à superfície $\alpha(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$. Como $\boldsymbol{\omega}_0$ é um ponto de mínimo, então T_0 é paralelo a $\Omega \subset \mathbb{R}^n$. Cada vetor unitário $\mathbf{h} \in \Omega$ determina um plano que é ortogonal a T_0 . A interseção deste plano com a superfície é chamada de linha levantada e pode ser obtida considerando o gráfico $LD(\boldsymbol{\omega}_0 + a\mathbf{h})$ em função de a , $a \in \mathbb{R}$. A curvatura normal da linha levantada, denotada $C_{\mathbf{h}}$, é definida como a curvatura da curva plana $(a, LD(\boldsymbol{\omega}_0 + a\mathbf{h}))$ em $a = 0$.

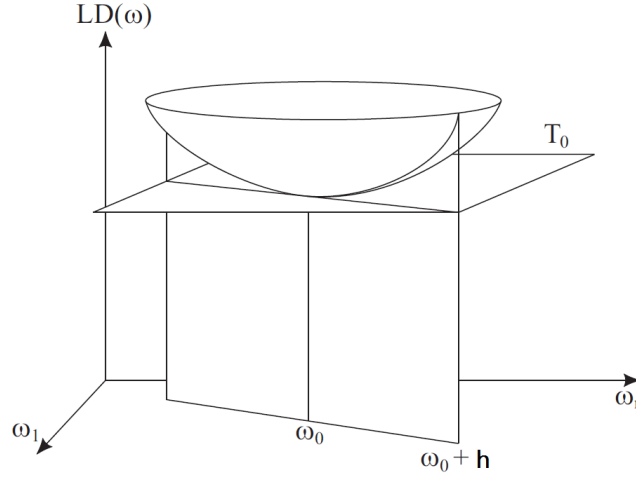


Figura 1.5: Representação do gráfico de influência local.

Cook (1986) mostra que a curvatura normal na direção \mathbf{h} é dada por

$$C_{\mathbf{h}}(\boldsymbol{\theta}) = 2|\mathbf{h}'\ddot{F}\mathbf{h}|, \quad (1.14)$$

em que $\ddot{F} = \boldsymbol{\Delta}'\ddot{L}^{-1}\boldsymbol{\Delta}$, $-\ddot{L}$ é a matriz de informação observada avaliada em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\Delta}$ é uma matriz $p \times n$ com elementos $\Delta_{ij} = \frac{d^2l(\boldsymbol{\theta}|\boldsymbol{\omega})}{d\theta_i d\omega_j}$, $i = 1, \dots, p$ e $j = 1, \dots, n$ avaliada em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$.

Há diversas maneiras de utilizar (1.14) para estudar a superfície de influência local, cada uma dependendo da escolha do vetor \mathbf{h} . Podemos analisar $\alpha(\boldsymbol{\omega})$ na direção de maior curvatura, C_{max} , que é o maior autovalor de \ddot{F} . Assim, \mathbf{h} será o autovetor associado ao autovalor C_{max} , denotado por \mathbf{h}_{max} . Os componentes de $|\mathbf{h}_{max}|$ são utilizados para identificar observações influentes. Outra escolha para o vetor \mathbf{h} é o vetor unitário \mathbf{h}_i , com 1 na posição correspondente a i -ésima observação e 0 nas demais. Então, podemos analisar a curvatura C_i correspondente à direção de cada observação separadamente. Segundo Paula (2010), a análise também pode ser feita considerando o autovetor, \mathbf{h}_{max2} , correspondente ao segundo maior autovalor. Assim, gráficos de C_i , $|\mathbf{h}_{max}|$ e $|\mathbf{h}_{max}|_2$ vs. ordem das observações podem revelar as observações influentes. As observações que apresentarem valores de C_i maiores que $2\bar{C}$, em que $\bar{C} = \sum_{i=1}^n C_i/n$, podem ser consideradas influentes (Paula, 2010).

Podemos introduzir a perturbação no modelo através do vetor $\boldsymbol{\omega}$ de várias maneiras. Por exemplo, podemos perturbar a variável resposta \mathbf{Y} , a matriz de delineamento \mathbf{X} , apenas uma covariável \mathbf{X}_i ou a matriz de escala $\boldsymbol{\Sigma}$. Neste trabalho utilizamos o esquema de perturbação aditiva na variável resposta, dada por $\mathbf{Y}_{\boldsymbol{\omega}} = \mathbf{Y} + b\boldsymbol{\omega}$, que permite identificar observações com grande influência nos estimadores de máxima verossimilhança. Nesse caso, $\boldsymbol{\omega}_0 = (0, \dots, 0)'$ e b é um fator de escala.

Capítulo 2

Modelo espacial linear com distribuição normal/independente

Este capítulo é destinado a desenvolver o estudo de um modelo geoestatístico com distribuição normal/independente. Nas Seções 2.1 e 2.2 apresentamos a descrição do modelo e a função de log-verossimilhança. Desenvolvemos a estimação dos parâmetros do modelo geoestatístico por máxima verossimilhança e algoritmo EM nas Seções 2.3 e 2.4. As matrizes de informação observada e esperada são apresentadas nas Seções 2.5 e 2.6. Nas Seções 2.7 e 2.8, discutimos o método de influência local e o método de predição para o modelo geoestatístico com distribuição normal/independente. O software utilizado para análises bem como pacotes requeridos são apresentados na Seção 2.9.

2.1 Descrição do modelo

Analogamente ao modelo descrito em (1.3), definimos o seguinte modelo geoestatístico

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (2.1)$$

em que tanto o componente determinístico $\mu(\mathbf{s})$ quanto o estocástico $\epsilon(\mathbf{s})$ dependem da localização espacial em que $Y(\mathbf{s})$ é observada. Assume-se que o componente de erro tenha distribuição normal/independente, isto é, $\epsilon \sim NI(\mathbf{0}, \Sigma, \nu)$ e função de covariância $Cov(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j))$. Na presença de covariáveis, o componente determinístico pode ser expresso como $\mu = \mathbf{X}\beta$. Pela Definição 1.1, temos que $\mathbf{Y} \sim NI(\mathbf{X}\beta, \Sigma, \nu)$.

2.2 Função de log-verossimilhança

Considere $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))'$ um vetor de realizações de $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'$, $\boldsymbol{\theta} = (\boldsymbol{\phi}', \boldsymbol{\beta}')$, em que $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$ ou $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3, \phi_4)'$ é o vetor de parâmetros especificados em (1.6) e $\boldsymbol{\beta}$ é o vetor de coeficientes do componente linear de (1.4). Desse modo, a função de log-verossimilhança é expressa por

$$l(\boldsymbol{\theta}, \mathbf{y}) = \log \int_0^\infty \frac{u^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{u\delta}{2}} dH(u). \quad (2.2)$$

Substituindo Σ por sua forma paramétrica dada em (1.6) e $\boldsymbol{\mu}$ em função das covariáveis como dado em (1.4), temos

$$l(\boldsymbol{\theta}, \mathbf{y}) = \log \int_0^\infty \frac{u^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} |\phi_1 \mathbf{I} + \phi_2 \mathbf{R}(\phi_3)|^{\frac{1}{2}}} e^{-\frac{u\delta}{2}} dH(u),$$

em que $\delta = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\phi_1 \mathbf{I} + \phi_2 \mathbf{R}(\phi_3))^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Considerando a função densidade de \mathbf{Y} na forma de distribuição de contornos elípticos dada em (1.1), a função de log-verossimilhança é expressa por

$$l(\boldsymbol{\theta}, \mathbf{y}) = -\frac{1}{2} \log |\Sigma| + \log g(\delta). \quad (2.3)$$

A forma da função de log-verossimilhança depende da distribuição de mistura $H(u)$ e a função é maximizada em relação ao vetor de parâmetros $\boldsymbol{\theta}$. A seguir, são apresentadas as funções de log-verossimilhança para as distribuições em estudo descritas na Seção 1.1.

(i) **Distribuição normal:**

$$l(\boldsymbol{\theta}, \mathbf{y}) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{\delta}{2}, \quad (2.4)$$

(i) **Distribuição t-Student:**

$$l(\boldsymbol{\theta}, \mathbf{y}) = \log C_t - \frac{1}{2} \log |\Sigma| - \left(\frac{\nu + n}{2}\right) \log \left(1 + \frac{\delta}{\nu}\right), \quad (2.5)$$

$$\text{em que } C_t = \frac{\Gamma((\nu + n)/2)}{\Gamma(\nu/2)(\nu\pi)^{n/2}}.$$

(i) **Distribuição slash:**

$$l(\boldsymbol{\theta}, \mathbf{y}) = \log C_s + \log \nu - \frac{1}{2} \log |\Sigma| - \left(\frac{n}{2} + \nu\right) \log \frac{\delta}{2} + \log F_S(1). \quad (2.6)$$

$$\text{em que } C_s = \frac{\Gamma(n/2 + \nu)}{(\nu\pi)^{n/2}}.$$

2.3 Estimação dos parâmetros via máxima verossimilhança

A estimação dos parâmetros se dá maximizando a função de log-verossimilhança (2.2) ou (2.3) em relação ao vetor de parâmetros $\boldsymbol{\theta}$. Por questões de simplicidade, será utilizada a forma em

Tabela 2.1: Valores de $-2V_{g(\delta)}$ para as distribuições em estudo.

Distribuição	$-2V_{g(\delta)}$
Normal	1
t-Student	$(\nu + n)/(\nu + \delta)$
Slash	$G\left(\frac{n}{2} + \nu + 1, \frac{\delta}{2}\right) / G\left(\frac{n}{2} + \nu, \frac{\delta}{2}\right)$

(2.3). Considerando que $g(\cdot)$ é contínua e diferenciável, podemos definir as quantidades

$$V_{g(\delta)} = \frac{d \log g(\delta)}{d\delta} = \frac{g'(\delta)}{g(\delta)}, \quad (2.7)$$

$$V'_{g(\delta)} = \frac{dV_{g(\delta)}}{d\delta}, \quad (2.8)$$

que dependem da distribuição de interesse. A Tabela 2.1 apresenta o valor de $-2V_{g(\delta)}$ para as distribuições em estudo sendo $G(\cdot)$ a função gama incompleta.

A função score é definida como o vetor gradiente da função de log-verossimilhança. Para o modelo em 2.1, é expressa como

$$U(\boldsymbol{\theta}) = (U(\boldsymbol{\beta}), U(\boldsymbol{\phi}))' = \left(\frac{dl(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\beta}}, \frac{dl(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\phi}} \right)'.$$

Quando igualamos a função score a zero, podemos encontrar o estimador do parâmetro de interesse caso esse tenha forma fechada. A função score para $\boldsymbol{\beta}$ é expressa por

$$U(\boldsymbol{\beta}) = \frac{dl(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\beta}} = -2V_{g(\delta)} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon},$$

em que $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$. Como resultado, o estimador de $\boldsymbol{\beta}$ possui forma fechada e pode ser expresso como

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y}. \quad (2.9)$$

A função score para o parâmetro $\boldsymbol{\phi}$ é expressa por

$$U(\boldsymbol{\phi}) = \frac{dl(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\phi}},$$

com elementos

$$U(\phi_j) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \right) - V_{g(\delta)} \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}, \quad j = 1, 2 \text{ e } 3. \quad (2.10)$$

O desenvolvimento pode ser observado no Apêndice A. Não é possível calcular o estimador de ϕ_j de forma fechada. Portanto, ϕ é estimado através da maximização da função de log-verossimilhança dado $\hat{\beta}$, ou seja,

$$\hat{\phi} = \operatorname{argmax}(l(\phi, \hat{\beta}, \mathbf{y})). \quad (2.11)$$

Note que a estimação do parâmetro β depende do parâmetro ϕ e vice-versa. Assim, é realizado um processo iterativo no qual, dados valores iniciais de $\beta^{(0)}$ e $\phi^{(0)}$, estimamos os parâmetros

$$\begin{aligned} \hat{\beta}^{(k+1)} &= (\mathbf{X}'\Sigma^{(k)-1}\mathbf{X})^{-1}\mathbf{X}\Sigma^{(k)-1}\mathbf{y}, \\ \hat{\phi}^{(k+1)} &= \operatorname{argmax}(l(\phi, \hat{\beta}^{(k)}, \mathbf{y})), \end{aligned}$$

em que $\hat{\beta}^{(k)}$ e $\hat{\phi}^{(k)}$ são os vetores das estimativas de β e ϕ na k -ésima iteração, respectivamente. Diggle and Ribeiro (2007) apresentaram uma reparametrização da matriz escala Σ dada em (1.6) para a distribuição normal, diminuindo o número de variáveis de três para duas no processo de otimização. A reparametrização é feita da seguinte maneira: sejam $\omega = \phi_1/\phi_2$ e $\mathbf{V} = \omega\mathbf{I} + \mathbf{R}(\phi_3)$. Reescreva a matriz de escala como $\Sigma = \phi_2\mathbf{V}$ na função de verossimilhança apresentada na Seção 2.2. O estimador do vetor de parâmetros β é dado por

$$\hat{\beta}(\mathbf{V}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}. \quad (2.12)$$

O estimador de ϕ_2 possui forma fechada (ver detalhes de contas no Apêndice B) e é dado por

$$\hat{\phi}_2(\mathbf{V}) = \delta^*/n, \quad (2.13)$$

em que $\delta^* = (\mathbf{Y} - \beta\mathbf{X})'\mathbf{V}^{-1}(\mathbf{Y} - \beta\mathbf{X})$.

Note que os estimadores $\hat{\beta}$ e $\hat{\phi}_2$ ficam em função de \mathbf{V} , que é de interesse estimar seus componentes. Para isso, temos que considerar \mathbf{V} conhecida e substituir (2.12) e (2.13) na expressão da função de log-verossimilhança apresentada em (2.4). Assim a função de log-verossimilhança concentrada é dada por

$$l(\omega, \phi_3) = -\frac{1}{2} \log |\mathbf{V}| - \frac{n}{2} \log \hat{\phi}_2(\mathbf{V}). \quad (2.14)$$

A expressão em (2.14) é otimizada numericamente em relação a ω e ϕ_3 , seguido por substituição para obter $\hat{\beta}$ e $\hat{\phi}_2$ (processo iterativo).

2.4 Estimação dos parâmetros via algoritmo EM

Dempster et al. (1977) propuseram um algoritmo iterativo de estimação de máxima verossimilhança quando as observações podem ser vistas como dados incompletos. Este método envolve duas etapas, uma de esperança (*Expectation*) e outra de maximização (*Maximization*), chamando-o de algoritmo EM. O algoritmo EM pode ser usado em duas situações: uma quando há realmente dados faltantes (devido a limitações de observação, por exemplo) e outra quando consideramos uma parte adicional dos dados como faltantes para assim poder simplificar a maximização da função de verossimilhança.

Para o modelo em estudo consideramos os dados de U (distribuição de mistura) como dados faltantes. Seja $\mathbf{y}_c = (\mathbf{y}', u)$ o vetor observado de dados completos e \mathbf{Y}_c o vetor aleatório que representa os dados completos. A função de verossimilhança completa de \mathbf{Y} e U é expressa por

$$f_{\mathbf{Y},U}(\mathbf{y}, u) = f_{\mathbf{Y}|U}(\mathbf{y}|u) \times h_U(u).$$

Considerando a forma hierárquica dada pela Definição 1.1, temos que

$$\begin{aligned} \mathbf{Y}|U = u &\sim N_n(\mathbf{X}\boldsymbol{\beta}, u^{-1}\boldsymbol{\Sigma}), \\ U &\sim h_U(u), \end{aligned}$$

e a função de log-verossimilhança completa fica

$$\begin{aligned} l_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \log f_{\mathbf{Y},U}(\mathbf{y}, u) = \log f_{\mathbf{Y}|U}(\mathbf{y}|u) + \log h_U(u) \\ &= -\frac{n}{2} \log(2\pi) + \frac{n}{2} \log(u) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{u}{2} \delta + \log h_U(u). \end{aligned} \quad (2.15)$$

A esperança condicional da função de log-verossimilhança para dados completos, também chamada de função Q , para o modelo descrito em (2.1) é dada por

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) &= -\frac{n}{2} \log(2\pi) + \frac{n}{2} E(\log(U)|\mathbf{Y}, \boldsymbol{\theta}^{(n)}) - \frac{1}{2} \log |\boldsymbol{\Sigma}| \\ &\quad - \frac{\delta}{2} E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)}) + E(\log h_U(u)|\mathbf{Y}, \boldsymbol{\theta}^{(n)}). \end{aligned}$$

em que $\boldsymbol{\theta}^{(n)}$ é a estimativa do vetor de parâmetros $\boldsymbol{\theta}$ na n -ésima iteração.

O termo $E(\log h_U(u)|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$ será necessário apenas se apresentar parâmetros a serem estimados, como graus de liberdade da distribuição t-Student ou slash. Podemos reescrever a função

Q desconsiderando os termos que não dependem do vetor de parâmetros $\boldsymbol{\theta}$ como

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{\delta}{2} E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)}). \quad (2.16)$$

No passo E, precisamos conhecer a distribuição de $U|\mathbf{Y}$, a qual dependerá da distribuição de mistura. A seguir apresentamos o valor do termo $E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$ para as distribuições em estudo e detalhes podem ser observados no Apêndice C.

(i) **Distribuição normal:** Nesse caso, U é uma variável aleatória degenerada em 1 e, então, $U|\mathbf{Y} = 1$ com probabilidade 1.

(ii) **Distribuição t de Student:** Nesse caso, temos que

$$U|\mathbf{Y} \sim Gama\left(\frac{n}{2} + \frac{\nu}{2}, \frac{1}{2}(\nu + \delta)\right),$$

e seu valor esperado

$$E(U|\mathbf{Y}) = \frac{n + \nu}{\nu + \delta}.$$

(iii) **Distribuição Slash:**

$$E(U|\mathbf{Y}) = \frac{G\left(\frac{n}{2} + \nu + 1, \frac{\delta}{2}\right)}{G\left(\frac{n}{2} + \nu, \frac{\delta}{2}\right)}.$$

No passo M, maximizamos a função Q . Derivando a função Q em relação a $\boldsymbol{\beta}$ e ϕ_j , $j = 1, 2$ e 3 , obtemos

$$\begin{aligned} \frac{dQ(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})}{d\boldsymbol{\beta}} &= -\frac{1}{2} E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)}) \frac{d\delta}{d\boldsymbol{\beta}} \\ \frac{dQ(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})}{d\phi_j} &= \frac{1}{2} \omega_1 \boldsymbol{\epsilon}' \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{\epsilon} - \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \right) \end{aligned}$$

em que $\omega_1 = E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$ e $\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

Nesse passo obtemos as seguintes expressões

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}, \quad (2.17)$$

$$\text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \right) = \omega_1 \boldsymbol{\epsilon}' \left(-\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{\epsilon}. \quad (2.18)$$

O desenvolvimento dessa solução pode ser observado no Apêndice D. Note que a distribuição de mistura influenciará na estimação dos parâmetros apenas pelo termo $E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$ na estimação de $\boldsymbol{\phi}$ (considerando os parâmetros da distribuição de U como fixos).

Note que a relação (2.18) não possui fórmula analítica e $\boldsymbol{\phi}$ deve ser estimado por

$$\hat{\boldsymbol{\phi}} = \operatorname{argmax} \left(Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}, \hat{\boldsymbol{\beta}}) \right). \quad (2.19)$$

Utilizando a reparametrização da matriz escala proposta por Diggle and Ribeiro (2007) e considerando $\omega = \phi_1/\phi_2$ e $\mathbf{V} = \omega\mathbf{I} + \mathbf{R}(\phi_3)$, os estimadores de $\boldsymbol{\beta}$ e ϕ_2 possuem forma fechada e são expressos por

$$\hat{\boldsymbol{\beta}}(\mathbf{V}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \quad (2.20)$$

$$\hat{\phi}_2(\mathbf{V}) = \frac{\delta^*}{n}E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)}), \quad (2.21)$$

em que $\delta^* = (\mathbf{Y} - \boldsymbol{\beta}\mathbf{X})'\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\beta}\mathbf{X})$. Substituindo (2.20) e (2.21) em (2.16), obtemos a função Q concentrada expressa por

$$Q_C(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) = -\frac{1}{2} \log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2} \log |\boldsymbol{\Sigma}|. \quad (2.22)$$

A expressão em (2.22) é otimizada numericamente em relação a ω e ϕ_3 . Os dois passos do algoritmo são repetidos até que haja convergência do vetor de parâmetros ou atinja o limite de iterações.

Alguns problemas podem ser encontrados quanto sua utilização. Por exemplo, o cálculo do termo $E(U|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$ para a distribuição slash pode causar alguns problemas computacionais pois, no processo iterativo de otimização, é possível apresentar "valores" como " ∞ " ou *NaN* (*Not a number*, normalmente ocasionado por uma divisão por 0).

2.5 Matriz de informação observada

A matriz de informação observada de Fisher é definida como $I(\boldsymbol{\theta}) = -\ddot{L}(\boldsymbol{\theta})$, em que

$$\ddot{L}(\boldsymbol{\theta}) = \ddot{L}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{d^2l(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\theta}d\boldsymbol{\theta}'} = \begin{pmatrix} \ddot{L}_{\beta\beta} & \ddot{L}_{\beta\phi} \\ \ddot{L}_{\phi\beta} & \ddot{L}_{\phi\phi} \end{pmatrix}.$$

Os cálculos de $\ddot{L}(\boldsymbol{\theta})$ para o modelo em (2.1), considerando a função de log-verossimilhança da forma (2.2) são inviáveis. Considerar a função de log-verossimilhança da forma de contornos elípticos (2.3) facilita os cálculos, além de fornecer expressões gerais de $\ddot{L}(\boldsymbol{\theta})$ para a classe de distribuições normal/independente. Assim, De Bastiani (2012) mostra que

$$\ddot{L}_{\beta\beta} = \frac{d^2 l(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\beta}d\boldsymbol{\beta}'} = 2\mathbf{X}'\boldsymbol{\Sigma}^{-1}(V_{g(\delta)}\boldsymbol{\Sigma} + 2V'_{g(\delta)}\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\boldsymbol{\Sigma}^{-1}\mathbf{X},$$

$$\ddot{L}_{\beta\phi} = \frac{d^2 l(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\beta}d\boldsymbol{\phi}'}, \text{ com elementos}$$

$$\frac{d^2 l(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\beta}d\boldsymbol{\phi}_j} = 2\mathbf{X}'\boldsymbol{\Sigma}^{-1} \left(V_{g(\delta)} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} + V'_{g(\delta)} \boldsymbol{\epsilon}\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} \right) \boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}, \text{ para } j = 1, 2 \text{ e } 3.$$

$$\ddot{L}_{\phi\beta} = \left(\ddot{L}_{\beta\phi} \right)',$$

$$\ddot{L}_{\phi\phi} = \frac{d^2 l(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\phi}d\boldsymbol{\phi}'}, \text{ com elementos}$$

$$\begin{aligned} \frac{d^2 l(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\phi}_i d\boldsymbol{\phi}_j} &= \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left(\frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i} \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} - \frac{d^2 \boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i d\boldsymbol{\phi}_j} \right) \right] + \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \left[V'_{g(\delta)} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}\boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} + \right. \\ &\quad \left. V_{g(\delta)} \left(\frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i} \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} - \frac{d^2 \boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i d\boldsymbol{\phi}_j} + \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}, \text{ para } j = 1, 2 \text{ e } 3. \end{aligned}$$

Considerando a forma paramétrica da matriz escala $\boldsymbol{\Sigma}$, suas derivadas de primeira e segunda ordem em relação a $\boldsymbol{\phi}$ são apresentadas no Apêndice E.

Para calcular o erro padrão das estimativas dos parâmetros, precisamos conhecer a matriz de informação esperada de Fisher, definida por $E[I(\boldsymbol{\theta})]$. Um resultado importante que pode ser visto em Geyer (2007) é sobre a convergência da matriz de informação observada para a matriz de informação esperada quando $n \rightarrow \infty$.

2.6 Matriz de informação esperada

O erro padrão das estimativas de máxima verossimilhança pode ser obtido através da matriz de informação esperada de Fisher, $\mathbf{K}(\boldsymbol{\theta})$, dada por

$$\mathbf{K}(\boldsymbol{\theta}) = E[I(\boldsymbol{\theta})] = \begin{pmatrix} E[-\ddot{L}_{\beta\beta}] & E[-\ddot{L}_{\beta\phi}] \\ E[-\ddot{L}_{\phi\beta}] & E[-\ddot{L}_{\phi\phi}] \end{pmatrix} = \begin{pmatrix} \mathbf{K}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\phi}) \end{pmatrix}.$$

De Bastiani (2012) apresenta a forma da matriz $\mathbf{K}(\boldsymbol{\theta})$ para a distribuição de contornos elípticos, que está relacionada à distribuição normal/independente, obtendo $\mathbf{K}(\boldsymbol{\beta}) = \frac{4d_g}{n} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$, e $\mathbf{K}(\boldsymbol{\phi})$ com elementos: $\kappa_{ij}(\boldsymbol{\phi}) = \frac{b_{ij}}{4} \left(\frac{4f_g}{n(n+2)} - 1 \right) + \frac{2f_g}{n(n+2)} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i} \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} \right)$, em que $d_g = E[V_{g(\delta)}^2 W]$, $f_g = E[V_{g(\delta)}^2 W^2]$, com $W = \|\mathbf{Z}\|^2$, $\mathbf{Z} \sim NI(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\nu})$ e $b_{ij} = \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_i} \right) \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\phi}_j} \right)$.

As esperanças d_g e f_g possuem formas fechadas para algumas distribuições, tais como

- Distribuição normal: $d_g = \frac{n}{4}$ e $f_g = \frac{n(n+2)}{4}$.

- Distribuição t-Student: $d_g = \frac{n}{4} \frac{\nu + n}{\nu + n + 2}$ e $f_g = \frac{n(n+2)}{4} \frac{\nu + n}{\nu + n + 2}$.

Para a distribuição slash, o cálculo de d_g e f_g envolve integrais complexas e a esperança é aproximada por técnicas de Monte Carlo (Garay, 2014). O erro padrão do vetor de parâmetros $\boldsymbol{\theta}$ é obtido através de $\text{diag}(K(\boldsymbol{\theta})^{-1})^{1/2}$.

2.7 Influência local no modelo espacial linear com distribuição normal/independente

Considere o modelo geoestatístico dado em (2.1) em que \mathbf{Y} possui distribuição normal/independente. Seja $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)'$ o vetor de perturbações, $\boldsymbol{\omega} \in \Omega$. Considerando a função densidade de \mathbf{Y} na forma de distribuição de contornos elípticos apresentada em (1.1), a função de log-verossimilhança do modelo perturbado é dada por

$$l(\boldsymbol{\theta}, \mathbf{y}|\boldsymbol{\omega}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| + \log g(\delta_{\boldsymbol{\omega}}), \quad (2.23)$$

em que $\delta_{\boldsymbol{\omega}} = (\mathbf{Y}_{\boldsymbol{\omega}} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{\boldsymbol{\omega}} - \mathbf{X}\boldsymbol{\beta})$ e $\mathbf{Y}_{\boldsymbol{\omega}} = \mathbf{Y} + b\boldsymbol{\omega}$ (esquema de perturbação aditiva). Para desenvolver a análise da curvatura normal $C_{\mathbf{h}}$ em (1.14), precisamos conhecer a matriz de informação observada (apresentada na Seção 2.5) e a matriz $\boldsymbol{\Delta}$, que possui elementos (Apêndice F)

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \frac{d^2 l(\boldsymbol{\theta}, \mathbf{y}|\boldsymbol{\omega})}{d\boldsymbol{\beta} d\boldsymbol{\omega}'} = -2\mathbf{X}'\boldsymbol{\Sigma}^{-1} \left(V_{g(\delta_{\boldsymbol{\omega}})}\boldsymbol{\Sigma} + 2V'_{g(\delta_{\boldsymbol{\omega}})}\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \right) \boldsymbol{\Sigma}^{-1}, \\ \boldsymbol{\Delta}_{\phi_i} &= \frac{d^2 l(\boldsymbol{\theta}, \mathbf{y}|\boldsymbol{\omega})}{d\phi_i d\boldsymbol{\omega}'} = -2\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_i} \boldsymbol{\Sigma}^{-1} \left(V'_{g(\delta_{\boldsymbol{\omega}})}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\boldsymbol{\Sigma}^{-1} + V_{g(\delta_{\boldsymbol{\omega}})}\mathbf{I}_n \right), \quad j = 1, 2 \text{ e } 3. \end{aligned}$$

O vetor \mathbf{h}_{max} é o autovetor normalizado correspondente ao maior autovalor da matriz \ddot{F} e o vetor \mathbf{h}_i fornece a curvatura C_i da linha levantada na direção da i -ésima observação. As componentes de ambos vetores contém informações sobre quais observações exercem maior influência sobre as estimativas dos parâmetros.

2.8 Predição

Na Seção 1.3.1 foi discutida a forma geral de predição para o modelo geoestatístico. Nessa seção, apresentaremos a forma de predição quando considerada a classe de distribuição normal/independente, ou seja, para o modelo definido em (2.1).

Seja $y_p = y(\mathbf{s}_p)$ o valor da variável resposta no local \mathbf{s}_p não amostrado, ou seja, no local em

que queremos prever seu valor. Seja também $\mathbf{y}_o = \mathbf{y}(\mathbf{s}_o)$ o vetor da variável resposta nos locais observados $\mathbf{s}_o = (\mathbf{s}_1, \dots, \mathbf{s}_n)$. Estamos interessados em encontrar a distribuição condicional $p(y_p|\mathbf{y}_o)$. Para as distribuições normal e t-Student, é fácil encontrar $p(y_p|\mathbf{y}_o)$ de forma fechada. No caso da distribuição slash, não há forma fechada para essa distribuição condicional.

2.8.1 Distribuição normal

Considere que (y_p, \mathbf{y}_o) tem distribuição normal multivariada, isto é,

$$(y_p, \mathbf{y}_o) \sim NM_{(1+n)} \left(\left(\begin{array}{c} \mu_p \\ \boldsymbol{\mu}_o \end{array} \right), \left(\begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right) \right),$$

em que μ_p é o valor médio de y_p , $\boldsymbol{\mu}_o$ é o vetor de médias do vetor \mathbf{y}_o , $\boldsymbol{\Sigma}_{11}$ corresponde à variância de y_p , $\boldsymbol{\Sigma}_{22}$ corresponde à matriz de covariâncias do vetor \mathbf{y}_o , $\boldsymbol{\Sigma}_{12}$ corresponde à matriz de covariâncias de y_p e \mathbf{y}_o e $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$. Note que podemos escrever os termos de média como função de covariáveis como em (1.4), ficando $\mu_p = \mathbf{X}_p\boldsymbol{\beta}$ e $\boldsymbol{\mu}_o = \mathbf{X}_o\boldsymbol{\beta}$.

Através de resultados relacionados à distribuição normal multivariada, é possível mostrar que $p(y_p|\mathbf{y}_o)$ é normal com

$$\begin{aligned} E[y_p|\mathbf{y}_o] &= \mu_p + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o), \\ Var[y_p|\mathbf{y}_o] &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned}$$

2.8.2 Distribuição t-Student

Considere que (y_p, \mathbf{y}_o) tem distribuição t-Student multivariada, isto é,

$$(y_p, \mathbf{y}_o) \sim tM_{(1+n)} \left(\left(\begin{array}{c} \mu_p \\ \boldsymbol{\mu}_o \end{array} \right), \left(\begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right), \nu \right),$$

em que ν corresponde ao graus de liberdade e os demais termos são como descritos na Seção (2.8.1). Roth (2013) mostra que a $p(y_p|\mathbf{y}_o)$ é uma distribuição t-Student com

$$\begin{aligned} E[y_p|\mathbf{y}_o] &= \mu_p + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o), \\ \boldsymbol{\Sigma}_{[y_p|\mathbf{y}_o]} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \\ Var[y_p|\mathbf{y}_o] &= \frac{\nu + (\mathbf{y}_o - \boldsymbol{\mu}_o)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o)}{\nu + n} \boldsymbol{\Sigma}_{[y_p|\mathbf{y}_o]} \\ \nu_{[y_p|\mathbf{y}_o]} &= \nu + n. \end{aligned} \tag{2.24}$$

Note que a expressão para a esperança e para a matriz escala é igual a do caso Gaussiano, mas o valor da variância é influenciado pelos valores observados \mathbf{y}_o .

2.8.3 Distribuição slash

Considerando a forma hierárquica da distribuição da variável resposta dada pela Definição 1.1, podemos escrever a distribuição condicional $p(y_p|\mathbf{y}_o)$ (também chamada de distribuição preditiva) na forma de distribuição marginal como

$$\begin{aligned} p(y_p|\mathbf{y}_o) &= \int p(y_p, u|\mathbf{y}_o) du \\ &= \int p(y_p|\mathbf{y}_o, u)p(u) du \end{aligned} \quad (2.25)$$

No caso da distribuição slash, não há forma fechada para a expressão em (2.25). A distribuição preditiva será aproximada utilizando amostras Monte Carlo (Gamerman, 1997). Lembrando que, para a distribuição slash, $U \sim Beta(\nu, 1)$. A distribuição $p(y_p|\mathbf{y}_o, u)$ é normal com

$$\begin{aligned} E[y_p|\mathbf{y}_o] &= \mu_p + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o), \\ Var[y_p|\mathbf{y}_o] &= u^{-1}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \end{aligned}$$

Então, para termos uma aproximação da distribuição preditiva, geramos um valor da distribuição de U e, a partir dele, geramos uma observação de $p(y_p|\mathbf{y}_o, u)$. Realizando esse processo N vezes, teremos uma amostra de tamanho N que se aproxima da distribuição preditiva. Desse modo, podemos calcular as medidas de interesse, no caso, a média e a variância de predição. Note que esse método pode ser utilizado para qualquer distribuição desde que se tenha as distribuições $p(u)$ e $p(y_p|\mathbf{y}_o, u)$, ou seja, poderia ser utilizado para as distribuições normal e t-Student. Mas as formas exatas das distribuições condicionais resultam em melhor precisão e, além disso, são computacionalmente mais eficientes, uma vez que não é necessário o uso de simulação de Monte Carlo.

2.8.4 Comparação

A classe de distribuição normal/independente é também chamada de mistura de escala normal devido ao fato de que a alteração é feita somente no fator de escala das distribuições. Note que a expressão para o valor esperado da variável resposta no local a ser predito é a mesma na predição do modelo para as distribuições apresentadas. Ou seja, se o vetor de parâmetros $\boldsymbol{\theta}$ fosse conhecido, os valores preditos seriam os mesmos independente da distribuição. O que

diferenciaria uma distribuição de outra seria a variância de predição, uma vez que estamos considerando fatores de escala distintos.

Na prática, não conhecemos o verdadeiro valor de θ , e então pode haver variações nas estimativas de predição pois é necessário, primeiro, encontrar $\hat{\theta}$ para então realizar a predição.

2.9 Considerações computacionais

Foram utilizados recursos computacionais para o desenvolvimento do trabalho. Os algoritmos foram desenvolvidos utilizando a linguagem de programação R (R Core Team, 2015). A seguir apresentamos o método de otimização numérica aplicado, assim como problemas que surgiram com sua aplicação ao longo do estudo, o método utilizado para simulação de processos espaciais e os principais pacotes necessários para as análises neste trabalho.

2.9.1 Método Numérico BFGS

Em problemas de otimização, quando se quer obter o máximo ou mínimo de uma função, muitas vezes não encontramos soluções analíticas para encontrar o zero da função derivada. Desse modo, se faz necessário o uso de métodos numéricos iterativos que tentam fornecer uma solução aproximada da verdadeira. Um ponto crítico pode ser local ou global e, segundo Press et al. (1988), encontrar uma solução global é um problema difícil. Press et al. (1988) apresentam vários métodos numéricos para maximização/minimização de uma função, tanto para funções unidimensionais como para multidimensionais. Observe que o nosso interesse é maximizar a função de log-verossimilhança em (2.11) e a função Q em (2.19) em relação ao vetor de parâmetros ϕ e a função de log-verossimilhança concentrada em (2.14) em relação ao vetor de parâmetros $(\omega, \phi_3)'$. Para isso, utilizaremos o método quase-Newton BFGS (Broyden-Fletcher-Goldfarb-Shanno) que pertence à classe de métodos gradientes. Os algoritmos dessa classe tem por objetivo minimizar uma função $f(\theta)$, com a seguinte sequência iterativa, dado $\theta^{(0)}$,

$$\theta^{(k+1)} = \theta^{(k)} + \lambda^{(k)} \mathbf{A}^{(k)} \Delta f(\theta^{(k)}) \quad (2.26)$$

em que \mathbf{A} é uma matriz positiva definida, $\Delta f(\theta^{(k)})$ é o gradiente da função $f(\theta)$ aplicado em $\theta^{(k)}$ e $\lambda^{(k)}$ é o tamanho do passo entre iteração, todas na iteração k . Dentro dessa classe, esses métodos pertencem à classe de métricas variáveis, e o método BFGS é descrito em Press et al. (1988). A vantagem dessa classe de métodos está no fato de que não é necessário calcular a matriz hessiana (de segundas derivadas) da função ao passo que é feita uma aproximação dessa matriz.

Método BFGS

Fazendo expansão de Taylor em $f(\boldsymbol{\theta})$ em torno de $\boldsymbol{\theta}_0$ e utilizando aproximação quadrática, temos

$$f(\boldsymbol{\theta}) \approx f(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \Delta f(\boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{A} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \quad (2.27)$$

em que $\Delta f(\boldsymbol{\theta}_0)$ é o gradiente de $f(\boldsymbol{\theta})$ aplicado em $\boldsymbol{\theta}_0$, \mathbf{A} é uma matriz positiva definida e uma aproximação da matriz Hessiana.

Dado, então, $\boldsymbol{\theta}^{(0)}$ e $\mathbf{A}^{(0)}$ (normalmente a matriz identidade), os procedimentos a seguir são repetidos até que haja a convergência de $\boldsymbol{\theta}^{(k)}$.

- (Passo 1). Calcule $\Delta f(\boldsymbol{\theta}^{(k)})$.
- (Passo 2). Obtenha uma direção $p^{(k)}$ resolvendo $p^{(k)} = -\mathbf{A}^{(k)-1} \Delta f(\boldsymbol{\theta}^{(k)})$.
- (Passo 3). Faça uma busca linear na direção de $p^{(k)}$ para achar um passo $\lambda^{(k)}$, isto é, faça

$$\lambda^{(k)} = \operatorname{argmin} f(\boldsymbol{\theta}^{(k)} + \lambda p^{(k)}).$$

- (Passo 4). Atualize $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \lambda^{(k)} p^{(k)}$.
- (Passo 5). Atualize a matriz \mathbf{A} e retorne ao Passo 1.

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \frac{y^{(k)} y^{(k)'}}{y^{(k)' s^{(k)}}} + \frac{\mathbf{A}^{(k)} s^{(k)} s^{(k)' } \mathbf{A}^{(k)}}{s^{(k)' } \mathbf{A}^{(k)} s^{(k)}},$$

em que $y^{(k)} = \Delta f(\boldsymbol{\theta}^{(k+1)}) - \Delta f(\boldsymbol{\theta}^{(k)})$ e $s^{(k)} = \lambda^{(k)} p^{(k)}$.

A convergência pode ser verificada pelo seguinte critério $|\Delta f(\boldsymbol{\theta}^{(k)})| < e$, e um valor pequeno e Press et al. (1988) utilizam $u = 10^{-8}$. A direção $p^{(k)}$ é chamada de direção de decréscimo, pois é a direção em que a função decresce de valor. A sequência de matrizes $\mathbf{A}^{(k)}$ são todas definidas positivas, o que garante que a direção $p^{(k)}$ será de fato a direção de decréscimo, diferentemente da matriz hessiana que pode ser negativa definida se o ponto atual estiver longe de um ponto de mínimo. Por isso o uso de aproximações da hessiana é melhor que o uso da matriz verdadeira.

Restrição do método computacional

Métodos numéricos estão sujeitos a erros no processo iterativo, principalmente quando se trata de funções complexas. Um problema recorrente durante o estudo foi a estimação de componentes negativas do vetor ϕ' . Durante o processo de otimização, a busca pelos valores também se dá na região negativa que não pertence ao espaço paramétrico de ϕ' . O maior problema é, em um passo do algoritmo, essa busca resultar em $|V|$ negativo e, então, o termo $\log |V|$ é inconsistente (não existe, de fato). Para podermos contornar esse problema, foi utilizada a reparametrização logarítmica para os componentes de ϕ' . Dessa forma, $k_i = \log(\phi_i)$ e podemos reescrever a forma paramétrica de Σ dada em (1.6)

$$\Sigma = e^{k_1} \mathbf{I} + e^{k_2} \mathbf{R}(e^{k_3}). \quad (2.28)$$

Assim, a otimização é feita em relação à (k_1, k_2, k_3) e o espaço paramétrico de cada componente é \mathbb{R} . No caso da reparametrização apresentada na Seção 2.3, temos que $k_\omega = \log(\omega) = \log(\phi_1) - \log(\phi_2)$ e a função em (2.14) é otimizada em relação à (k_ω, k_3) .

Porém, outro problema foi encontrado durante o processo de otimização considerando essa reparametrização. Podemos obter um vetor $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$ tal que, computacionalmente, a matriz \mathbf{V} se torna singular e, portanto, não invertível. Isso pode ocorrer, por exemplo, quando o processo resulta em um valor de ϕ_1 extremamente pequeno e um valor de ϕ_3 extremamente alto. A matriz Σ resultante teria, então, apenas o valor de ϕ_2 em todas suas entradas. Contornamos esse problema numérico da seguinte forma: na k -ésima iteração do algoritmo, teste se a matriz \mathbf{V} é invertível. Caso positivo, a otimização é feita. Caso contrário, atribuímos à função objetivo o valor $-\infty$ (pois estamos maximizando) e o algoritmo continua o processo.

2.9.2 Simulação de processos espaciais

Nos processos espaciais, as observações são correlacionadas e essa correlação é determinada por uma lei regida pelo fator de escala Σ . Considerando a forma paramétrica da matriz escala dada em (1.6), as funções de covariância descritas na Seção 1.3 determinam a associação das observações. Nos estudos de simulação, se torna necessário que a amostra seja simulada respeitando tal lei. Novamente considerando a forma hierárquica dada pela Definição 1.1, temos

que

$$\begin{aligned}\mathbf{Y}|U = u &\sim N_n(\mathbf{X}\boldsymbol{\beta}, u^{-1}\boldsymbol{\Sigma}), \\ U &\sim h_U(u).\end{aligned}$$

No caso da distribuição t-Student, $U \sim \text{Gama}(\nu/2, \nu/2)$, no caso da distribuição slash, $U \sim \text{Beta}(\nu, 1)$, e no caso da distribuição normal, $U = 1$ com probabilidade 1. Assim, gera-se um valor da variável aleatória U e, a partir desse valor, é gerado um vetor multivariado da variável resposta \mathbf{Y} . Dessa forma garantimos que a amostra seja correlacionada pela lei regida pelo fator de escala $\boldsymbol{\Sigma}$.

2.9.3 Software e pacotes utilizados

Todos os procedimentos computacionais foram feitos utilizando o software R (Ribeiro Jr. and Diggle, 2001), versão 3.2.2. Alguns pacotes auxiliaram no desenvolvimento dos algoritmos, são eles

- geoR (Ribeiro Jr. and Diggle, 2001): utilizado para construir a matriz de correlação $R(\phi_3)$ e gráficos de semivariograma.
- fields (Douglas Nychka et al., 2015): utilizado na construção de mapas temáticos.
- MASS (Venables and Ripley, 2002): utilizado para gerar um vetor aleatório normal multivariado no processo de simulação.
- sp (Pebesma and Bivand, 2005): utilizado para criar uma malha regular em áreas irregulares.

Capítulo 3

Estudo de simulações

Apresentamos a seguir estudos de simulações para avaliar a metodologia apresentada. A partir de valores conhecidos do vetor de parâmetros θ , foram geradas amostras seguindo o procedimento descrito na Seção 2.9.2 e os parâmetros do modelo foram estimados utilizando os métodos de máxima verossimilhança e algoritmo EM descritos nas Seções 2.3 e 2.4, respectivamente.

As simulações foram desenvolvidas em malha regular com uma unidade de distância. Foram simulados dados de tamanho 25, 49 e 100 em regiões quadradas de tamanho 5, 7 e 10, respectivamente. Na Figura 3.1 podemos observar a região correspondente à amostra de tamanho 100 identificando as localizações dos pontos amostrados.

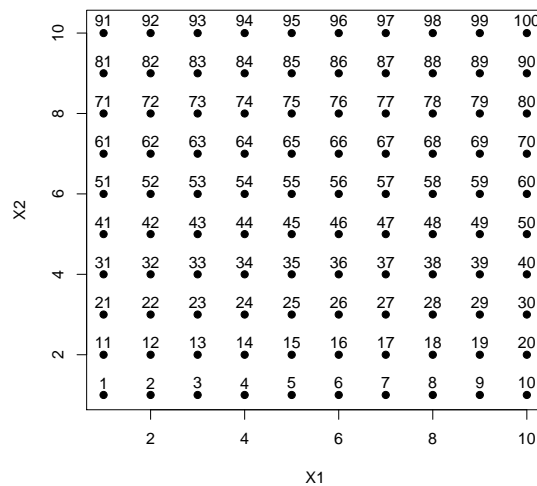


Figura 3.1: Malha regular de tamanho 100 em uma região quadrada com identificação dos locais amostrados.

Para facilitar notação, denote o vetor de média $\mu = \mu \mathbf{1}_n$, em que μ é um escalar e $\mathbf{1}_n$ é o vetor de 1's de tamanho n . Consideramos dois modelos, um com média constante e outro com covariáveis, expressos da seguinte forma

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (3.1)$$

$$Y(\mathbf{s}) = \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (3.2)$$

em que X_1 e X_2 representam as coordenadas no eixo x e y , respectivamente. Apresentamos na Seção 3.1 o estudo do desempenho dos algoritmos. Um estudo sobre o desempenho das es-

timizações por MV é feito na Seção 3.2. As análises de predição e influência local são apresentadas nas Seções 3.3 e 3.4.

3.1 Desempenho dos algoritmos

Para avaliar o desempenho do algoritmo, foram realizadas 500 simulações Monte Carlo de amostras de tamanho 25 e estimados os parâmetros por máxima verossimilhança (MV) e algoritmo EM. As amostras foram geradas considerando uma distribuição de interesse e os parâmetros foram estimados considerando a mesma distribuição. Por exemplo, amostras foram geradas e os parâmetros foram estimados ambos considerando a distribuição normal. As distribuições utilizadas neste estudo de simulações foram as distribuições normal, t-Student com 3, 5 e 10 g.l. e slash com $\nu = 3, 5$ e 10, função de covariância exponencial e vetor de parâmetros $\boldsymbol{\theta} = (\mu, \phi_1, \phi_2, \phi_3) = (2, 1, 1, 2)$. Para avaliar o desempenho das estimativas, foi utilizado o viés relativo absoluto (*absolute relative bias - ARB*) dado por

$$ARB(\hat{\theta}_j) = \left| \frac{\bar{\theta}_j - \theta_j}{\theta_j} \right| \times 100,$$

em que $\bar{\theta}_j = (1/N) \sum_{k=1}^N \hat{\theta}_{jk}$ e $\hat{\theta}_{jk}$ é a estimativa de $\theta_j \in \boldsymbol{\theta}$ da k -ésima simulação Monte Carlo. A Figura 3.2 apresenta o tempo de execução em segundos para realização da estimação por MV e algoritmo EM. Observamos que, para todas as distribuições consideradas, a estimação por MV apresentou menor tempo de execução do que a estimação por algoritmo EM. Para a distribuição slash, algumas amostras resultaram em não convergência do algoritmo EM, sendo 7 amostras para $\nu = 3$ e 2 amostras para $\nu = 5$ e 10. Essas amostras foram desconsideradas e novas foram simuladas para atingir o número de 500 simulações. As estimativas do vetor $\boldsymbol{\theta}$ e seus respectivos erros padrão entre parênteses se encontram na Tabela 3.1. A Figura 3.3 apresenta o gráfico de ARB (%) de $\hat{\boldsymbol{\theta}}$ e note que o ARB de $\hat{\mu}$ é pequeno, apresentando valores menores que 5% (Garcia-Papani et al., 2016). Para os demais parâmetros, observamos valores de ARB grandes, sendo maiores para $\hat{\phi}_1$ e $\hat{\phi}_2$ considerando a distribuição t-Student com 3 g.l. Notamos que não há um comportamento padrão de qual algoritmo teve melhor desempenho através do ARB . Desse modo, neste trabalho as análises foram feitas considerando a estimação por MV devido ao seu melhor desempenho em relação ao tempo de execução.

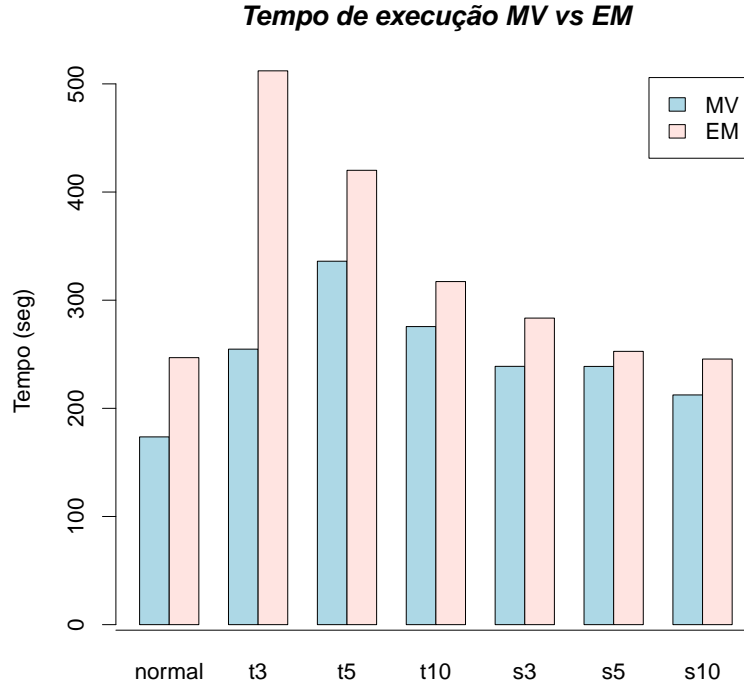
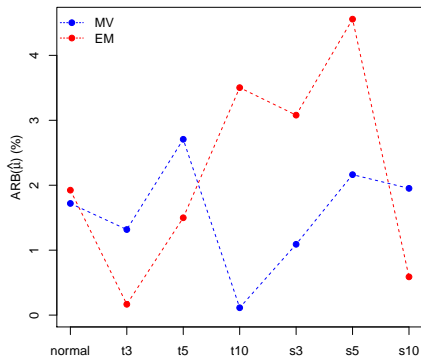


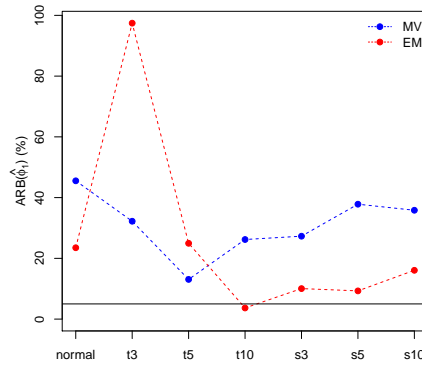
Figura 3.2: Gráfico de tempo de execução (em segundos) para estimação do vetor de parâmetros θ por MV e algoritmo EM de 500 simulações considerando tamanho de amostra 25 e função de covariância exponencial.

Tabela 3.1: Resultado da estimação do vetor de parâmetros θ e erros padrão (entre parênteses) por MV e algoritmo EM para o estudo de simulação considerando as distribuições normal, t-Student com 3, 5 e 10 g.l. e slash com $\nu = 5$.

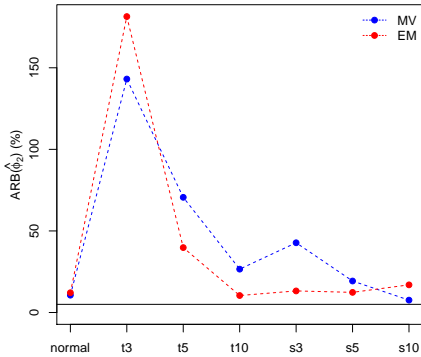
Distribuição	Algoritmo	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
Normal	MV	1,965 (0,627)	0,544 (0,627)	1,106 (0,807)	0,764 (0,706)
	EM	1,961 (0,613)	0,765 (0,697)	0,878 (0,895)	0,661 (0,707)
t-Student 3 g.l.	MV	2,026 (0,958)	1,322 (3,234)	2,431 (6,030)	0,687 (0,606)
	EM	1,996 (0,982)	1,974 (5,700)	2,813 (9,511)	0,643 (0,729)
t-Student 5 g.l.	MV	2,054 (0,752)	0,869 (1,612)	1,706 (2,093)	0,725 (0,718)
	EM	2,029 (0,716)	1,249 (1,903)	1,397 (2,230)	0,688 (0,747)
t-Student 10 g.l.	MV	1,997 (0,644)	0,737 (0,948)	1,265 (1,276)	0,727 (0,678)
	EM	2,070 (0,710)	0,963 (1,028)	1,104 (1,862)	0,633 (0,719)
Slash $\nu = 3$	MV	1,978 (0,806)	0,727 (0,900)	1,427 (1,912)	0,724 (0,675)
	EM	1,938 (0,746)	0,899 (1,086)	1,132 (1,499)	0,634 (0,813)
Slash $\nu = 5$	MV	2,043 (0,702)	0,621 (0,737)	1,193 (1,012)	0,767 (0,710)
	EM	1,908 (0,654)	0,907 (0,869)	0,876 (0,967)	0,639 (0,739)
Slash $\nu = 10$	MV	2,039 (0,623)	0,641 (0,692)	1,075 (0,926)	0,761 (0,719)
	EM	1,988 (0,646)	0,839 (0,721)	0,830 (0,905)	0,589 (0,673)



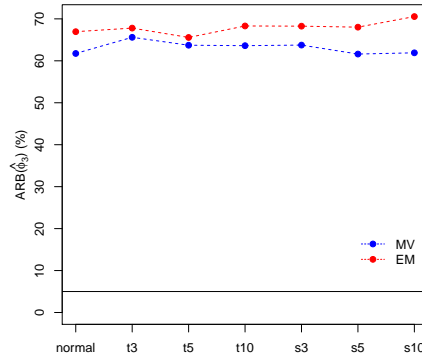
(a)



(b)



(c)



(d)

Figura 3.3: Gráfico de ARB (%) das estimativas de (a) $\hat{\mu}$, (b) $\hat{\phi}_1$, (c) $\hat{\phi}_2$ e (d) $\hat{\phi}_3$ realizadas por MV e algoritmo EM de 500 simulações considerando tamanho de amostra 25 e função de covariância exponencial.

3.2 Desempenho das estimativas realizadas por máxima verossimilhança

O objetivo deste estudo de simulações é avaliar o desempenho das estimativas dos parâmetros do modelo descrito em (2.1) a partir de variações no tamanho de amostra, distribuições e nos valores dos parâmetros verdadeiros.

Para analisar o bom desempenho das estimativas dos parâmetros da estrutura de média do modelo descrito em (2.1), fixamos o vetor de parâmetros $\phi = (1, 1, 2)$. Foram considerados os modelos com média constante (3.1) e com covariáveis (3.2) e vários valores para os parâmetros da média de cada modelo, tais como:

- Modelo sem covariável. μ : -2; -1; 0; 1 e 2.
- Modelo com covariável. $(\beta_0, \beta_1, \beta_2)$: (0; -2; -1); (0; -1; -0,5); (0; 0; 0); (0; 1; 0,5); (0; 2; 1).

Para analisar o bom desempenho das estimativas dos parâmetros da estrutura de covariância, foi considerado um modelo com média constante $\mu = 2$. Tendo como base o vetor de parâmetros $\phi = (1, 1, 2)$, cada parâmetro ϕ_i , $i = 1, 2$ e 3 , foi variado mantendo os demais fixos. Os valores considerados de ϕ_i na análise foram:

- ϕ_1 : 0; 0,2; 0,4; 0,6; 0,8; 1,0; 1,5; 2; 5 e 10.
- ϕ_2 e ϕ_3 : 0,1; 0,2; 0,4; 0,6; 0,8; 1,0; 1,5; 2; 5 e 10.

Por exemplo, para a análise do desempenho da estimativa de ϕ_1 , utilizamos os vetores de parâmetros (2;0;1;2), (2;0,2;1;2), (2;0,4;1;2) e assim por diante. Chamaremos de cenário cada combinação do vetor de parâmetros θ .

Foram realizadas 500 simulações para cada cenário e foram consideradas as funções de covariância exponencial e Gaussiana. Esse processo foi realizado para as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$.

3.2.1 Variação no parâmetro da média

Considerando os tamanhos de amostra 25, 49 e 100 e funções de covariância exponencial e Gaussiana, foi desenvolvida a análise do desempenho das estimativas do termo de média para os modelos de média constante e de covariáveis.

Sem covariável

Como pode ser observado na Figura 3.4, o parâmetro da média é bem estimado em todos os

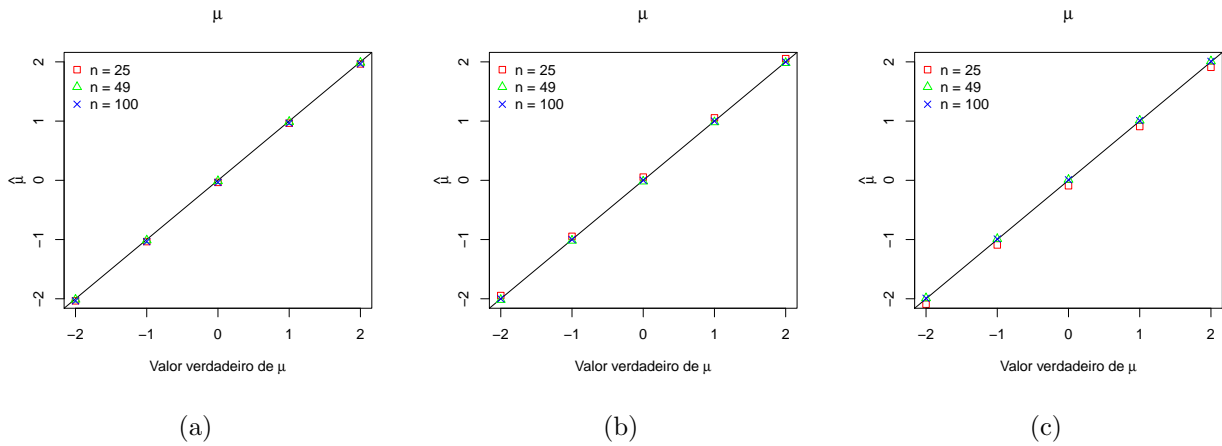


Figura 3.4: Gráfico de $\hat{\mu}$ vs. μ considerando a função de covariância exponencial e distribuição (a) normal, (b) t-Student com 5 g.l. e (c) slash com $\nu = 5$.

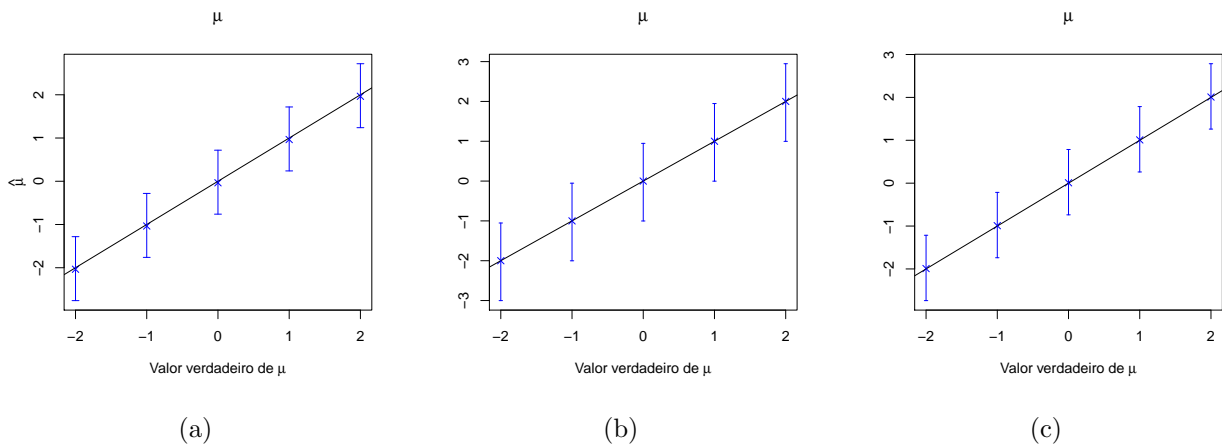


Figura 3.5: Intervalo de 95% de confiança empírico para $\hat{\mu}$, considerando a função de covariância exponencial, distribuição (a) normal, (b) t-Student com 5 g.l. e (c) slash com $\nu = 5$ e tamanho de amostra 100.

cenários apresentando valores bem próximos da bissetriz, o que significa que o valor estimado está próximo do valor verdadeiro. Observamos também que as estimativas do parâmetro para todos os tamanhos de amostra considerados foram similares. Além disso, foi construído um intervalo de 95% de confiança empírico para o parâmetro μ considerando o tamanho de amostra 100. Notamos, pela Figura 3.5, que os intervalos de confiança possuem a mesma amplitude independente do cenário. Note também que para diferentes valores do parâmetro da média a estimativa de $\hat{\phi}$ foi a mesma para todos os cenários (vide Tabela 3.2). Isso indica que o estimador do parâmetro da média é robusto.

Com covariável

Note que o parâmetro β_0 do modelo com covariáveis não foi alterado, tendo valor 0 em todos os cenários. Para cada distribuição, as estimativas de β_0 foram praticamente iguais para todos

Tabela 3.2: Estimativa do vetor de parâmetros ϕ considerando as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$ e parâmetros verdadeiros $\phi = (1, 1, 2)$.

Distribuição	Normal			t-Student			Slash		
n	25	49	100	25	49	100	25	49	100
$\hat{\phi}_1$	0,445	0,543	0,695	0,795	0,828	1,405	0,667	0,608	0,770
$\hat{\phi}_2$	1,205	1,245	1,185	1,779	1,991	2,085	1,116	1,312	1,264
$\hat{\phi}_3$	0,752	1,208	1,596	0,704	1,137	1,569	0,706	1,153	1,673

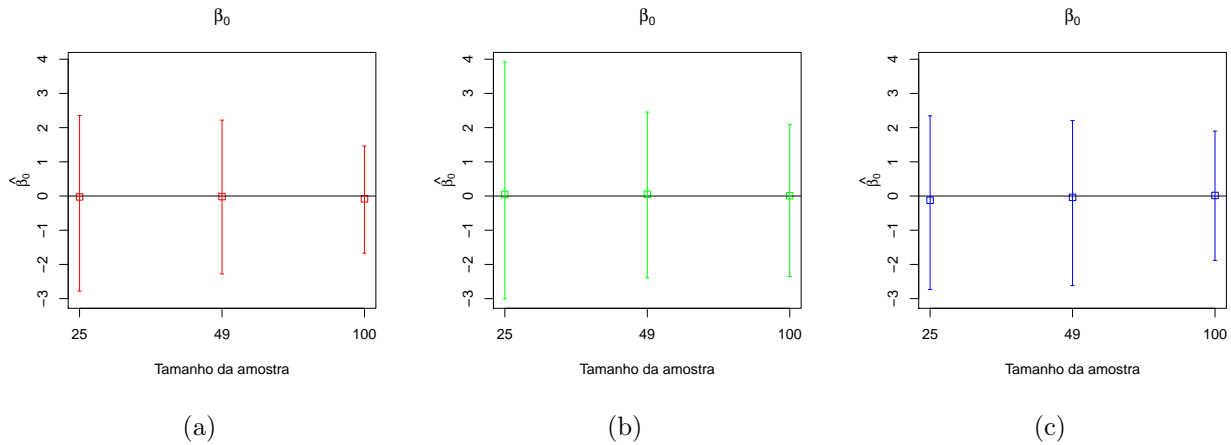


Figura 3.6: Gráfico de $\hat{\beta}_0$ vs. β_0 considerando a função de covariância exponencial e (a) distribuição normal, (b) t-Student com 5 g.l. e (c) slash com $\nu = 5$.

os tamanhos de amostra e observamos na Figura 3.6 que o parâmetro é bem estimado, apresentando estimativas em torno do valor verdadeiro 0. Além disso, a amplitude do intervalo de 95% de confiança empírico é reduzida quando se aumenta o tamanho da amostra, o que indica melhor precisão na estimação. Pela Figura 3.7, notamos que os parâmetros β_1 e β_2 também são bem estimados, apresentando estimativas próximas aos valores verdadeiros. Assim como no modelo sem covariável, os intervalos de confiança possuem a mesma amplitude para todos os cenários considerando um tamanho de amostra fixo. Desse modo, a Figura 3.8 apresenta o intervalo de 95% de confiança empírico para β_1 e β_2 considerando o cenário com $\beta = (0, 2, 1)$. Notamos que a amplitude do intervalo de confiança é reduzida com o aumento do tamanho de amostra, indicando maior precisão das estimativas. A variação no vetor de parâmetros β não resultou em estimativas diferentes do vetor de parâmetros de escala ϕ , que foi mantido constante $\phi = (1, 1, 2)$, sendo que a estimativa $\hat{\phi}$ foi a mesma para todos os cenários (vide Tabela 3.3). Isso indica que o estimador do vetor de parâmetros β é robusto.

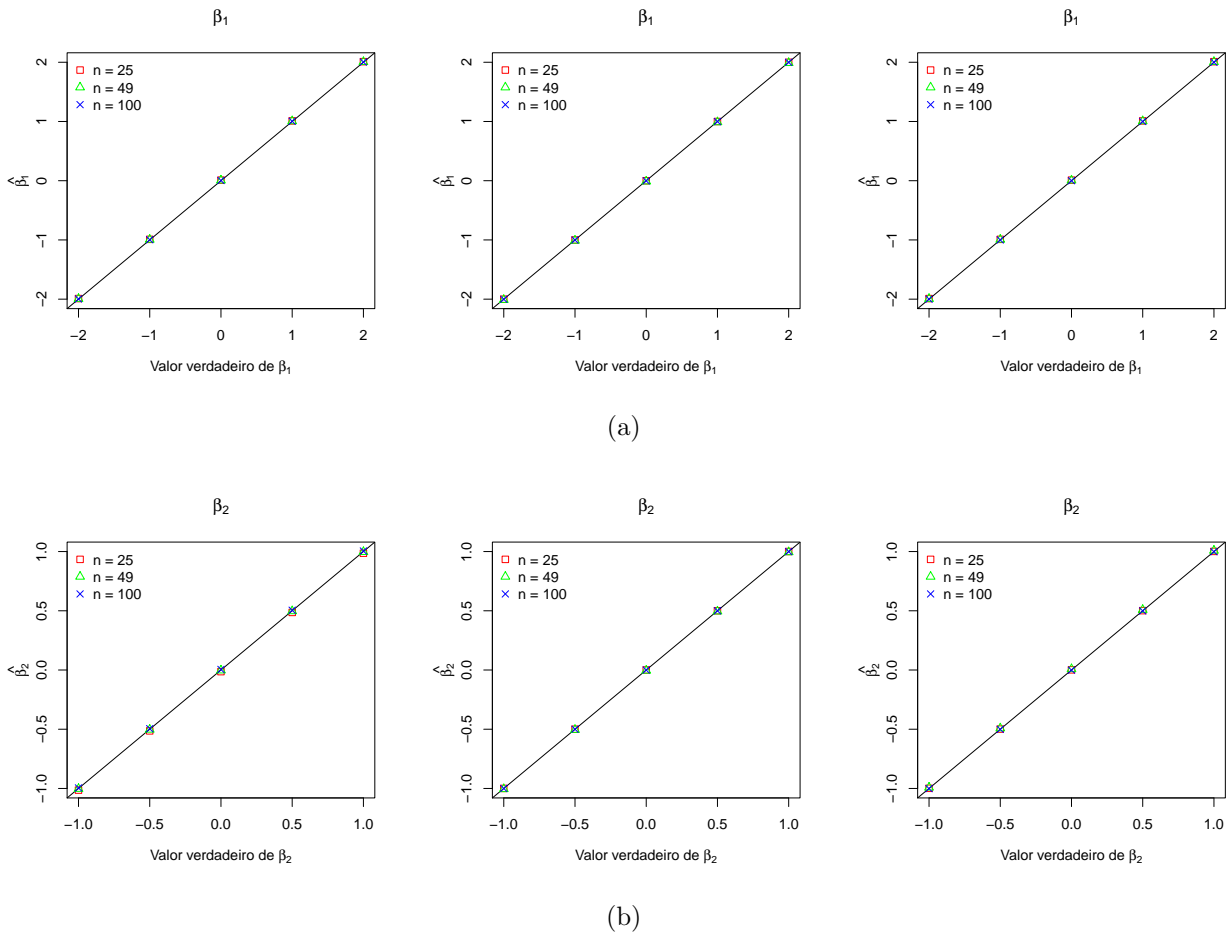
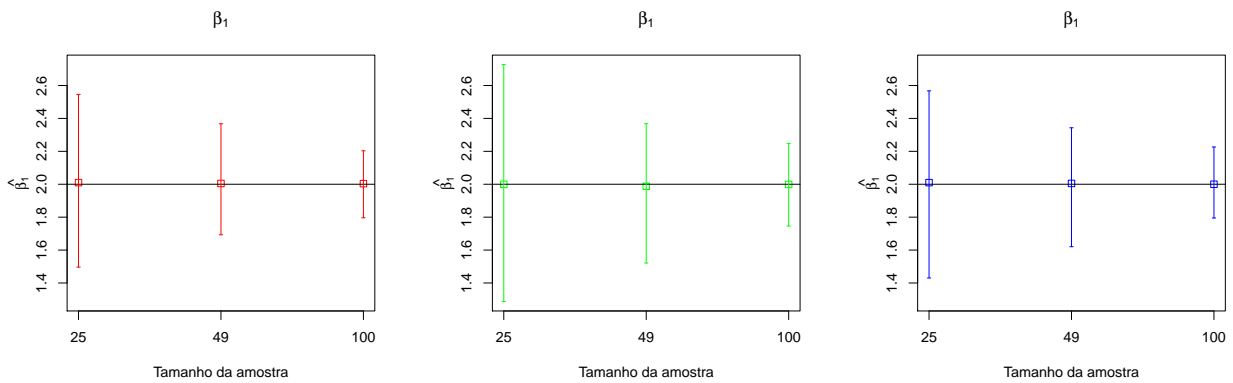


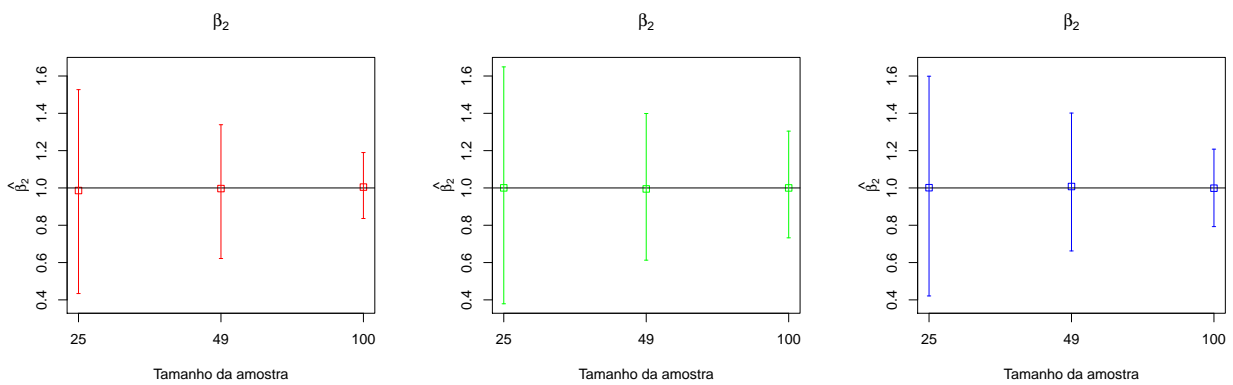
Figura 3.7: Gráficos de (a) $\hat{\beta}_1$ vs. β_1 e (b) $\hat{\beta}_2$ vs. β_2 considerando a função de covariância exponencial e distribuição normal, t-Student com 5 g.l. e slash com $\nu = 5$, respectivamente.

Tabela 3.3: Estimativa do vetor de parâmetros ϕ considerando estimação com covariáveis para a estrutura da média, distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$ e parâmetros verdadeiros $\phi = (1, 1, 2)$.

Distribuição	Normal			t-Student			Slash		
	25	49	100	25	49	100	25	49	100
$\hat{\phi}_1$	0,500	0,438	0,549	0,962	0,805	1,169	0,800	0,556	0,571
$\hat{\phi}_2$	0,803	1,061	1,134	1,108	1,628	1,924	0,623	1,058	1,246
$\hat{\phi}_3$	0,237	0,449	0,787	0,230	0,427	0,780	0,228	0,433	0,834



(a)



(b)

Figura 3.8: Intervalo de 95% de confiança empírico para (a) β_1 e (b) β_2 considerando $\beta = (0, 2, 1)'$, função de covariância exponencial e distribuição normal, t-Student com 5 g.l. e slash com $\nu = 5$, respectivamente.

3.2.2 Variação nos parâmetros de escala

Considerando os tamanhos de amostra 25, 49 e 100 e funções de covariância exponencial e Gaussiana, foi desenvolvida a análise do desempenho das estimativas do vetor de parâmetros de escala ϕ para o modelo de média constante.

Considerando a função de covariância exponencial, as Figuras 3.9, 3.10 e 3.11 apresentam as estimativas de ϕ para os cenários que variam os valores de ϕ_1 , ϕ_2 e ϕ_3 , respectivamente. Podemos observar nas Figuras 3.9 (a), 3.10 (b) e 3.11 (c), que, para valores de ϕ_1 , ϕ_2 e ϕ_3 próximos de zero, $\hat{\phi}_1$, $\hat{\phi}_2$ e $\hat{\phi}_3$ são menos viesadas, ou seja, se aproximam mais do valor verdadeiro em comparação com valores maiores, como 5 e 10. Entretanto, ao se aumentar o tamanho da amostra, esse vício diminui. Pelas Figuras 3.9 (b) e (c), note que, ao variar o valor do parâmetro ϕ_1 , as estimativas dos parâmetros ϕ_2 e ϕ_3 sofrem variações consideráveis. Para tamanho de amostra 25, os valores de $\hat{\phi}_2$ e $\hat{\phi}_3$ passam de 0,735 e 1,604 quando $\phi_1 = 0$ para 5,45 e 0,341 quando $\phi_1 = 10$, respectivamente. Nas Figuras 3.10 (a) e (c), observe que, ao variar o valor do parâmetro ϕ_2 , as estimativas do parâmetro ϕ_1 não sofrem grandes alterações enquanto que as estimativas do parâmetro ϕ_3 sofrem variações consideráveis, passando de 0,337 para 1,360 quando ϕ_2 passa de 0,1 para 10 (considerando tamanho de amostra 25). Note nas Figuras 3.11 (a) e (b) que, ao variar o valor do parâmetro ϕ_3 , as estimativas dos parâmetros ϕ_1 e ϕ_2 sofrem variações consideráveis. Então, observe que se o valor de um parâmetro ϕ_i é alterado quando os demais são mantidos fixos, a estimativa de ϕ é alterada e não somente a do parâmetro que foi variado. Isso ocorre devido a correlação existente entre os parâmetros ϕ_i 's, como pode ser observado na Figura 3.12, que apresenta o resultado de simulação para 5000

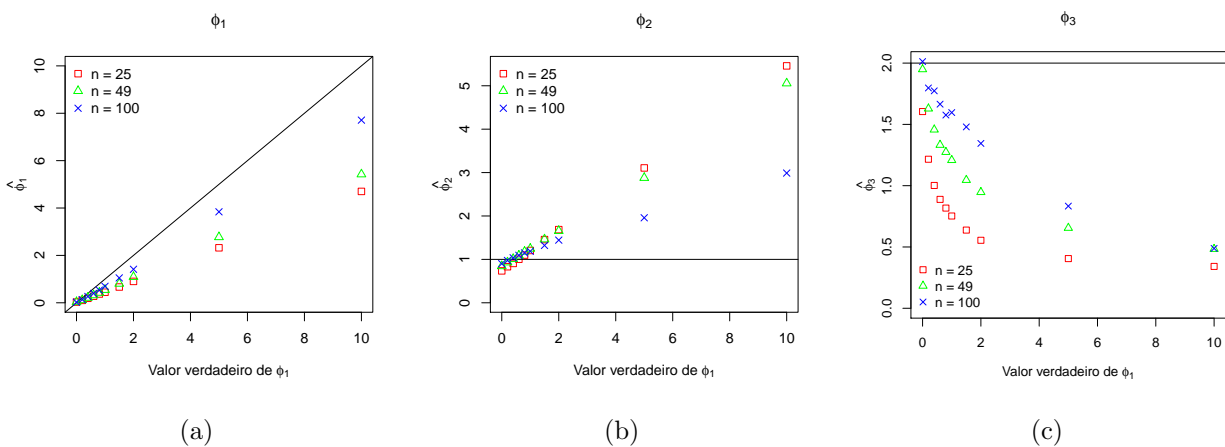


Figura 3.9: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de ϕ_1 considerando a função de covariância exponencial e distribuição normal. Gráfico de (a) $\hat{\phi}_1$, (b) $\hat{\phi}_2$ e (c) $\hat{\phi}_3$ em função do verdadeiro valor de ϕ_1 .

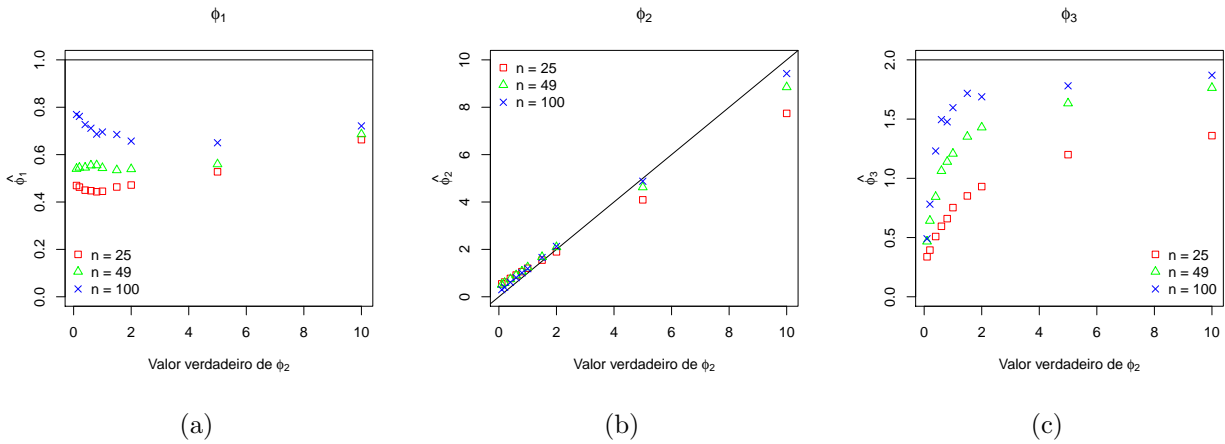


Figura 3.10: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de ϕ_2 considerando a função de covariância exponencial e distribuição normal. Gráfico de (a) $\hat{\phi}_1$, (b) $\hat{\phi}_2$ e (c) $\hat{\phi}_3$ em função do verdadeiro valor de ϕ_2 .

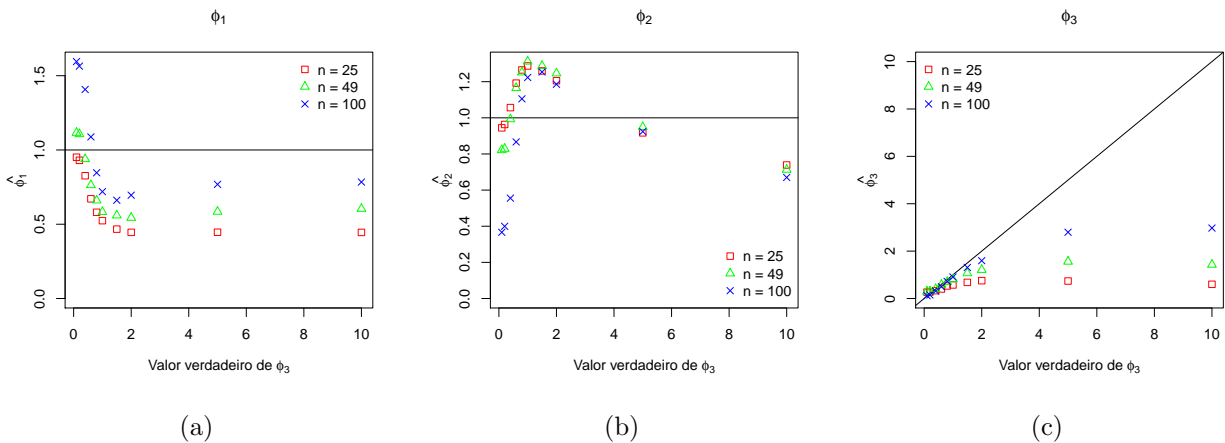


Figura 3.11: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de ϕ_3 considerando a função de covariância exponencial e distribuição normal. Gráfico de (a) $\hat{\phi}_1$, (b) $\hat{\phi}_2$ e (c) $\hat{\phi}_3$ em função do verdadeiro valor de ϕ_3 .

simulações considerando a distribuição normal com vetor de parâmetros $\theta = (2, 1, 1, 2)$, função de covariância exponencial e tamanho de amostra 25.

Os resultados obtidos considerando a função de covariância Gaussiana apresentaram o mesmo comportamento, ver Figuras 3.13, 3.14 e 3.15.

Como as estimativas $\hat{\phi}$ para o tamanho de amostra 100 apresentaram menor viés comparado às estimativas para os tamanhos de amostra 25 e 49, foram construídos intervalos de 95% confiança empíricos para o vetor ϕ considerando o tamanho de amostra 100. Esse resultado pode ser observado na Figura 3.16 para as funções de covariância exponencial e Gaussiana, respectivamente. Note que os intervalos de confiança para ϕ_1 e ϕ_3 possuem grande amplitude, com limite inferior próximo de zero para quase todos os cenários. Isso ocorre pelo fato de termos

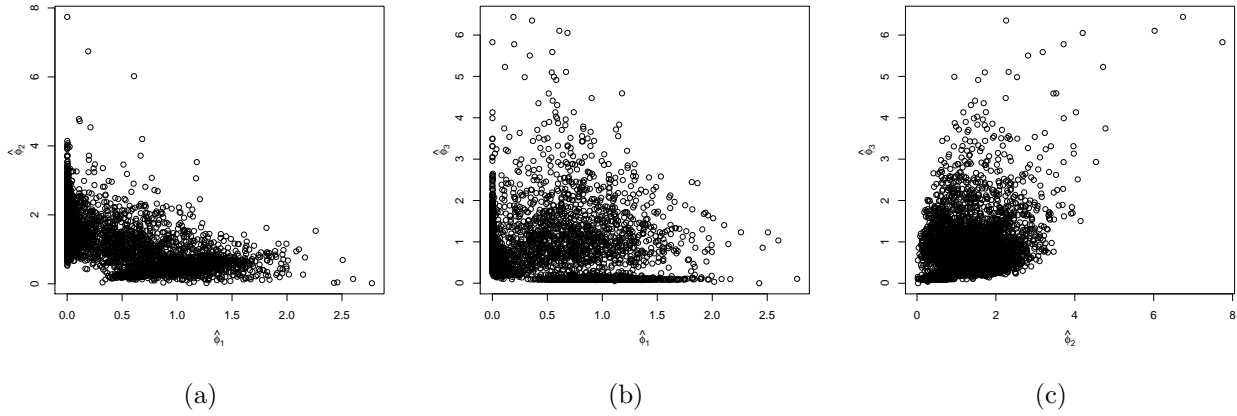


Figura 3.12: Gráficos de (a) $\hat{\phi}_2$ vs. $\hat{\phi}_1$, (b) $\hat{\phi}_3$ vs. $\hat{\phi}_1$ e (c) $\hat{\phi}_3$ vs. $\hat{\phi}_2$ de 5000 simulações de um modelo ajustado com distribuição normal, função de covariância exponencial, tamanho de amostra 25 e vetor de parâmetros $\theta = (2, 1, 1, 2)$.

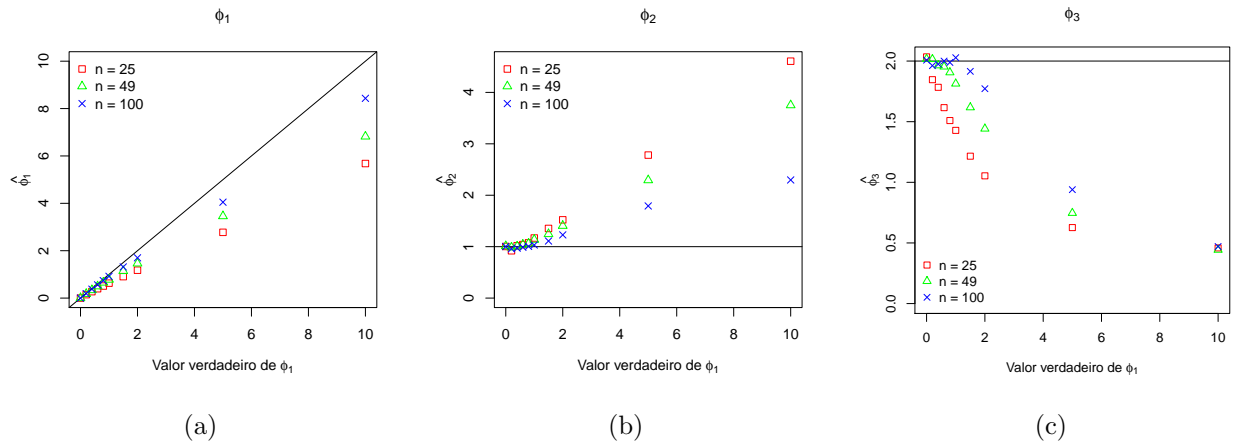


Figura 3.13: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de ϕ_1 considerando a função de covariância Gaussiana e distribuição normal. Gráfico de (a) $\hat{\phi}_1$, (b) $\hat{\phi}_2$ e (c) $\hat{\phi}_3$ em função do verdadeiro valor de ϕ_1 .

grandes quantidades de estimativas próximas de zero, mesmo para valores desses parâmetros como 5 e 10. O intervalo de confiança para ϕ_2 possui menor amplitude comparado aos intervalos de confiança para ϕ_1 e ϕ_3 , indicando que sua estimação é mais precisa. Além disso, observe que o intervalo de confiança contém o verdadeiro valor do parâmetro para os três parâmetros de escala.

Também foi analisada a estimação dos valores σ_h da matriz Σ . Os valores considerados para ϕ são os mesmo mencionados no início da seção. Pelo modelo apresentado em (1.6), notamos que a variação de ϕ_1 afeta apenas variâncias, a variação de ϕ_2 afeta variâncias e covariâncias e a variação de ϕ_3 afeta apenas covariâncias. Dessa forma, foram construídos gráficos de $\hat{\sigma}_h$ vs. σ_h para cada variação de parâmetros do vetor ϕ .

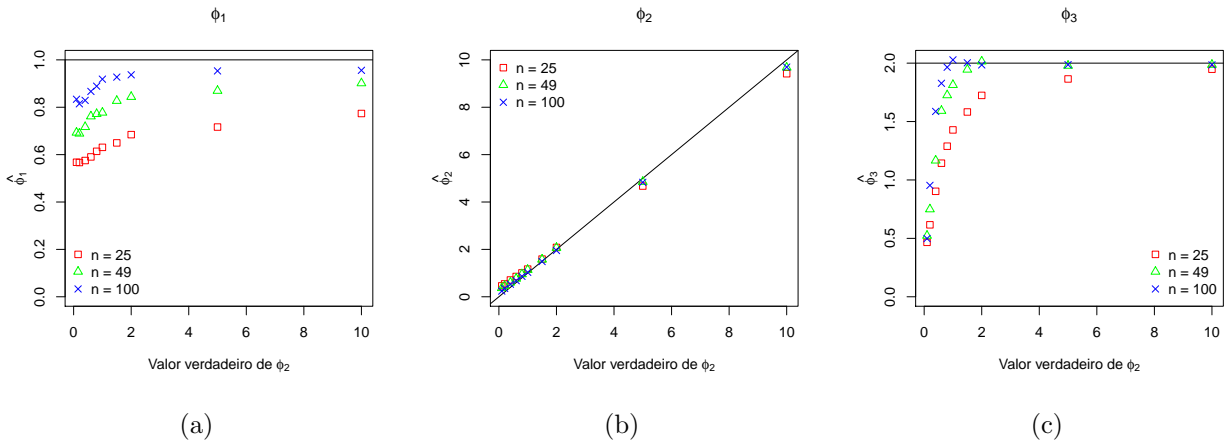


Figura 3.14: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de ϕ_2 considerando a função de covariância Gaussiana e distribuição normal. Gráfico de (a) $\hat{\phi}_1$, (b) $\hat{\phi}_2$ e (c) $\hat{\phi}_3$ em função do verdadeiro valor de ϕ_2 .

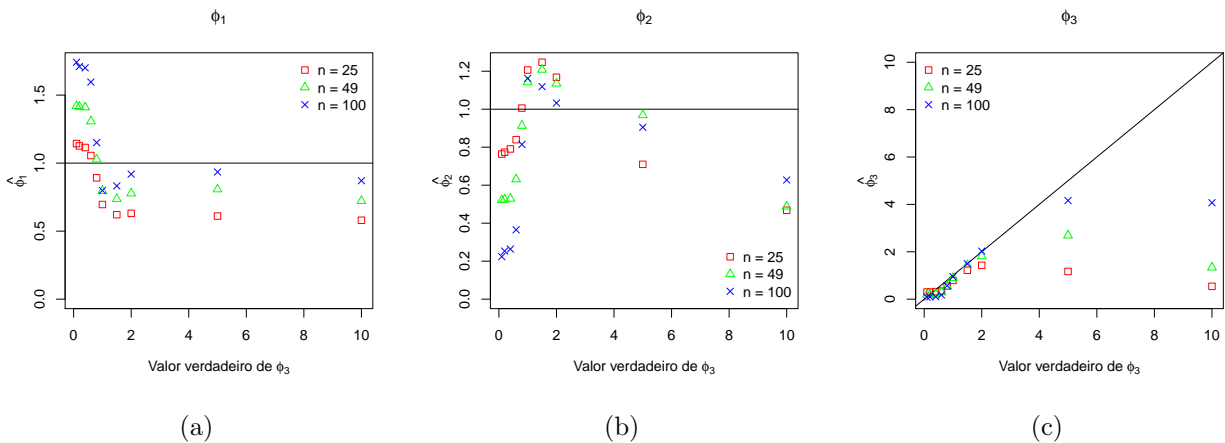
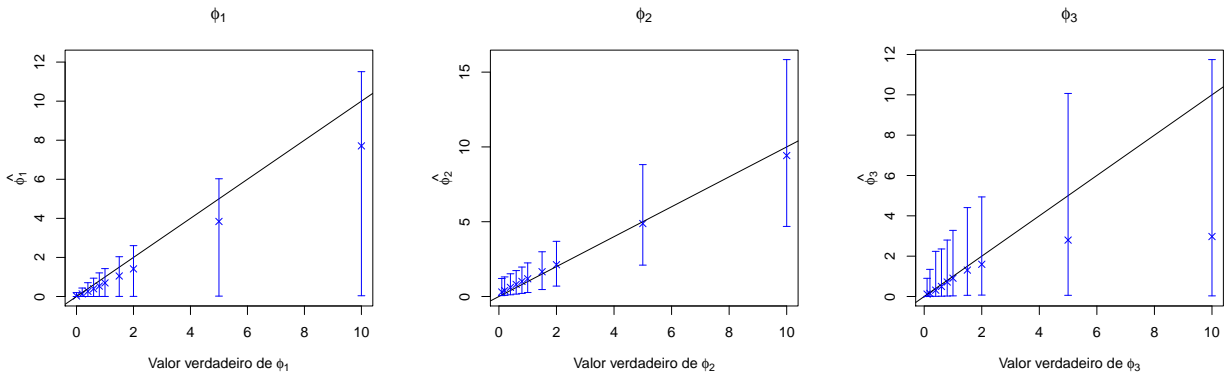


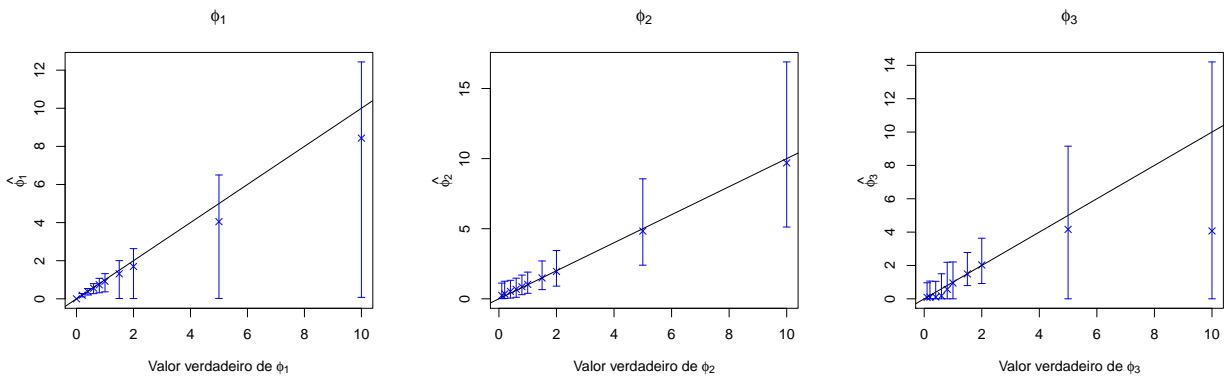
Figura 3.15: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de ϕ_3 considerando a função de covariância Gaussiana e distribuição normal. Gráfico de (a) $\hat{\phi}_1$, (b) $\hat{\phi}_2$ e (c) $\hat{\phi}_3$ em função do verdadeiro valor de ϕ_3 .

Na Figura 3.17 podemos observar que os componentes de variância são bem estimados quando variamos o valor do parâmetro ϕ_1 e mantemos os demais fixos, apresentando melhores resultados para maiores valores de tamanho de amostra. Isso indica que, apesar do viés presente nas estimativas de ϕ , os valores das estimativas dos componentes de variância são próximos dos valores verdadeiros.

Note na Figura 3.18 que os componentes de variância são bem estimados para qualquer tamanho de amostra quando variamos o valor do parâmetro ϕ_2 e mantemos os demais fixos. Para a distância $h = 1$ e função de covariância Gaussiana, os componentes de covariância são bem estimados, apresentando valores de $\hat{\sigma}_1$ próximos do valor verdadeiro σ_1 . Para a mesma distância e função de covariância exponencial, notamos um viés crescente chegando a quase 50% quando



(a)



(b)

Figura 3.16: Intervalo de 95% de confiança empírico para $\phi = (\phi_1, \phi_2, \phi_3)'$, considerando a função de covariância (a) exponencial e (b) Gaussiana, distribuição normal e tamanho de amostra 100.

$\sigma_1 = 6$ para o tamanho de amostra 25. Esse viés é reduzido com o aumento do tamanho de amostra. Para $h = 3$, notamos que as estimativas são próximas do valor zero. Vale notar que os gráficos dos componentes de covariância considerando a função de covariância exponencial e Gaussiana estão em escalas diferentes. Isso indica que, apesar do viés presente nas estimativas de ϕ , os valores das estimativas dos componentes de variância e covariância são próximos dos valores verdadeiros para a função de covariância Gaussiana. Percebemos que o viés não é corrigido quando calculamos os componentes da matriz escala para a função de covariância exponencial, sendo corrigido apenas com o aumento do tamanho de amostra.

Observamos na Figura 3.19 que, quando variamos o valor do parâmetro ϕ_3 e mantemos os demais fixos, $\hat{\sigma}_h$ possui comportamento semelhante ao de $\hat{\phi}_3$ (ver Figura 3.11). Novamente o viés não é corrigido quando obtemos $\hat{\phi}_3$ mas sim com o aumento do tamanho de amostra.

A análise do desempenho da estimação do vetor de parâmetros $\phi = (\phi_1, \phi_2, \phi_3)$ considerando as distribuições t-Student e slash segue um comportamento semelhante apresentado para a

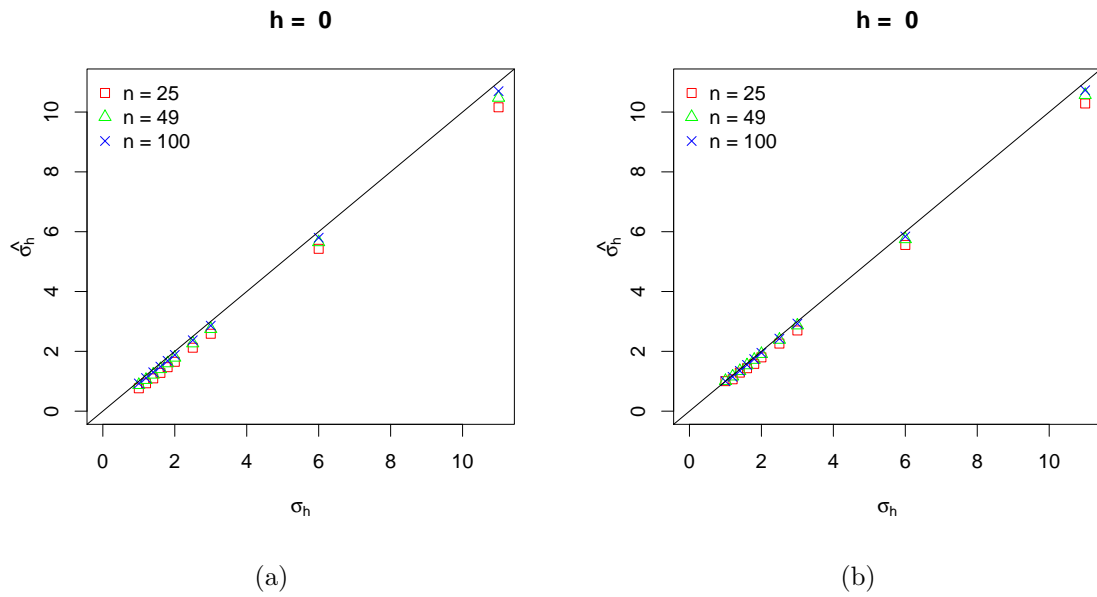


Figura 3.17: Gráfico de componentes de variância ($h = 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_1 considerando distribuição normal e as funções de covariância (a) exponencial e (b) Gaussiana.

distribuição normal. As Figuras 3.20 a 3.24 apresentam os resultados de simulação para a distribuição t-Student com 5 g.l. e as funções de covariância exponencial e Gaussiana. Notamos que os parâmetros ϕ_1 e ϕ_3 são superestimados, sendo ϕ_2 superestimado para todos os tamanhos de amostra. As estimativas de ϕ_3 seguem comportamento similar ao caso normal, sendo mais próximas ao valor verdadeiro para tamanho de amostra 100. Em relação aos componentes de variância e covariância, observamos que são superestimados para todos os tamanhos de amostra, exceto quando variamos ϕ_3 e fixamos os demais parâmetros cujo comportamento é similar ao caso normal.

As Figuras 3.25 a 3.29 apresentam os resultados de simulação para a distribuição slash com $\nu = 5$ e as funções de covariância exponencial e Gaussiana. Observamos que os parâmetros ϕ_1 e ϕ_3 são subestimados, mas ϕ_2 é bem estimado para tamanhos de amostra 49 e 100. As estimativas de ϕ_3 seguem comportamento similar ao caso normal. Note que as estimativas dos componentes de variância e covariância não apresentam grandes valores de viés, apenas quando variamos ϕ_3 e fixamos os demais parâmetros, comportamento semelhante ao observado no caso normal.

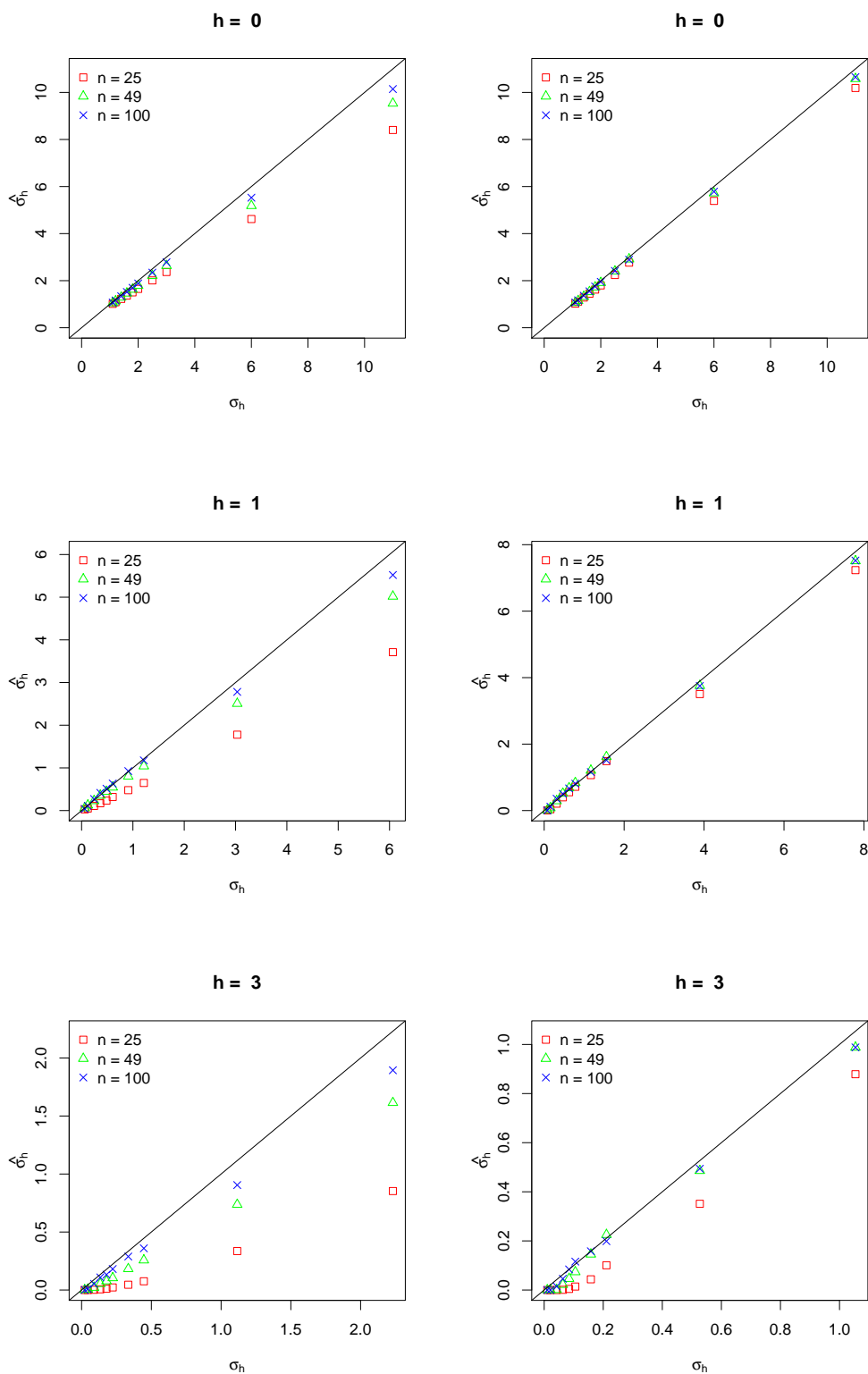


Figura 3.18: Gráfico de componentes de variância ($h = 0$) e covariância ($h > 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_2 considerando distribuição normal e as funções de covariância exponencial (coluna esquerda) e Gaussiana (coluna direita).

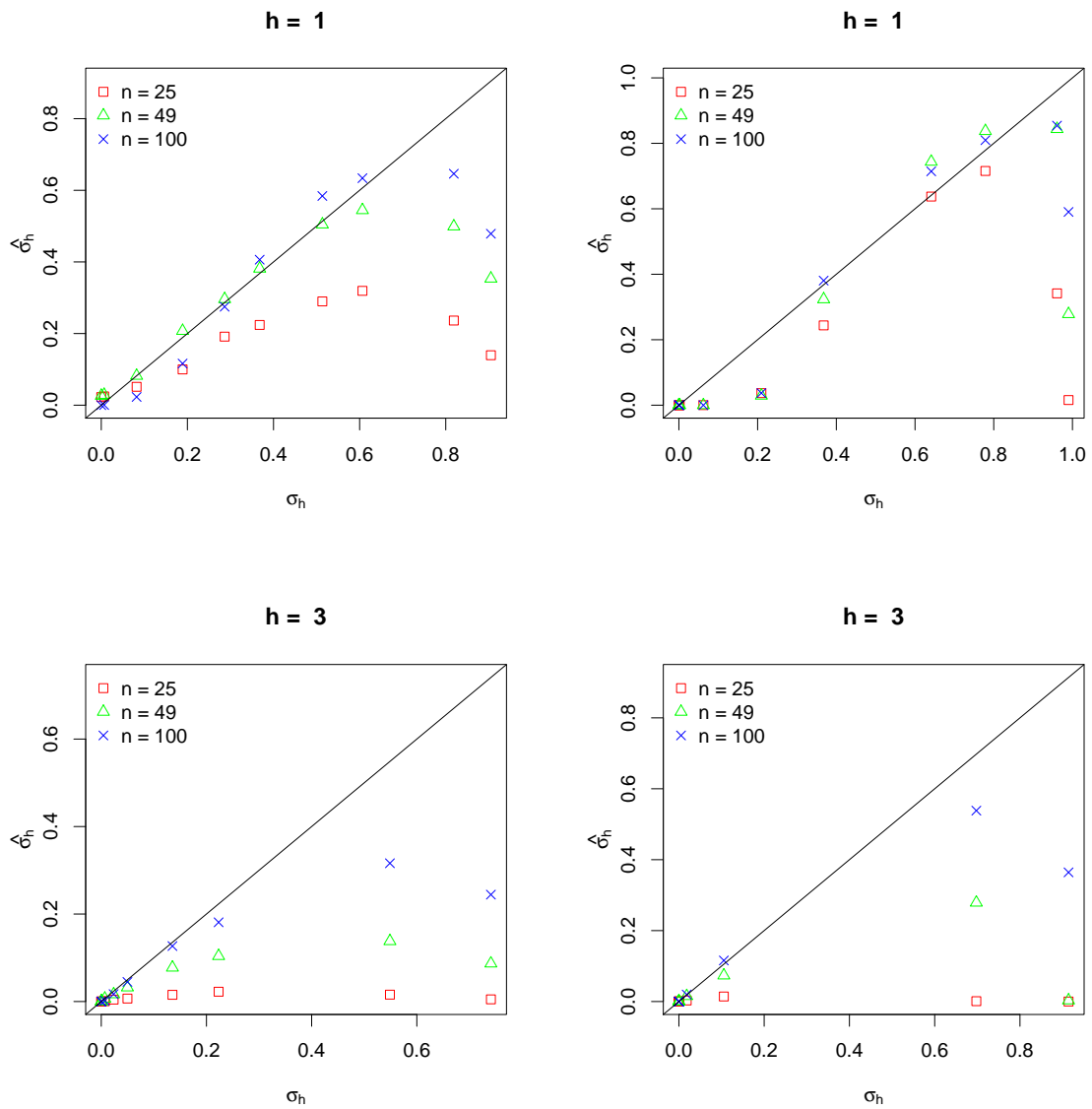


Figura 3.19: Gráfico de componentes de covariância ($h > 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_3 considerando distribuição normal e as funções de covariância exponencial (coluna esquerda) e Gaussiana (coluna direita).

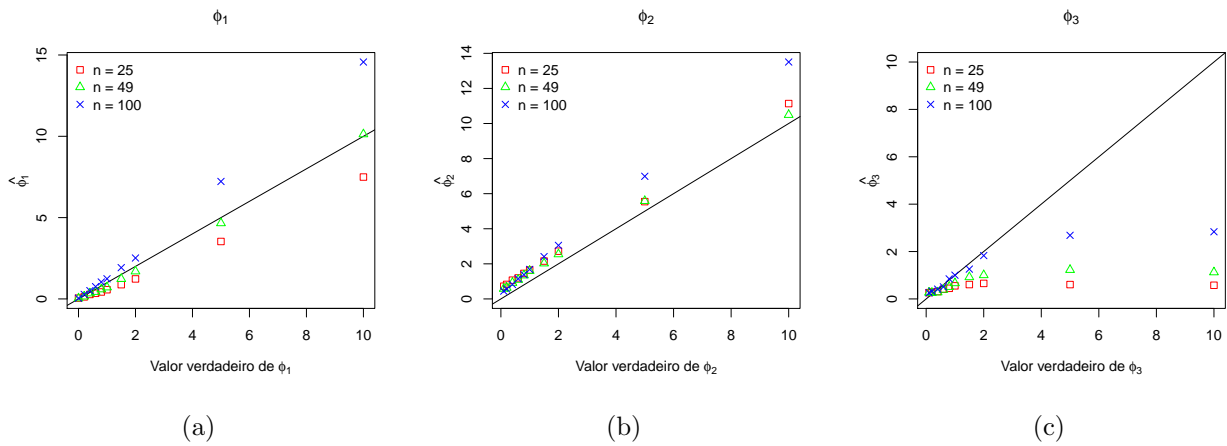


Figura 3.20: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância exponencial e distribuição t-Student com 5 g.l.

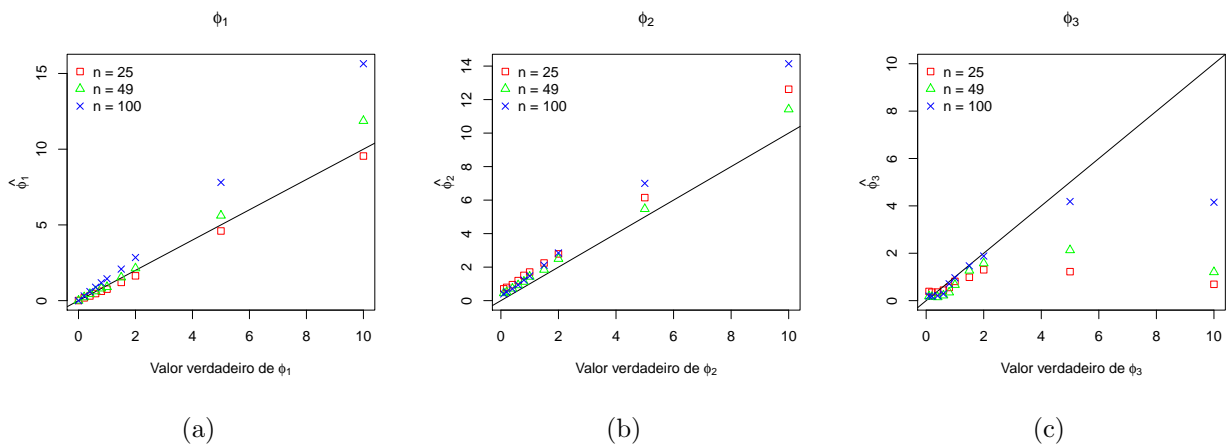


Figura 3.21: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância Gaussiana e distribuição t-Student com 5 g.l.

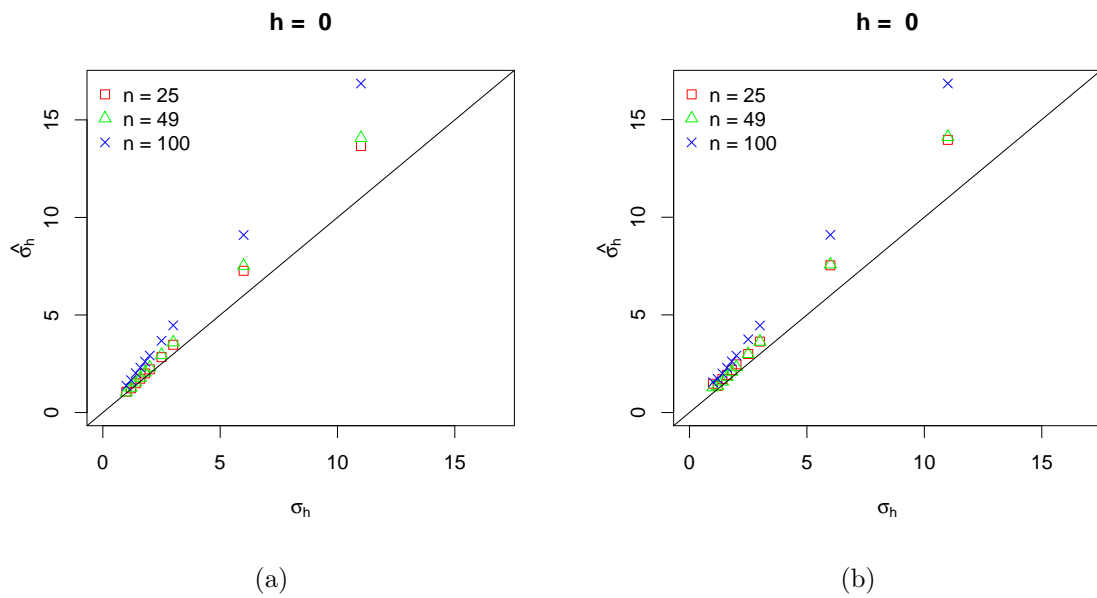


Figura 3.22: Gráfico de componentes de variância ($h = 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_1 , distribuição t-Student com 5 g.l. e considerando as funções de covariância (a) exponencial e (b) Gaussiana.

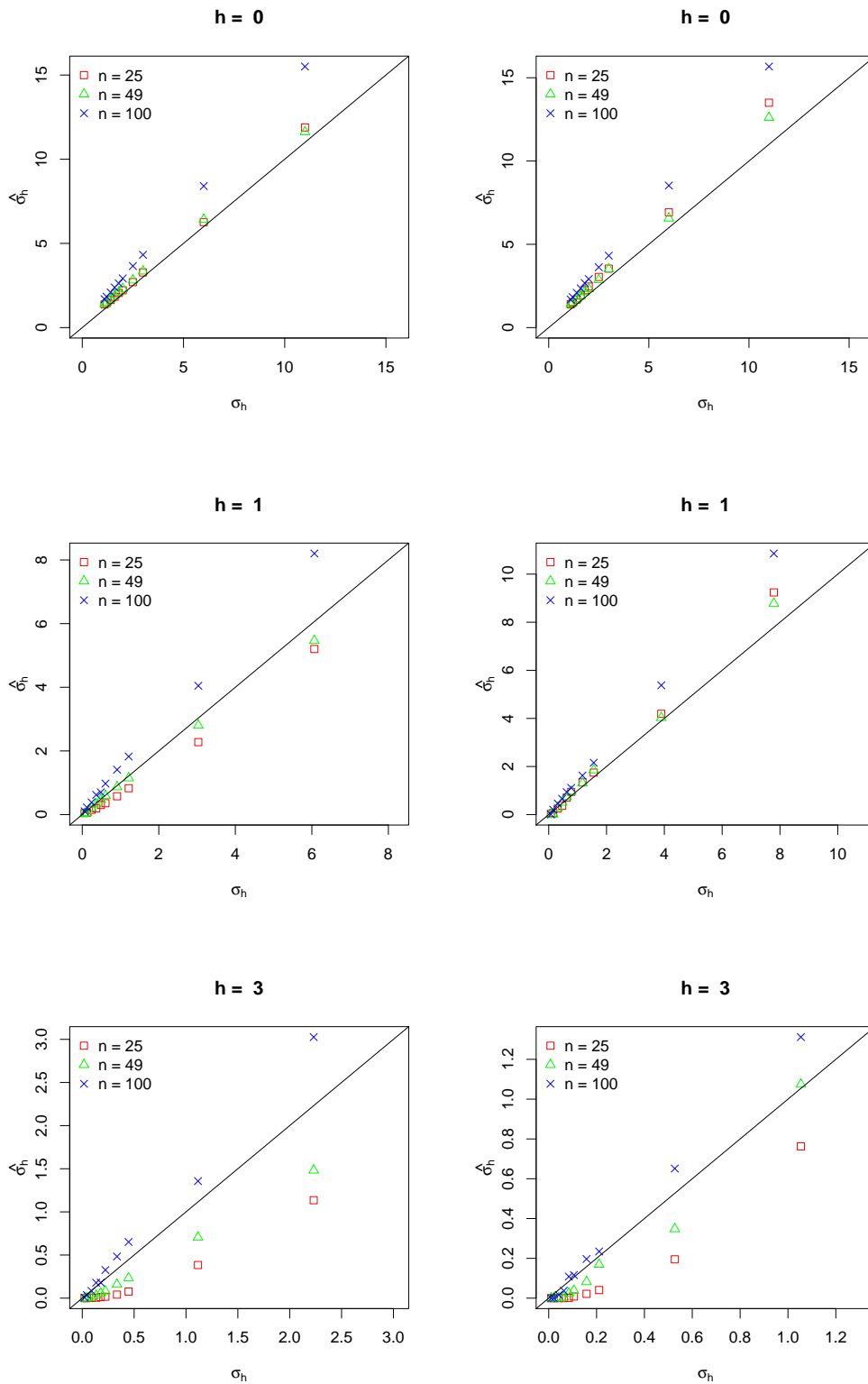


Figura 3.23: Gráfico de componentes de variância ($h = 0$) e covariância ($h > 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_2 , distribuição t-Student com 5 g.l. e considerando as funções de covariância exponencial (coluna esquerda) e Gaussiana (coluna direita).

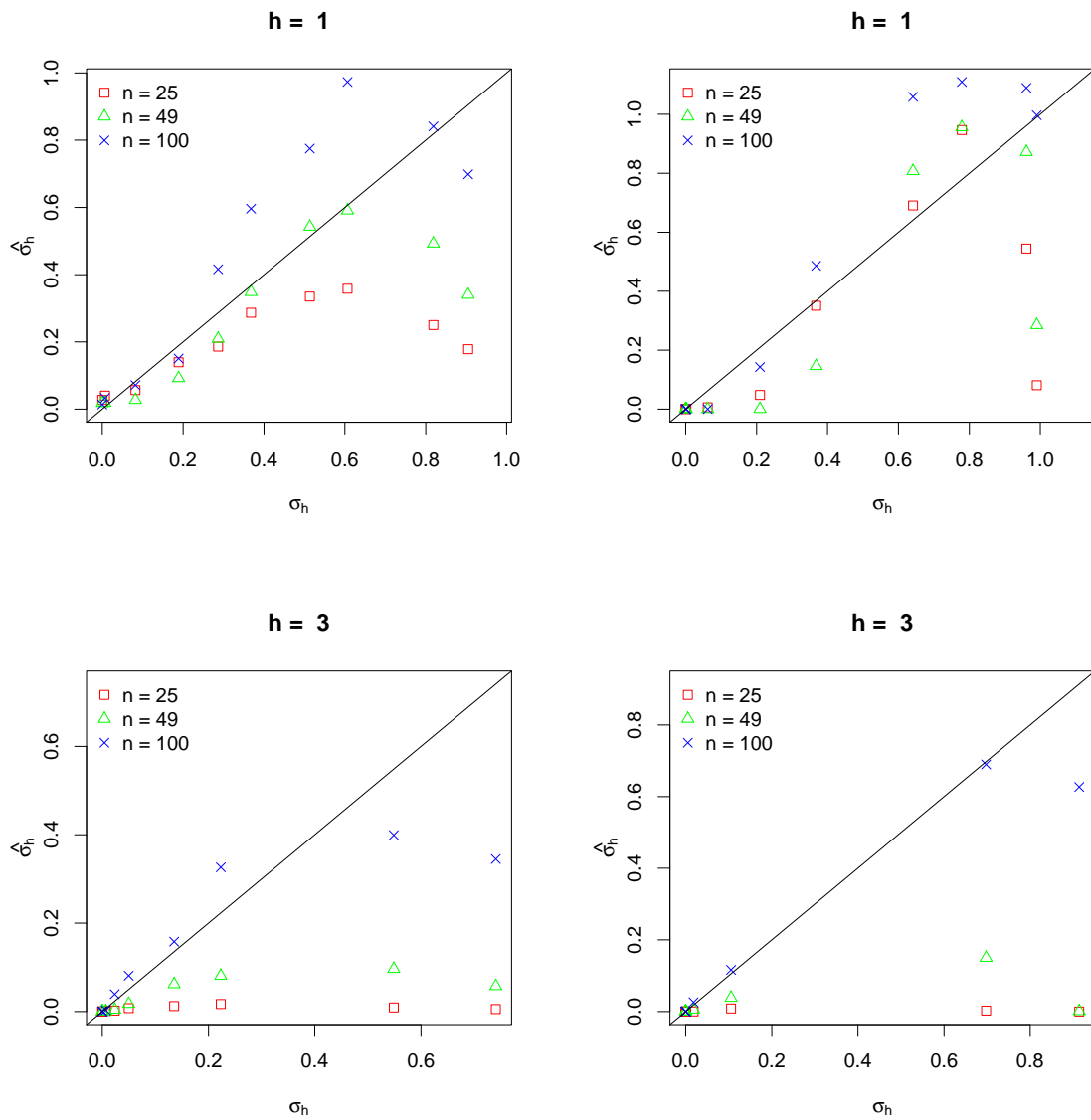


Figura 3.24: Gráfico de componentes de covariância ($h > 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_3 , distribuição t-Student com 5 g.l. e considerando as funções de covariância exponencial (coluna esquerda) e Gaussiana (coluna direita).

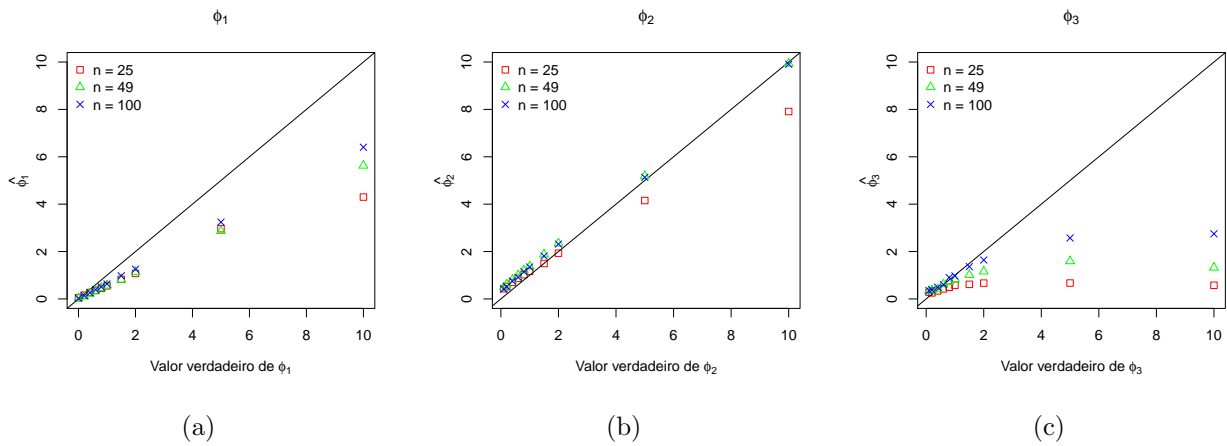


Figura 3.25: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância exponencial e distribuição slash com $\nu = 5$.

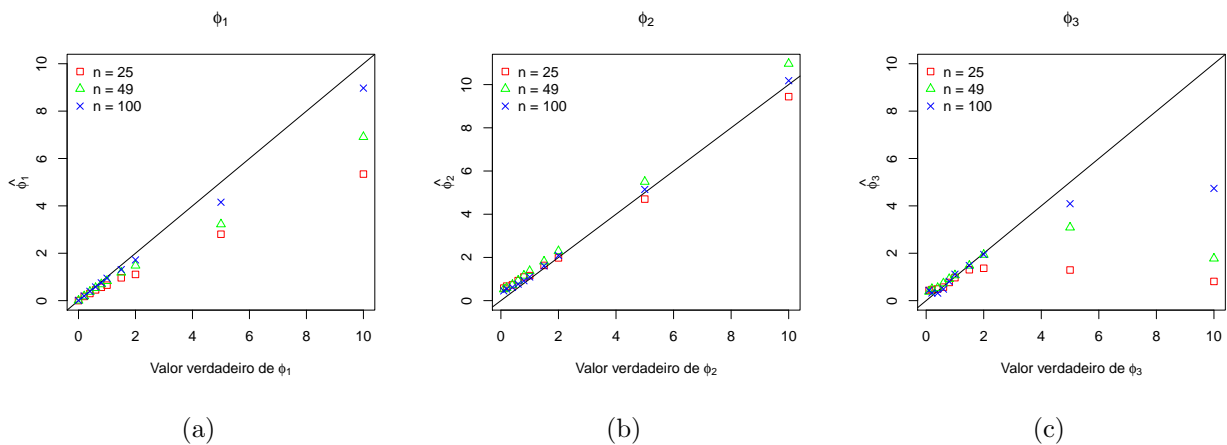


Figura 3.26: Análise do desempenho da estimação dos parâmetros $\phi = (\phi_1, \phi_2, \phi_3)'$ para alguns valores de (a) ϕ_1 , (b) ϕ_2 e (c) ϕ_3 considerando a função de covariância Gaussiana e distribuição slash com $\nu = 5$.

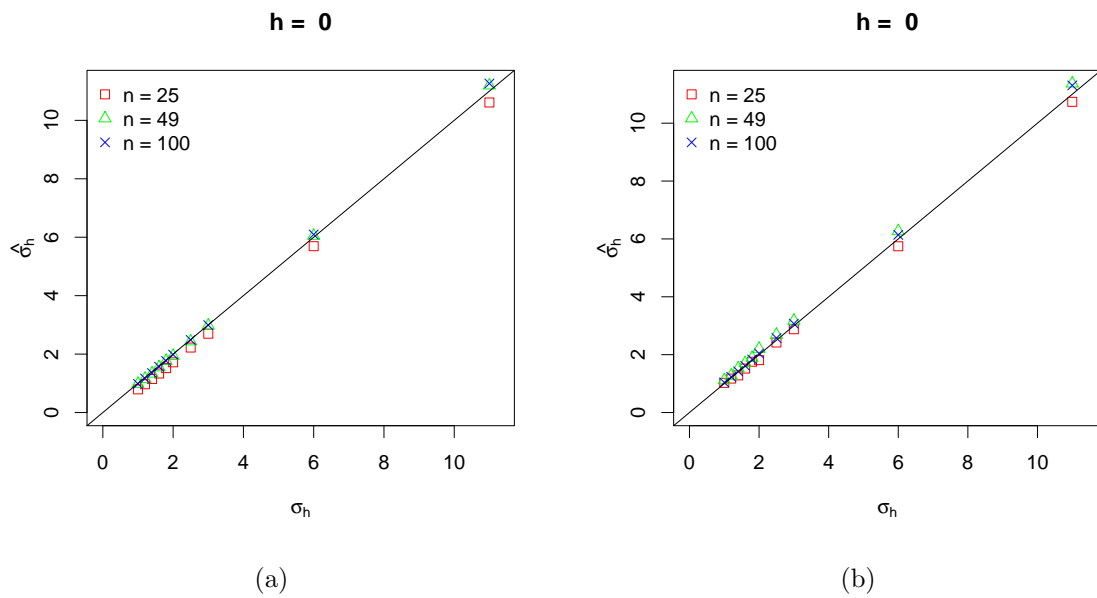


Figura 3.27: Gráfico de componentes de variância ($h = 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_1 , distribuição slash com $\nu = 5$ e considerando as funções de covariância (a) exponencial e (b) Gaussiana.

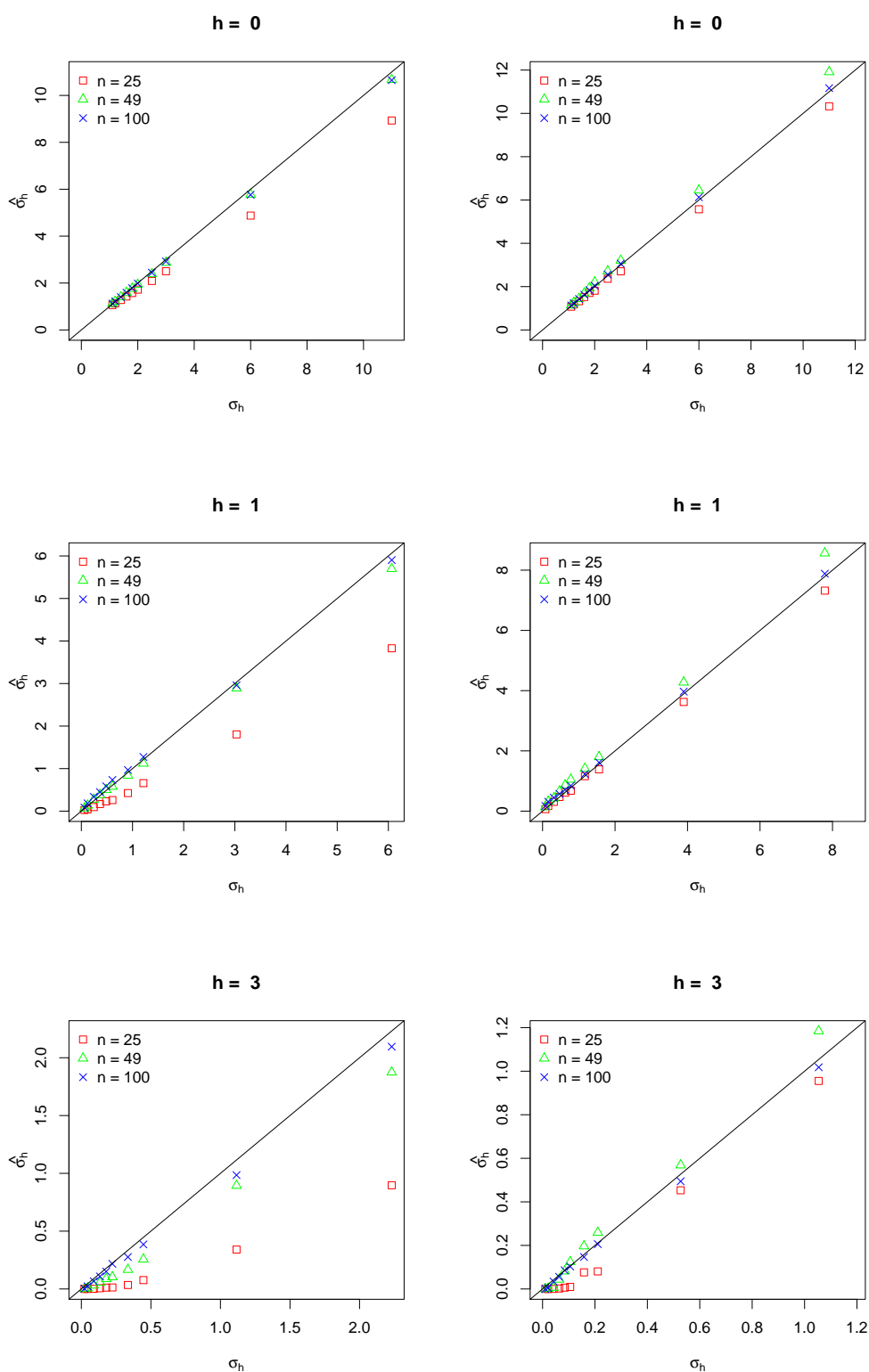


Figura 3.28: Gráfico de componentes de variância ($h = 0$) e covariância ($h > 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_2 , distribuição slash com $\nu = 5$ e considerando as funções de covariância exponencial (coluna esquerda) e Gaussiana (coluna direita).

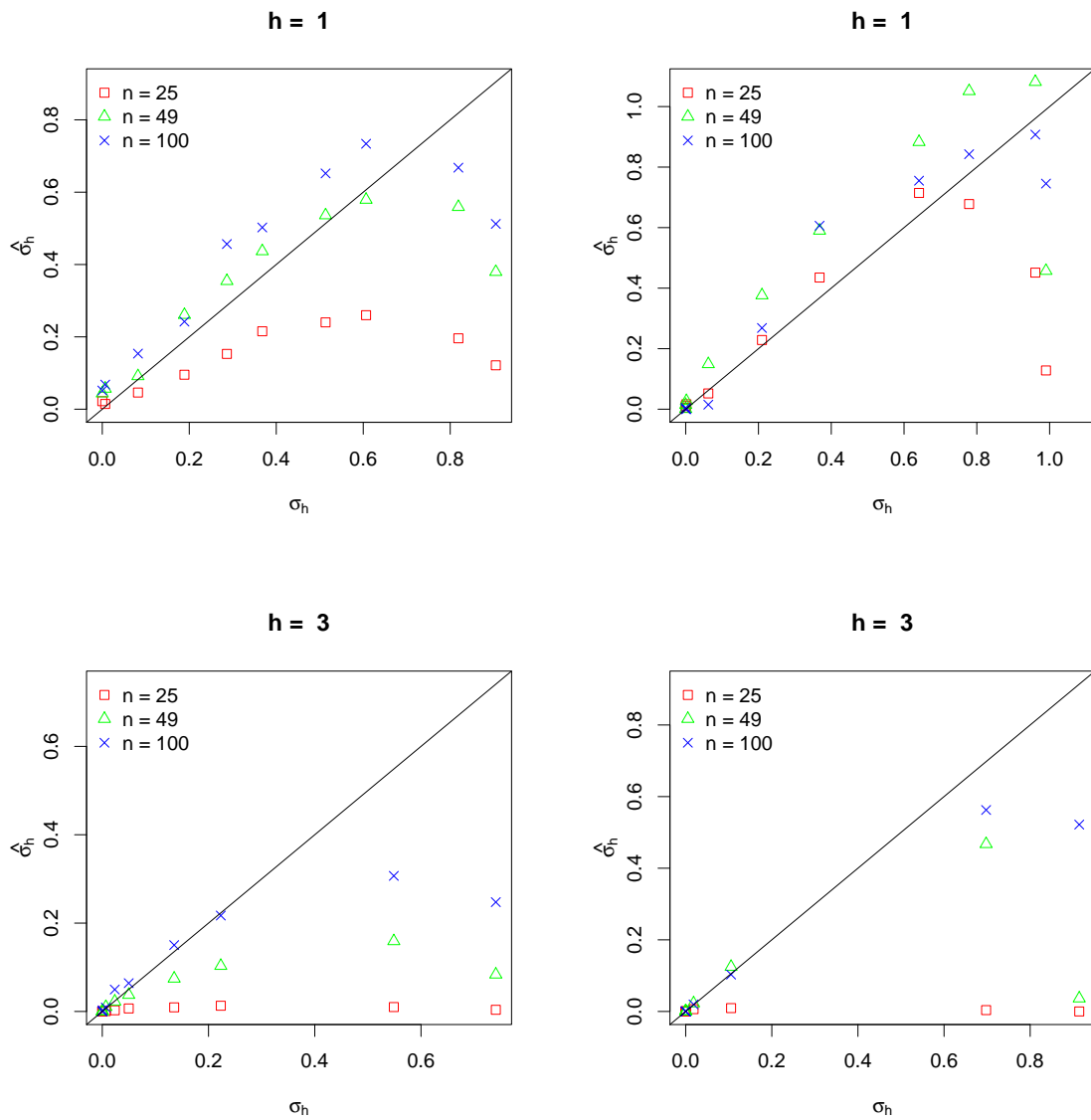


Figura 3.29: Gráfico de componentes de covariância ($h > 0$) estimados ($\hat{\sigma}_h$) em função do verdadeiro valor (σ_h) para alguns valores de ϕ_3 , distribuição slash com $\nu = 5$ e considerando as funções de covariância exponencial (coluna esquerda) e Gaussiana (coluna direita).

3.3 Predição

Para desenvolver o estudo de simulação para análise de predição, foi gerada uma amostra de tamanho 100 a partir de uma das distribuições em estudo e função de covariância exponencial. Utilizamos as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$ e malha regular (ver Figura 3.1). Foi utilizado o modelo de média constante com $\mu = 2$ e $\phi = (1, 1, 2)$.

A predição foi realizada de duas formas: (1) considerando os valores verdadeiros de θ e (2) considerando os valores estimados $\hat{\theta}$. Dessa forma, observamos o comportamento da predição utilizando diferentes distribuições. Para realizar a predição considerando a distribuição slash, precisamos recorrer a simulação Monte Carlo (vide Seção (2.8.3)) e foram realizadas 500 simulações para a distribuição preditiva.

Foi gerada uma amostra a partir da distribuição normal e função de covariância exponencial. Observamos o gráfico de boxplot do desvio padrão de predição considerando a predição realizada das formas (1) e (2) nas Figuras 3.30 (a) e (b), respectivamente. O boxplot em vermelho representa o resultado considerando a distribuição que originou os dados, no caso a distribuição normal. Considerando o verdadeiro valor de θ , observamos que o desvio padrão de predição é ligeiramente maior quando consideramos a distribuição t-Student em comparação com a distribuição normal, decrescendo devagar com o aumento dos g.l., e maior quando consideramos a distribuição slash. A maior variabilidade no boxplot de desvio padrão de predição para a distribuição slash é explicada pelo fato de utilizarmos simulações Monte Carlo, ao contrário das distribuições normal e t-Student que conhecemos a distribuição preditiva exata. Quando

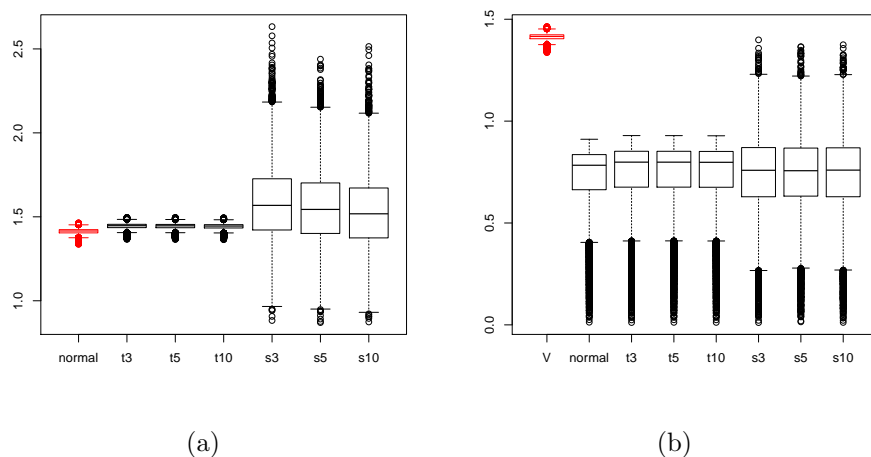


Figura 3.30: Gráficos de boxplot do desvio padrão de predição para as distribuições em estudo e função de covariância exponencial considerando (a) os parâmetros verdadeiros e (b) os parâmetros estimados e amostra gerada da distribuição normal e função de covariância exponencial.

consideramos os parâmetros estimados para realizar a predição, notamos um comportamento similar para todas as distribuições, sendo que a distribuição t-Student apresentou desvio padrão mediano maior que as distribuições normal e slash, mas todas subestimaram o desvio padrão de predição. Esperávamos que o comportamento do desvio padrão de predição observado no modelo considerando o vetor de parâmetros verdadeiros θ fosse refletido quando os modelos foram ajustados e os parâmetros estimados, o que não foi observado. Esse é um tópico que deve ser analisado mais profundamente.

Para amostras geradas das distribuições t-Student e slash, ambas com $\nu = 5$, o comportamento foi semelhante ao apresentado para o caso da amostra gerada pela distribuição normal.

3.4 Influência local

Para desenvolver o estudo de simulação para análise de influência local, foi gerada uma amostra de tamanho 100 em malha regular (Figura 3.1) a partir da distribuição normal com parâmetros $\mu = 2$ e $\phi = (1, 1, 2)'$. Utilizamos para o ajuste do modelo as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$. Para a amostra \mathbf{Y} gerada, o valor máximo foi perturbado sob o fator $b = 0,5\sqrt{\mathbf{Y}'\mathbf{Y}}$, sendo utilizada a perturbação aditiva. Desse modo, $\mathbf{Y}_\omega = \mathbf{Y} + b\omega$ em que ω é o vetor com 1 na posição correspondente a observação de valor máximo e 0 nas demais. A análise de diagnóstico foi realizada com base na metodologia de Cook (1986), discutida na Seção 1.4. A seguir são apresentados os resultados de simulação considerando função de covariância exponencial e Gaussiana.

3.4.1 Função de covariância exponencial

Geramos uma amostra de tamanho 100 considerando a distribuição normal e função de covariância exponencial. A Figura 3.31 (a) apresenta o gráfico de boxplot para a amostra original e a amostra após a perturbação feita na observação 14 (valor máximo). A Figura 3.31 (b) apresenta o gráfico de postplot da amostra gerada sem perturbação. Foram ajustados modelos de média constante considerando as distribuições normal, t-Student com 1 g.l. e slash com $\nu = 1$. Os gráficos de influência local são observados na Figura 3.32, sendo a curvatura normal C_i na primeira linha de gráficos e as componentes do autovetor associado ao maior autovalor, $|\mathbf{h}_{max}|$, na segunda linha. Note que, para o modelo ajustado com distribuição normal, a observação 14 é detectada como a mais influente pelos gráficos de diagnóstico e as observações 4, 13, 15 e 24, que são vizinhas à observação 14, também são identificadas no gráfico C_i mas com menor influência. Essas observações deixam de ser influentes para os modelos ajustados com

distribuição t-Student e slash. A Tabela 3.4 apresenta os valores de $\hat{\theta}$ para o conjunto de dados sem e com perturbação. Observamos que os valores de $\hat{\phi}_1$ e $\hat{\phi}_2$ aumentaram após a perturbação e o valor de $\hat{\phi}_3$ diminuiu. Porém, a variação foi menor para o modelo ajustado com distribuição slash, indicando a robustez dessa distribuição. Apesar de não haver grande diferença entre as variações dos parâmetros estimados pela distribuição normal e t-Student após a perturbação, notamos que essa última é mais robusta pelos gráficos de influência.

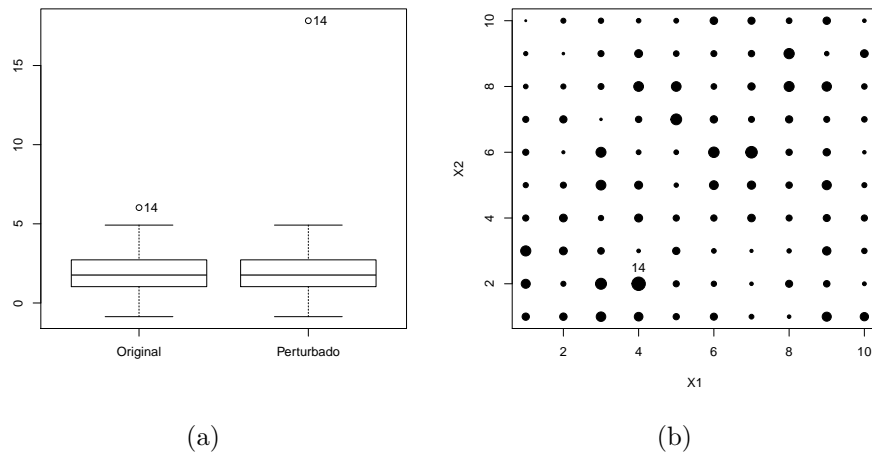
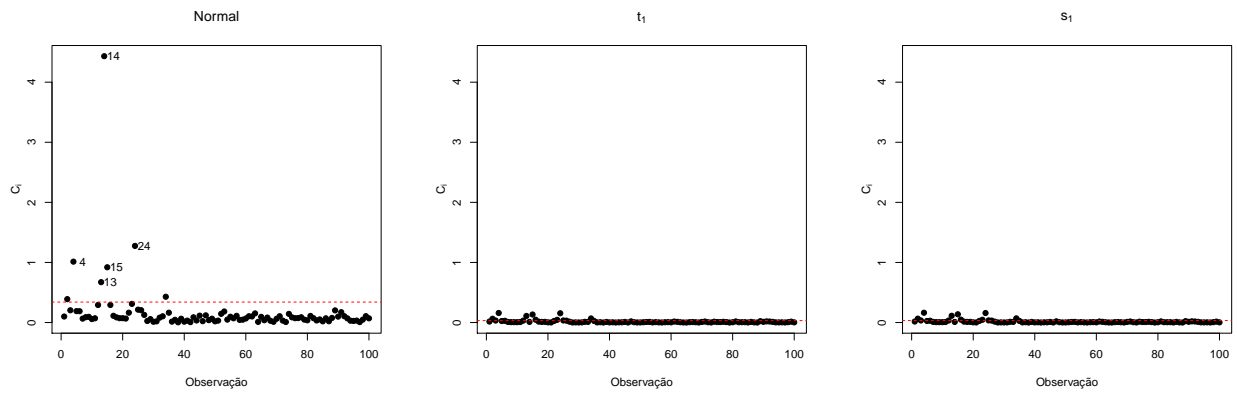


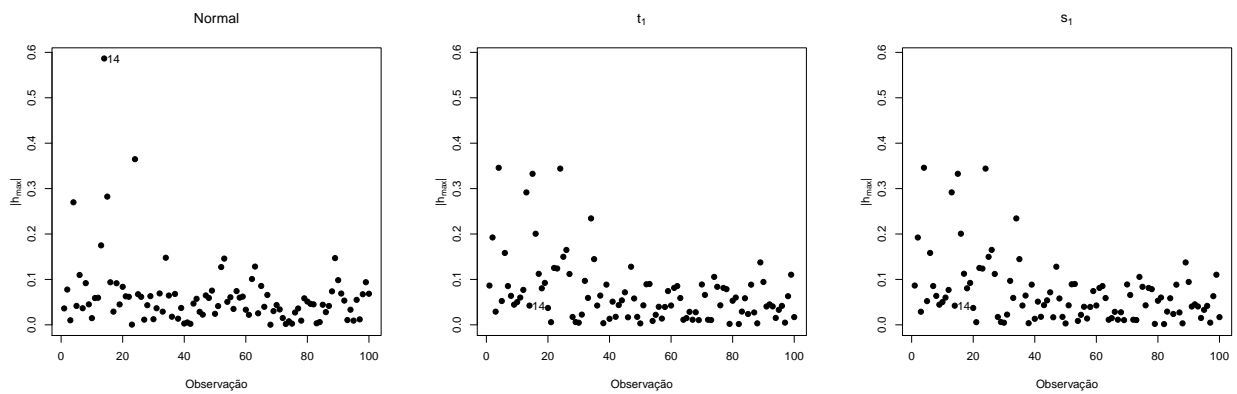
Figura 3.31: Gráficos de (a) boxplot e (b) postplot do conjunto de dados simulado considerando a distribuição normal e função de covariância exponencial.

Tabela 3.4: Estimativas de $\theta = (\mu, \phi_1, \phi_2, \phi_3)$ e seus erros padrão dos modelos sem e com perturbação considerando a distribuição normal, t-Student com 5 g.l., slash com $\nu = 5$ e função de covariância exponencial.

Modelo		$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
Normal	Sem perturbação	1,883 (0,246)	0,931 (0,670)	0,798 (0,692)	1,069 (0,844)
	Com perturbação	2,043 (0,322)	2,256 (2,681)	1,793 (2,738)	0,625 (0,533)
t_1	Sem perturbação	1,883 (0,246)	0,933 (1,097)	0,796 (1,041)	1,072 (1,330)
	Com perturbação	2,043 (0,263)	2,239 (5,014)	1,809 (4,947)	0,622 (0,979)
s_1	Sem perturbação	1,883 (0,246)	0,748 (0,828)	0,638 (0,814)	1,072 (1,331)
	Com perturbação	2,043 (0,263)	1,795 (4,287)	1,449 (4,288)	0,622 (0,980)



(a)



(b)

Figura 3.32: Gráficos de (a) C_i , (b) $|h_{max}|$ do conjunto de dados perturbado considerando a função de covariância exponencial e perturbação aditiva.

3.4.2 Função de covariância Gaussiana

Geramos uma amostra de tamanho 100 considerando a distribuição normal e função de covariância Gaussiana. A Figura 3.33 (a) apresenta o gráfico de boxplot para a amostra original e a amostra após a perturbação feita na observação 93 (valor máximo). A Figura 3.33 (b) apresenta o gráfico de postplot da amostra original. Foram ajustados modelos de média constante considerando as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$.

Os gráficos de influência local são observados na Figura 3.34. Note que a observação 93 é detectada como a mais influente pelos gráficos de diagnóstico para o modelo ajustado com distribuição normal. Essa observação deixa de ser influente para os modelos ajustados com distribuição t-Student e slash. As observações 83, 91, 95, 92 e 94 são vizinhas à observação 14 e são identificadas no gráfico de $|\mathbf{h}_{max}|$ para o modelo ajustado com distribuição normal, sendo que as duas últimas são identificadas também para os modelos com distribuição t-Student e slash.

A Tabela 3.5 apresenta os valores de $\hat{\boldsymbol{\theta}}$ para o conjunto de dados sem e com perturbação. Observamos que os valores de $\hat{\phi}$ aumentaram após a perturbação para todos os modelos ajustados. A variação foi menor para o modelo ajustado com distribuição slash. Apesar de não haver grande diferença entre as variações dos parâmetros estimados pela distribuição normal e t-Student após a perturbação, notamos que essa última é mais robusta pelos gráficos de influência.

Tabela 3.5: Estimativas de $\boldsymbol{\theta} = (\mu, \phi_1, \phi_2, \phi_3)$ dos modelos sem e com perturbação considerando a distribuição normal, t-Student com 5 g.l., slash com $\nu = 5$ e função de covariância Gaussiana.

Modelo		$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
Normal	Sem perturbação	1,938 (0,286)	1,053 (0,173)	0,971 (0,312)	1,889 (0,494)
	Com perturbação	2,125 (0,496)	2,802 (0,622)	1,874 (1,364)	2,154 (0,700)
t_1	Sem perturbação	1,938 (0,301)	1,053 (0,696)	0,971 (0,726)	1,890 (0,492)
	Com perturbação	2,125 (0,469)	2,797 (1,841)	1,870 (1,531)	2,153 (0,643)
s_1	Sem perturbação	1,938 (0,290)	0,844 (0,800)	0,778 (0,843)	1,890 (0,512)
	Com perturbação	2,125 (0,438)	2,245 (1,288)	1,502 (1,401)	2,154 (0,755)

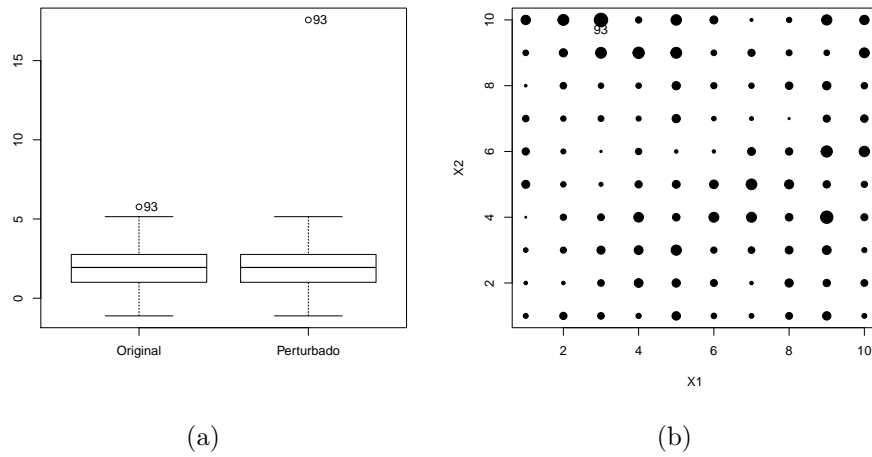


Figura 3.33: Gráficos de (a) boxplot e (b) postplot do conjunto de dados simulado considerando a distribuição normal e função de covariância Gaussiana.

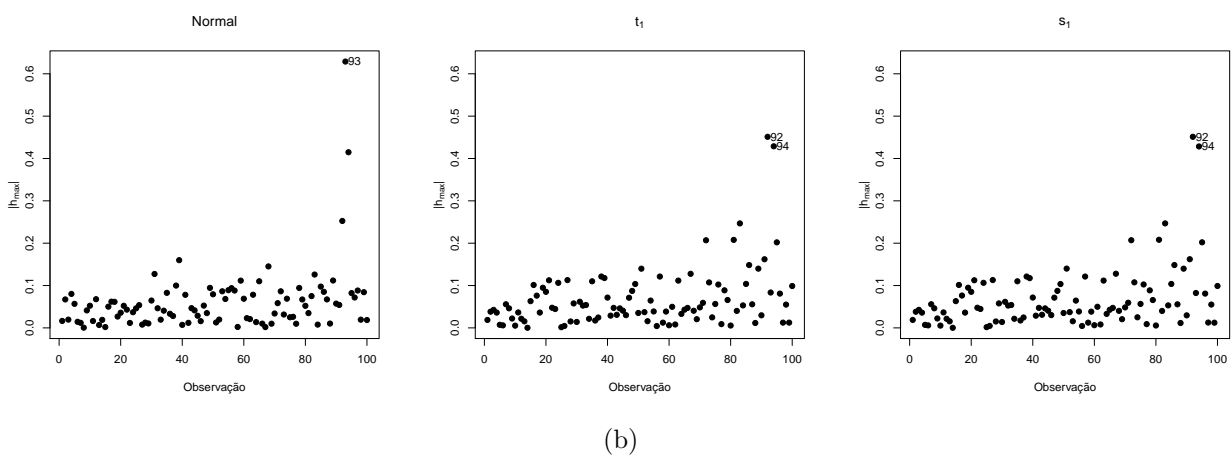
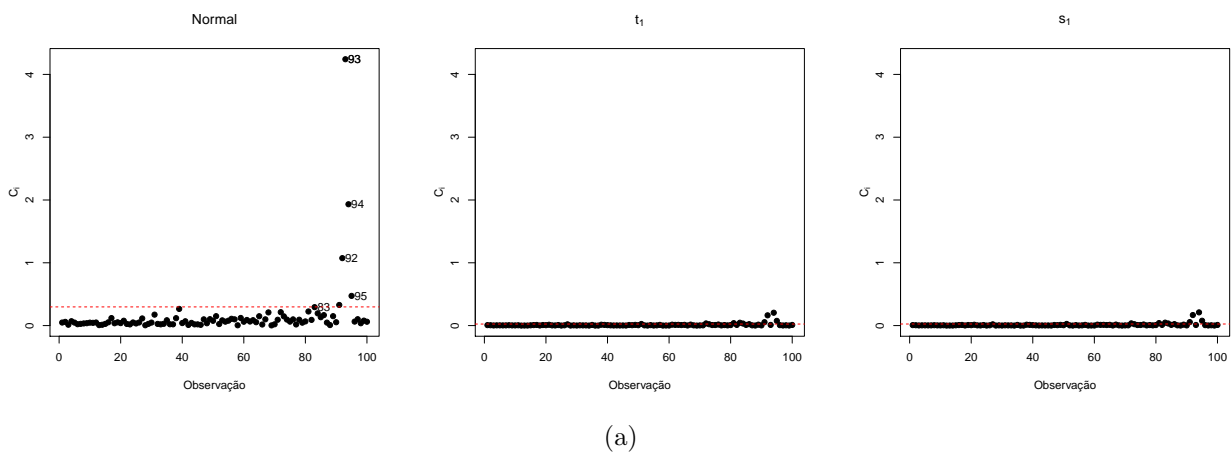


Figura 3.34: Gráficos de (a) C_i , (b) $|h_{max}|$ e (c) $|h_{max}|_2$ do conjunto de dados perturbado considerando a função de covariância Gaussiana e perturbação aditiva.

Capítulo 4

Aplicação a dados experimentais

Este capítulo é destinado a aplicação da metodologia discutida em um conjunto de dados real. A aplicação é realizada em um conjunto de dados de precipitação pluviométrica, apresentada na Seção 4.1.

4.1 Dados de precipitação pluviométrica

Em casos de eventos graves, como vazamento de poluente radioativo, é essencial que se tenha planos de urgência para combater ou restringir tal evento. Deve-se identificar as áreas afetadas e, se necessário, evacuar a população exposta até que a área seja limpa. A tomada de decisões deve ser feita de forma rápida e a estimativa da contaminação em locais não observados possui papel importante nos casos de emergências ambientais. Com essa motivação, o grupo de Monitoramento Ambiental Radioativo (Radioactive Environmental Monitoring - REM) do Instituto do Meio Ambiente de Joint Research Centre em Ispra, Itália, organizou um exercício de interpolação espacial na World Wide Web em setembro de 1997. A ideia foi fornecer uma parte de um banco de dados para que os participantes realizassem a predição para locais não observados. Os dados se referem ao caso do acidente da usina nuclear de Chernobyl, ocorrido em abril de 1986 no norte da Ucrânia. A precipitação pluviométrica foi um parâmetro que definiu o depósito radioativo no caso do desastre e, então, foram fornecidos os dados de precipitação diária de 100 locais da Suíça, medidas em 8 de maio de 1986 na unidade de 10^{-1} mm, a qual denotamos u.m. (unidade de medida). Para maiores informações, ver Dubois (2003). Neste trabalho, vamos utilizar esse conjunto de dados para aplicar a metodologia desenvolvida nas seções anteriores.

A Tabela 4.1 apresenta as estatísticas descritivas para os dados de precipitação. O valor mínimo é de 10 u.m. e o valor máximo é de 585 u.m. Notamos grande distância entre o valor do quartil 3 e o valor do máximo quando comparados a distância de valores entre o mínimo e o quartil 1, por exemplo. Observamos um coeficiente de variação (CV) de 64,7%, indicando comportamento heterogêneo dos dados de precipitação.

A Figura 4.1 (a) apresenta o gráfico de boxplot dos dados de precipitação e notamos que a observação 14 é detectada como possível observação atípica. A Figura 4.1 (b) apresenta o gráfico postplot dos dados e note que o valor da observação 14 é maior em relação aos dados

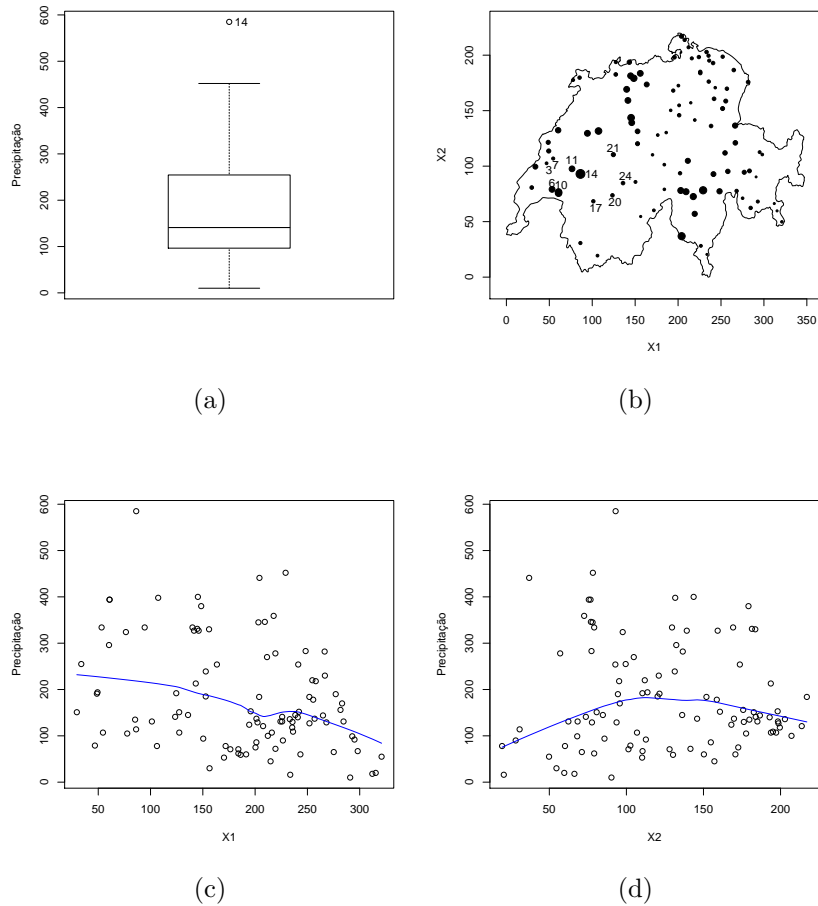


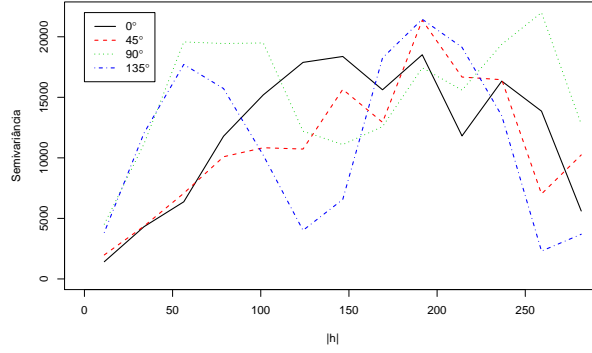
Figura 4.1: Gráficos de (a) boxplot, (b) postplot, (c) coordenada X1 (leste-oeste) e (d) coordenada X2 (norte-sul) para os dados de precipitação.

vizinhos, principalmente aos da direção leste 17, 20 e 24. Podemos observar o comportamento da precipitação em relação a cada uma das coordenadas espaciais pelas Figuras 4.1 (c) e (d). A linha azul representa a curva suavizada para os dados. Notamos uma relação decrescente fraca entre os dados de precipitação e a coordenada X1 (leste-oeste), apresentando maiores valores na região oeste. Em relação a coordenada X2 (norte-sul), não observamos uma possível relação com os dados.

A fim de verificar isotropia, isto é, se a semivariância entre duas localizações depende apenas da distância entre elas e não da direção, construímos o gráfico do semivariograma em função das direções principais que pode ser observado na Figura 4.2. Note que, em todas as quatro direções, o semivariograma indica que há dependência espacial e apresenta comportamento semelhante,

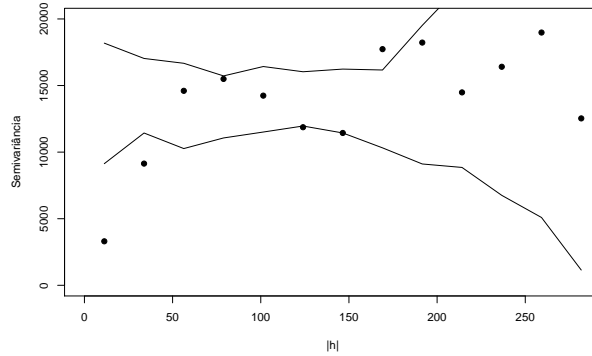
Tabela 4.1: Estatísticas descritivas para os dados de precipitação.

n	Min	Quartil 1	Mediana	Média	Quartil 3	Máximo	D.p.	CV
100	10,00	97,95	141,00	180,20	254,20	585,00	116,68	0,647



(a)

Figura 4.2: Variograma direcional para as quatro direções principais dos dados de precipitação.



(a)

Figura 4.3: Gráfico de envelope simulado para o variograma omnidirecional dos dados de precipitação.

o que indica isotropia. Podemos, então, trabalhar com o semivariograma omnidirecional, que considera apenas o módulo do vetor de distância sem considerar o sentido. A Figura 4.3 apresenta o gráfico de envelope simulado construído a partir de permutações aleatórias dos dados (foram consideradas 99). Para cada permutação, o valor do semivariograma empírico é calculado e a banda é construída a partir dos valores máximo e mínimo do semivariograma (considerando também os valores originais). Note que vários pontos se encontram fora dos limites do envelope, indicando claramente a existência de dependência espacial.

Desse modo, foram ajustados dois modelos, um com média constante e outro com covariáveis, expressos por:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (4.1)$$

$$Y(\mathbf{s}) = \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (4.2)$$

em que X_1 e X_2 representam as coordenadas na direção leste-oeste e norte-sul, respectivamente. Os modelos foram ajustados considerando as distribuições normal, t-Student com 5 g.l. e slash com $\nu = 5$. Também foram consideradas as funções de covariância exponencial e Gaussiana.

4.1.1 Análise sem covariável

Ajustamos modelos com média constante (4.1) para os dados de precipitação considerando as distribuições normal, t-Student com 1 g.l. e slash com $\nu = 1$ e funções de covariância exponencial e Gaussiana. A Tabela 4.2 apresenta as estimativas dos parâmetros dos modelo ajustados e seus respectivos erros padrão entre parênteses. O efeito pepita relativo (EPR) mede o impacto do efeito pepita no processo espacial e é calculado por $EPR = \hat{\phi}_1 / (\hat{\phi}_1 + \hat{\phi}_2)$. Cambardella et al. (1994) propõe uma escala de classificação, na qual a dependência espacial é considerada forte para $EPR \leq 0,25$, moderada se $0,25 < EPR \leq 0,75$ e fraca se $EPR > 0,75$. Note que a estimativa para o parâmetro da média é semelhante para todos os modelos ajustados. A estimativa de ϕ_1 é pequena se comparada à de ϕ_2 em quase todos os modelo ajustados, apresentando valores baixos de EPR , o que indica forte dependência espacial. Não se obteve convergência na estimação dos parâmetros do modelo com distribuição slash e função de covariância Gaussiana, não sendo então considerado. Sobre o modelo ajustado com distribuição t-Student e função de covariância Gaussiana, observamos pequeno valor da estimativa de ϕ_2 se comparado com os demais modelos, resultando em uma variância menor do que a esperada pelo gráfico de semivariograma (Figura 4.3).

A Tabela 4.3 apresenta as estatísticas utilizadas para seleção de modelos, sendo elas o erro médio

Tabela 4.2: Estimativas e erros padrão entre parênteses dos parâmetros dos modelos ajustados sem covariável.

Dist.	Função de cov.	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{a} = g(\hat{\phi}_3)$	EPR
Normal	Exponencial	179,345 (26,797)	0,000 (306,371)	6562,944 (5514,921)	39,390 (46,576)	118,172	*0,000
	Gaussiana	179,245 (16,228)	577,166 (430,133)	6607,215 (2868,842)	24,908 (8,361)	43,143	0,080
t_1	Exponencial	155,898 (18,113)	0,000 (135,591)	2906,558 (4835,899)	39,763 (47,575)	119,290	*0,000
	Gaussiana	180,132 (0,336)	0,000 (0,008)	4,864 (6,992)	16,432 (0,857)	28,461	0,000
s_1	Exponencial	176,398 (29,649)	0,516 (12,184)	2890,664 (216,293)	39,516 (4,215)	118,548	0,000

*Estimativa menor que 10^{-4} .

Tabela 4.3: Estatísticas de validação de modelo para os modelos ajustados sem covariável.

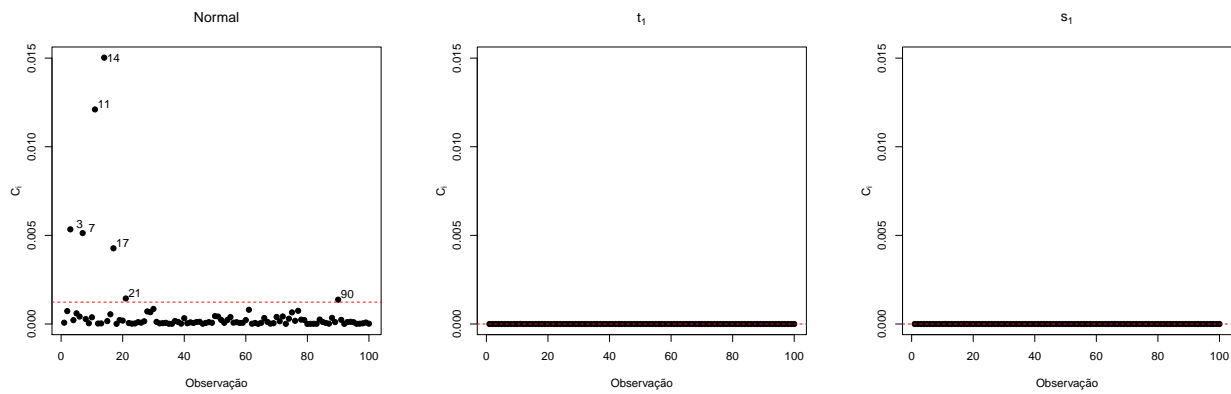
Dist	Função cov.	EMR	ρ	LMV	AIC
Normal	Exponencial	-0,000064	0,810	-504,011	1016,023
	Gaussiana	-0,000304	0,779	-496,414	999,829
t_1	Exponencial	0,000110	0,810	-667,766	1343,534
	Gaussiana	-0,000255	0,777	-670,619	1349,239
s_1	Exponencial	0,000359	0,803	-635,570	1279,14

reduzido (EMR), a correlação de Pearson entre os dados observados e os dados preditos de precipitação (ρ), o valor da função de log-verossimilhança (LMV) e o valor do AIC . Podemos notar que o modelo com menor valor de AIC e maior valor de LMV é o modelo com distribuição normal e função de covariância Gaussiana, sendo o melhor modelo sob esses critérios. Como os valores das estatísticas de validação para o modelo ajustado com distribuição t-Student foram próximos considerando função de covariância exponencial e Gaussiana, escolheu-se o primeiro devido aos melhores valores de LMV e AIC .

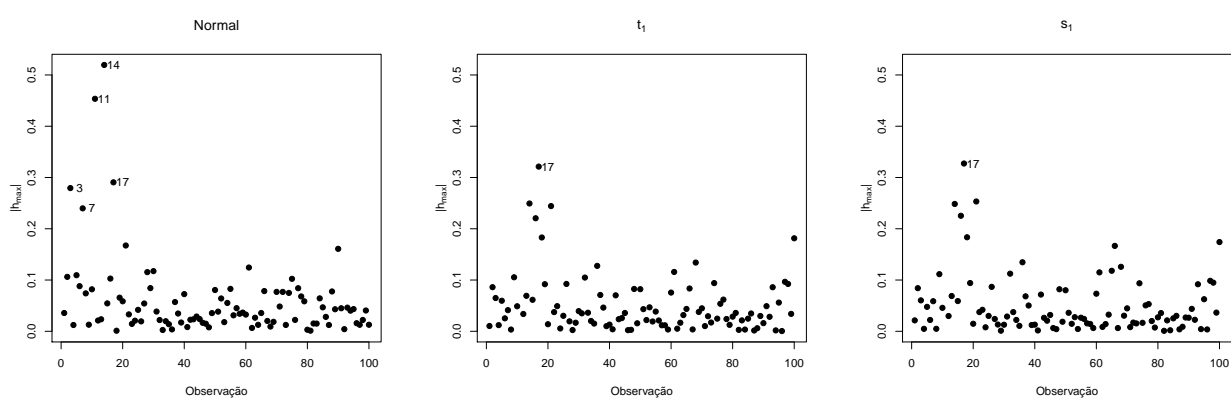
A seguir realizamos a análise de diagnóstico para os modelos com distribuição normal e função de covariância Gaussiana, t-Student com 1 g.l. e slash com $\nu = 1$, ambos com função de covariância exponencial.

A Figura 4.4 apresenta os gráficos de influência local para os modelos ajustados. Para o modelo ajustado com distribuição normal, é possível notar que a observação 14 e suas vizinhas 3, 7, 11 e 17 foram detectadas como influente nos gráficos de diagnóstico, sendo que a observação 14 apresentou maior valor de influência nos gráficos de C_i e $|\mathbf{h}_{max}|$. Essa observação deixa de ser influente quando ajustados modelos com distribuição t-Student e slash, assim como as observações vizinhas.

Os mapas temáticos de predição podem ser observados na Figura 4.5. Notamos que, considerando a distribuição normal, a região em torno da observação 14 apresenta valores maiores se comparados com aqueles considerando as distribuições t-Student e slash.



(a)



(b)

Figura 4.4: Gráficos de (a) C_i , (b) $|h_{max}|$ e (c) $|h_{max}|_2$ do conjunto de dados de precipitação para os modelos sem covariável ajustados com distribuição normal e t-Student com 5 g.l., ambos com função de covariância Gaussiana, e distribuição slash com $\nu = 5$ e função de covariância exponencial.

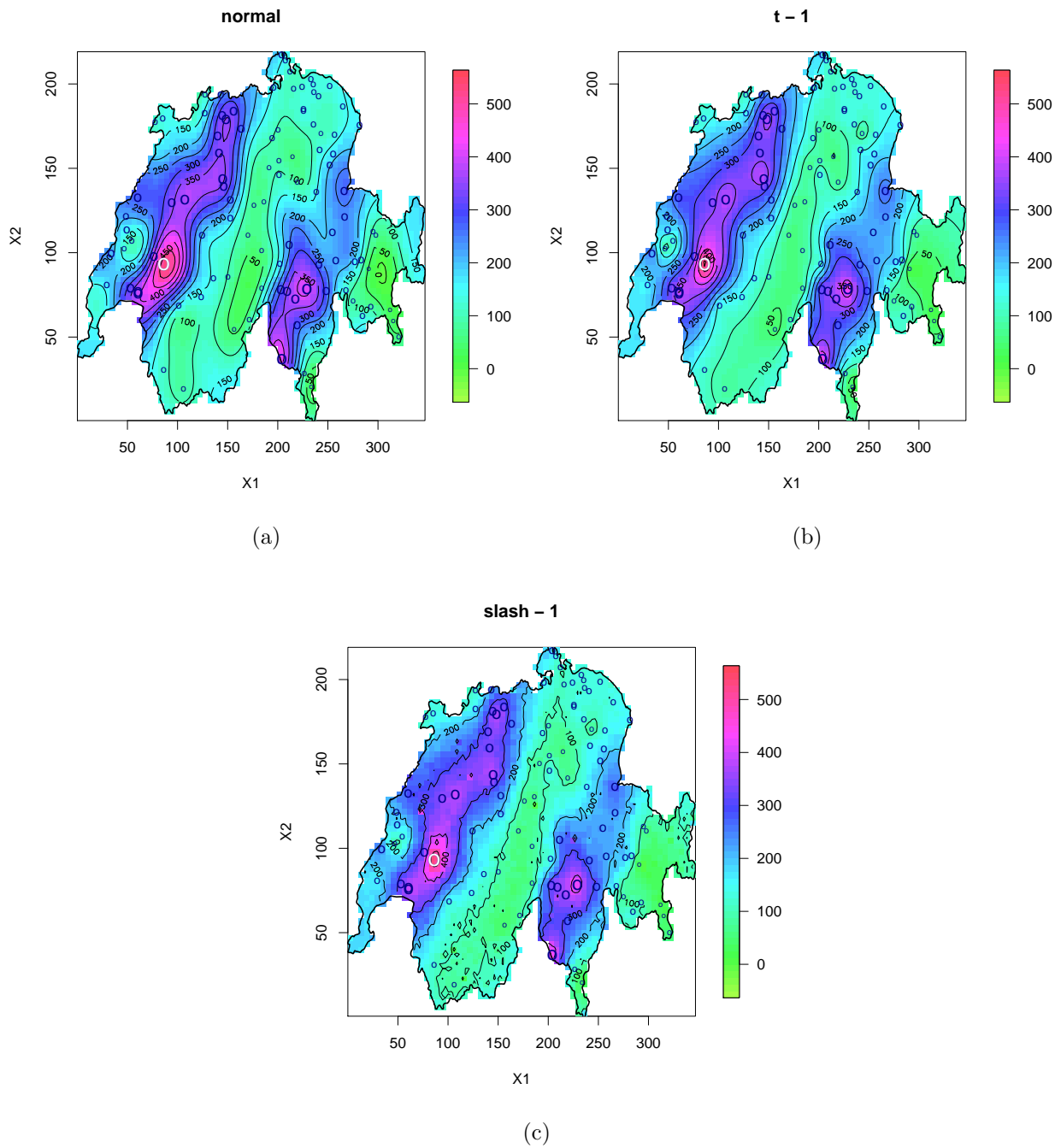


Figura 4.5: Mapas temáticos de previsão para os dados de precipitação considerando modelos de média constante ajustados com (a) distribuição normal e função de covariância Gaussiana, (b) distribuição t-Student e função de covariância Gaussiana e (c) distribuição slash e função de covariância exponencial.

4.1.2 Análise com covariável

Ajustamos modelos com covariáveis (4.2) para os dados de precipitação considerando as distribuições normal, t-Student com 1 g.l. e slash com $\nu = 1$ e funções de covariância exponencial e Gaussiana. A Tabela 4.4 apresenta os parâmetros estimados e respectivos erros padrão entre parênteses dos modelos ajustados. Note que as estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ foram semelhantes para todos os modelos ajustados. O valor de $\hat{\beta}_2$ é pequeno, o que corrobora com a análise descritiva na qual não observamos relação entre a coordenada X2 e os dados de precipitação. A estimativa de ϕ_1 dos modelos com função de covariância exponencial foi baixa comparada ao valor da estimativa de ϕ_2 . Em todos os modelos ajustados, o valor do *EPR* foi menor que 0,25, indicando que todos os modelos captaram a existência de dependência espacial.

A Tabela 4.5 apresenta as estatísticas utilizadas para seleção de modelos. É possível observar que os melhores valores de *AIC* e *LMV* estão relacionados aos modelos ajustados com função de covariância Gaussiana e, como as outras estatísticas apresentaram valores similares, escolheu-se realizar a análise de diagnóstico para os modelos com distribuição normal, t-Student com 1 g.l. e slash com $\nu = 1$ considerando a função de covariância Gaussiana.

A Figura 4.6 apresenta os gráficos de influência local para os modelos ajustados. Para o modelo ajustado com distribuição normal, é possível notar que a observação 14 e suas vizinhas 3, 7, 11 e 17 foram detectada como influente nos gráficos de diagnóstico, sendo que a observação 14 apresentou maior valor de influência nos gráficos de C_i e $|\mathbf{h}_{max}|$. Essa observação deixa de ser influente quando ajustados modelos com distribuição t-Student e slash, assim como as ob-

Tabela 4.4: Estimativas e erros padrão entre parênteses dos parâmetros dos modelos ajustados com covariável.

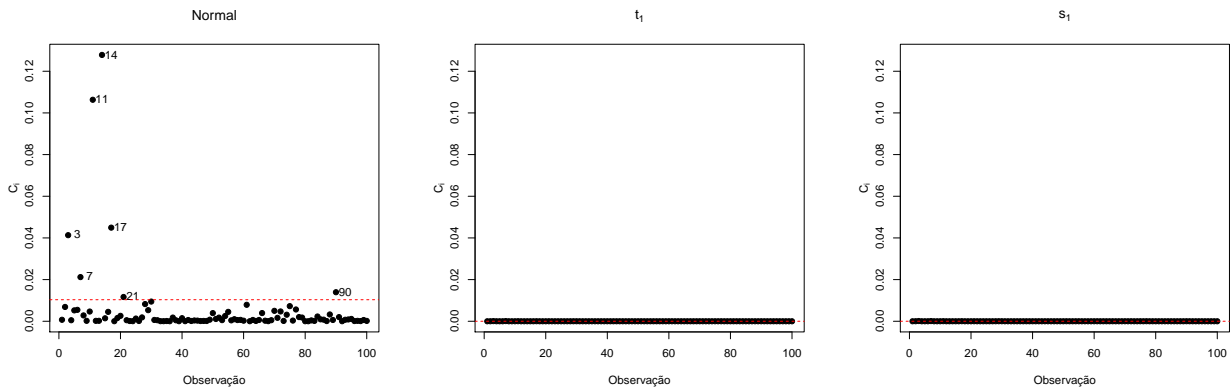
Dist.	$C(\cdot)$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{a} = g(\hat{\phi}_3)$	<i>EPR</i>
Normal	Exp.	284,694 (22,892)	-0,526 (0,091)	-0,049 (0,123)	0,250 (6,885)	5906,825 (103,525)	35,908 (0,937)	107,726	0,000
	Gauss.	285,124 (47,220)	-0,527 (0,188)	-0,049 (0,271)	528,795 (408,975)	5913,497 (2503,513)	23,899 (8,218)	41,394	0,082
t_1	Exp.	283,683 (20,170)	-0,617 (0,081)	-0,010 (0,109)	0,000 (32,953)	604,664 (984,027)	34,609 (40,248)	103,827	0,000
	Gauss.	278,420 (30,803)	-0,545 (0,123)	-0,000 (0,176)	241,808 (309,593)	2361,612 (3539,836)	24,771 (9,219)	42,905	0,092
s_1	Exp.	279,089 (67,934)	-0,607 (0,272)	-0,011 (0,366)	0,000 (6,996)	2160,551 (96,518)	34,224 (2,330)	102,673	0,000
	Gauss.	274,375 (64,148)	-0,525 (0,256)	0,009 (0,367)	173,402 (32,092)	2002,925 (192,649)	23,892 (1,960)	41,382	0,079

Tabela 4.5: Estatísticas de validação de modelo para os modelos ajustados com covariável.

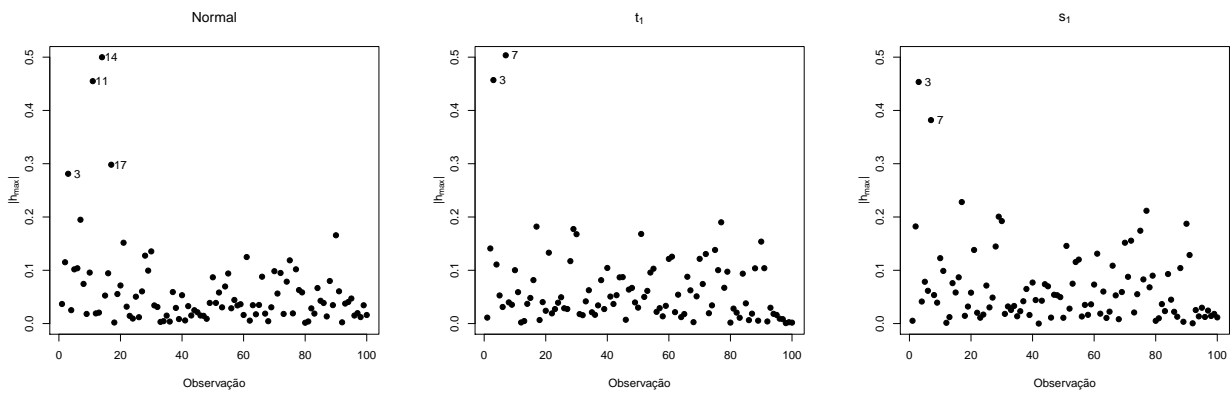
Dist.	Modelo	EMR	ρ	LMV	AIC
Normal	Exponencial	-0,000288	0,806	-505,563	1023,127
	Gaussiana	-0,000456	0,770	-496,692	1005,385
t_1	Exponencial	0,000106	0,806	-668,183	1348,368
	Gaussiana	0,000341	0,759	-666,453	1344,906
s_1	Exponencial	-0,000747	0,828	-635,313	1282,628
	Gaussiana	0.003003	0,729	-634,213	1280,427

servações vizinhas, sendo as observações 3 e 7 detectadas no gráfico de $|\mathbf{h}_{max}|$ para os modelos ajustados com ambas distribuições. Isso confirma a robustez dessas distribuições com caudas mais pesadas.

Os mapas temáticos de predição podem ser observados na Figura 4.7. Apesar de a observação 14 não ser considerada influente na análise de diagnóstico para os modelos com distribuição t-Student e slash, note que os mapas são bem semelhantes. Esse fato pode ser explicado pela presença de covariáveis, induzindo à mesma predição média para os três modelos.



(a)



(b)

Figura 4.6: Gráficos de (a) C_i , (b) $|h_{max}|$ e (c) $|h_{max}|_2$ do conjunto de dados de precipitação para os modelos com covariável ajustados com distribuição normal, t-Student com 5 g.l. e slash com $nu = 5$ e função de covariância Gaussiana.

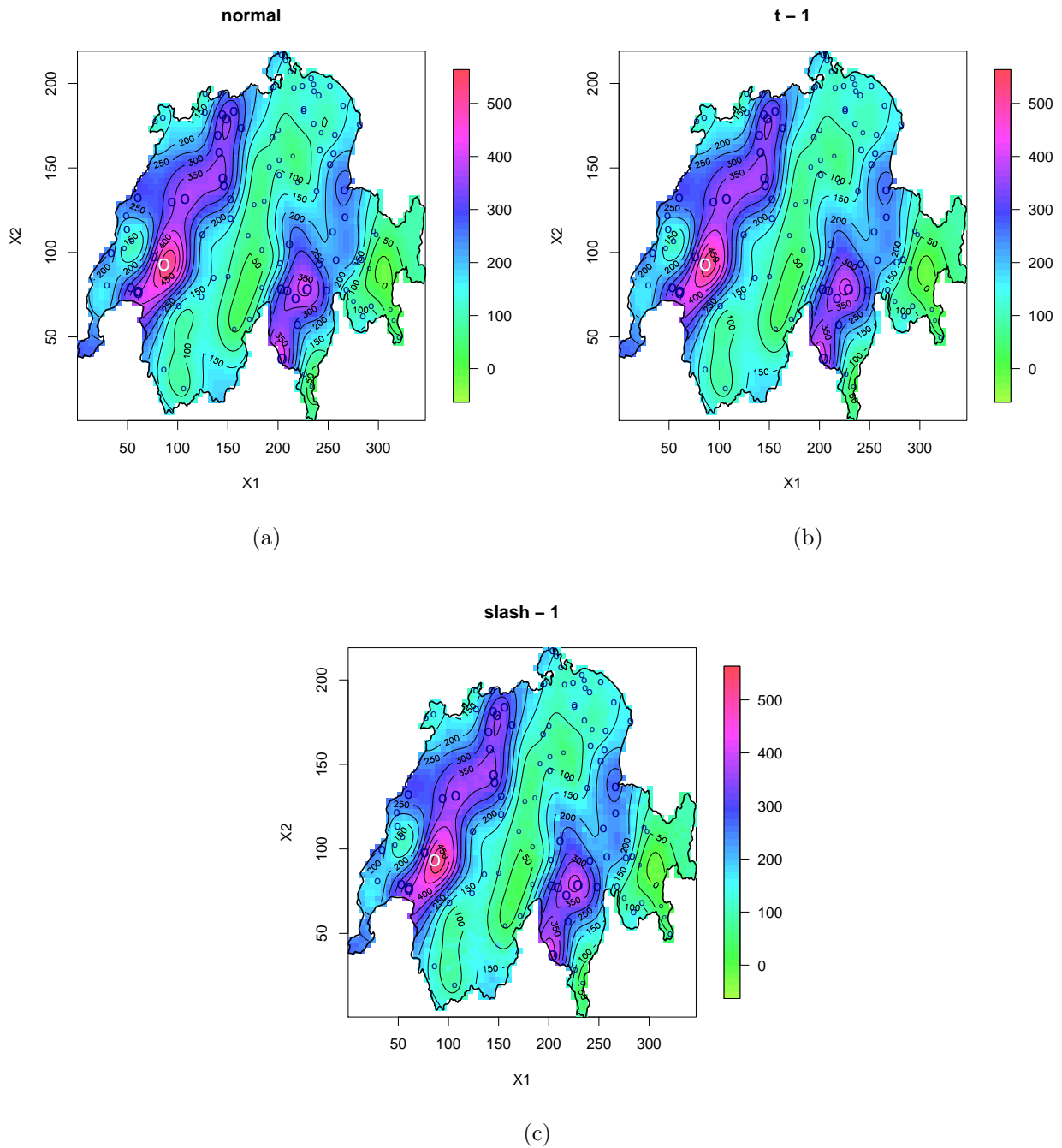


Figura 4.7: Mapas temáticos de previsão para os dados de precipitação considerando modelos com covariável ajustados com distribuição (a) normal, (b) t-Student e (c) distribuição slash e função de covariância Gaussiana.

4.1.3 Análise sem a observação 14

Foram ajustados os modelos que foram utilizados para a análise de diagnósticos sem a observação 14 e a Tabela 4.6 apresenta os parâmetros estimados. Note que a retirada dessa observação resultou em valores menores das estimativas de β e ϕ_1 , se comparados com as estimativas incluindo essa observação (Tabelas 4.2 e 4.4). $\hat{\phi}_2$ e $\hat{\phi}_3$ apresentaram pequeno aumento. Os valores de EPR dos modelos foram menores que 0,25 indicando forte dependência espacial. A Tabela 4.7 apresenta as estatísticas utilizadas para seleção de modelos. Notamos que, após retirar a observação 14, os valores de LMV e AIC sofreram aumento e decréscimo, respectivamente, em todos os modelos, indicando melhor qualidade de ajuste. O modelo com covariável é o melhor nos critérios de validação, entretanto os valores são semelhantes para os modelos sem e com covariável.

A Figura 4.8 apresenta os mapas temáticos de predição e notamos que são similares. Note que na região em que se localiza a observação 14 (ponto branco) não há mais cor vermelha, o que indica que essa observação era responsável pela predição de valores altos em localizações vizinhas. Considerando a distribuição normal, os valores preditos para essa região são menores se comparados aos valores preditos considerando as distribuições t-Student e slash, como pode ser visto pelas curvas de nível. Ou seja, ao retirar a observação 14, os modelos com distribuição t-Student e slash, ambos com $\nu = 1$, foram capazes de predizer melhores valores para essa região, já que era esperado maiores valores devido a observação 14, indicando a robustez dessas distribuições.

Tabela 4.6: Estimativas e erros padrão entre parênteses dos parâmetros dos modelos ajustados sem a observação 14.

Dist.	Modelo	$C(\cdot)$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	EPR
Normal	Sem cov.	Gauss	175,099 (15,413)	- -	- -	556,772 (411,417)	5834,008 (2576,536)	25,208 (8,867)	0,087
	Com cov.	Gauss	264,005 (45,782)	-0,461 (8,183)	-0,009 (0,261)	533,955 (401,693)	5333,885 (2346,038)	24,668 (8,959)	0,091
t_1	Sem cov.	Exp	163,410 (9,242)	- -	- -	0,000 (23,066)	606,201 (1038,690)	47,829 (59,461)	0,000
	Com cov.	Gauss	263,465 (31,746)	-0,499 (0,127)	0,028 (0,181)	236,766 (384,060)	2556,188 (3813,912)	24,371 (8,569)	0,085
s_1	Sem cov.	Exp	164,171 (26,997)	- -	- -	0,718 (12,020)	3255,414 (236,315)	42,121 (4,258)	0,000
	Com cov.	Gauss	258,487 (49,768)	-0,488 (0,198)	0,048 (0,284)	155,118 (9,691)	1442,153 (51,315)	24,503 (0,782)	0,097

Tabela 4.7: Estatísticas de validação de modelo para os modelos ajustados sem a observação 14.

Dist.	Modelo	$C(\cdot)$	EMR	ρ	LMV	AIC
Normal	Normal sem cov.	Gauss	-0,000042	0,822	-480,764	967,529
	Normal com cov.	Gauss	-0,000312	0,818	-480,452	966,904
t_1	Normal sem cov.	Exp	0,000039	0,839	-652,781	1313,563
	Normal com cov.	Gauss	-0,001171	0,797	-651,076	1314,152
s_1	Normal sem cov.	Exp	-0,000562	0,846	-619,790	1247,580
	Normal com cov.	Gauss	0,000367	0,795	-619,387	1250,776

4.1.4 Consideração geral

Com o auxílio das técnicas geoestatísticas foi possível realizar uma análise correta do problema. Foi possível construir mapas de predição dos dados de precipitação pluviométrica para as distribuições em estudo considerando a dependência espacial do processo. Notamos que a observação 14 foi detectada como influente nos modelos sem e com covariável considerando a distribuição normal. Para as distribuições t-Student com 1 g.l. e slash com $\nu = 1$, essa observação não foi detectada na análise de diagnóstico. Observamos o impacto de se retirar essa observação na Figura 4.8, notando que os valores preditos de precipitação na área em torno da observação 14 foram menores quando comparados aos mapas dos modelos considerando essa observação (Figuras 4.5 e 4.7). Podemos, então, optar pela retirada da observação (por exemplo, caso o dado esteja errado por erro de medida) ou o uso de uma distribuição para incorporar essa observação de forma consistente, como a t-Student e a slash. Essa decisão deve ser tomada em conjunto com o pesquisador da área.

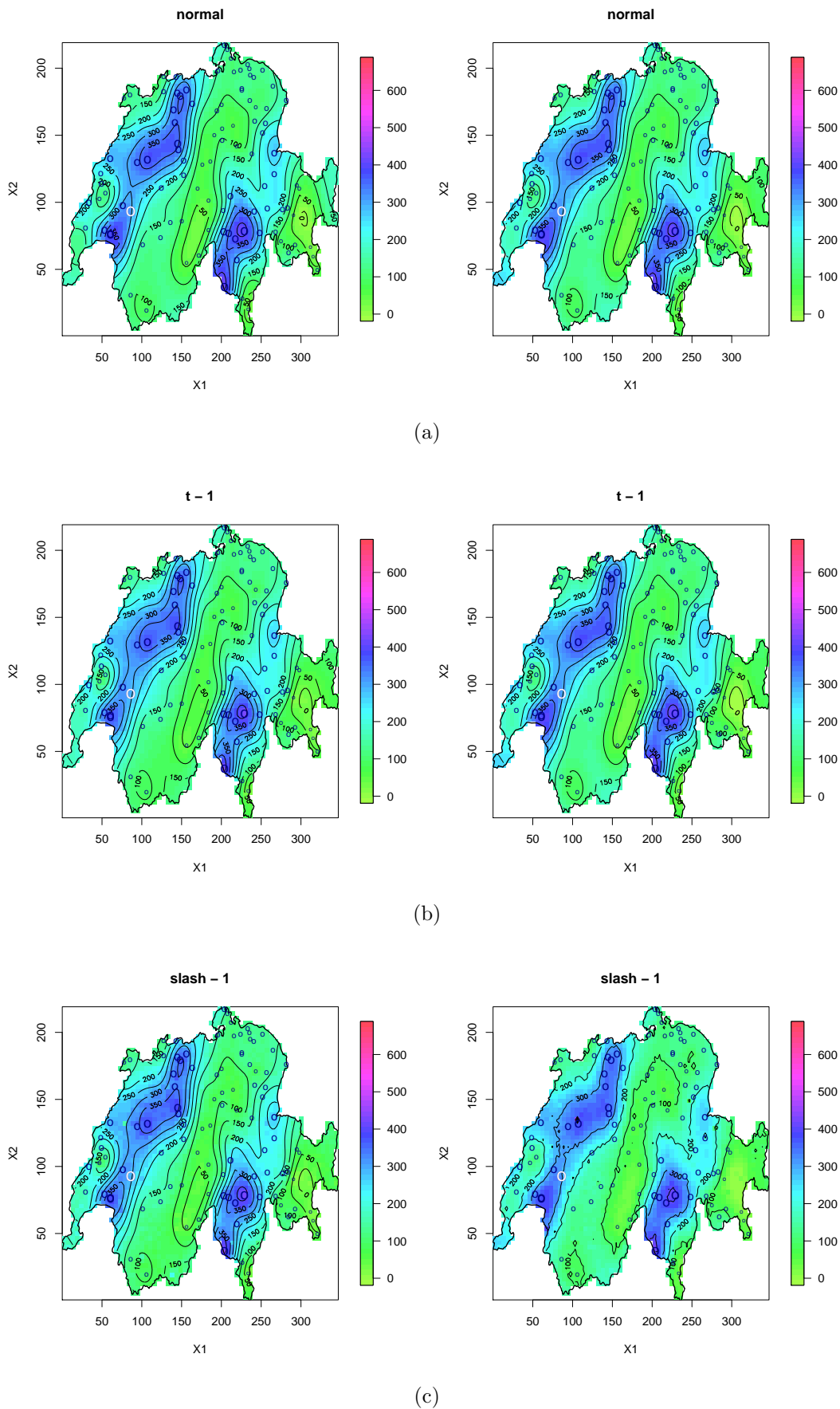


Figura 4.8: Mapas temáticos de predição considerando distribuição (a) normal, (b) t-Student com 1 g.l. e (c) slash com $\nu = 1$ para os modelos sem covariável (primeira coluna) e com covariável (segunda coluna) retirando a observação 14.

Capítulo 5

Considerações finais

Para fenômenos em que os dados são correlacionados espacialmente, faz-se necessário o uso de técnicas geoestatísticas para a análise correta do problema. Nesta dissertação apresentamos, como contribuição deste trabalho, um modelo geoestatístico considerando uma classe de distribuições com caudas mais pesadas que a distribuição normal. Além disso, foi desenvolvida a técnica de influência local para o modelo em estudo.

A estimação dos parâmetros desse tipo de modelo não é uma tarefa fácil, visto que envolve problemas de otimização numérica. Como exemplos de dificuldades observadas citamos a não convergência do algoritmo BFGS para algumas amostras simuladas e a correlação existente entre os parâmetros de variância. Também foi possível observar no estudo com dados simulados que o método de curvatura normal desenvolvido por Cook (1986) aplicado ao modelo geoestatístico com distribuição normal/independente foi eficiente na identificação de pontos influentes. Além disso, a observação perturbada deixou de ser influente quando aplicado os modelos ajustados com distribuições t-Student e slash, evidenciando a robustez de tais distribuições. A aplicação da metodologia estudada em um conjunto de dados reais reforça a justificativa desse trabalho, apresentando evidências empíricas de que o estudo é útil e contribui na análise de problemas reais.

Como pesquisas futuras podemos citar o uso de outras distribuições da classe de distribuição normal/independente, como as distribuições normal contaminada e exponencial potência, além de desenvolver a estimação do vetor de parâmetros ν da distribuição de mistura $H(u)$. Pretende-se também realizar o estudo de influência local através do Q -afastamento da verossimilhança para dados incompletos, proposto por Zhu and Lee (2001), e, assim, realizar a análise de diagnóstico de modelos ajustados através do algoritmo EM.

Referências Bibliográficas

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.
- Andriotti, J. (2003). *Fundamentos de Estatística e Geoestatística*. Editora Unisinos.
- Assumpção, R. (2010). Influência local em modelos espaciais geoestatísticos t-student com aplicações a dados agrícolas. Master’s thesis, Universidade Estadual do Oeste do Paraná.
- Cambardella, C., Moorman, T., Novak, J., Parkin, T., Karden, D., Turco, R., and Konopka, A. (1994). Field-scale variability of soil properties in central iowa soils. *Soil Science Society of America Journal*, 58:1501–1511.
- Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 48:133–169.
- Cook, R. D. (1987). Influence assessment. *Journal of Applied Statistics*, 14:117–131.
- De Bastiani, F. (2012). Influência local em modelos espaciais lineares com distribuição da família de contornos elípticos. Master’s thesis, Universidade Federal de Pernambuco.
- De Oliveira, V., Kedern, B., and Short, D. (1997). Bayesian prediction of transformed gaussian random fields. *Journal of the American Statistical Association*.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.
- Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.
- Diggle, P., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the royal statistical society. Series C*, pages 299–350.
- Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain (2015). *fields: Tools for spatial data*. R package version 8.10.
- Dubois, G. (2003). Spatial interpolation comparison 97: foreword and introduction. *Journal of geographic information and decision analysis*, 2:1–11.

- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Garay, A. (2014). *Modelos de regressão para dados censurados sob distribuições simétricas*. PhD thesis, Universidade de São Paulo.
- Garcia-Papani, F., Uribe-Opazo, M. A., Leiva, V., and Aykroyd, R. G. (2016). Birnbaum-saunders spatial modelling and diagnostics applied to agricultural engineering data. *Stochastic Environmental Research and Risk Assessment*, pages 1–20.
- Geyer, J. C. (2007). Stat5102 notes: Fisher information and confidence intervals using maximum likelihood.
- Hongyu, K. (2012). Seleção e ajuste de modelos espaciais de semivariograma aplicados a dados de ph do solo. Technical report, Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo.
- Jácomo, C. and Flores, E. (2009). Aplicação de geoestatística na caracterização dos casos de dengue no município de presidente prudente. In *Simpósio de Geoestatística Aplicada em Ciências Agrárias*.
- Lange, K. and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of computational and graphical statistics*, 2(2):175–198.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.
- Oda-Souza, M., Barbin, D., Ribeiro Jr., P. J., and Stape, J. (2008). Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados de um ensaio de espaçamento florestal em delineamento sistemático tipo leque. *Revista Árvore*, 32:499 – 509.
- Paula, G. A. (2010). Modelos de regressão com apoio computacional.
- Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA.

- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ribeiro Jr., P. and Diggle, P. (2001). geoR: a package for geostatistical analysis. *R-NEWS*, 1(2):15–18.
- Roth, M. (2013). On the multivariate t distribution. Technical report, Linköpings universitet.
- Shing, L., Oliveira, M. D., Macedo, J., Da Silva, E., Capeche, C., and Hissa, H. (2012). Aplicação de geoestatística na determinação da variabilidade espacial do teor de argila. In *VI Congresso brasileiro de soja*.
- Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.
- Valle, R. B. A. (1994). *Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão*. PhD thesis, Universidade de São Paulo.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.
- Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society Series B*, 63:111–126.

Apêndice A: Estimação dos parâmetros por máxima verossimilhança

Neste apêndice são obtidos os estimadores de máxima verossimilhança para o modelo geostatístico apresentado em (2.1). Apresentamos a seguir os cálculos que levaram às equações para $\boldsymbol{\beta}$ e ϕ dadas em (2.9) e (2.10).

Expandimos a forma de δ como

$$\begin{aligned}
 \delta &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= \mathbf{Y}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= \mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} - \mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \\
 &= \mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta},
 \end{aligned}$$

em que o último passo foi obtido pelo fato de que $\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$, pois são escalares. Utilizando resultados de diferenciação matricial, a derivada de δ em relação a $\boldsymbol{\beta}$ é expressa por

$$\begin{aligned}
 \frac{d\delta}{d\boldsymbol{\beta}} &= -2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} + 2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \\
 &= -2\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= -2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon},
 \end{aligned} \tag{A.1}$$

em que $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$. A função escore para $\boldsymbol{\beta}$ tem a forma

$$\begin{aligned}
 U(\boldsymbol{\beta}) &= \frac{dl(\boldsymbol{\theta}, \mathbf{y})}{d\boldsymbol{\beta}} \\
 &= \frac{1}{g(\delta)} \frac{dg(\delta)}{d\boldsymbol{\beta}} \\
 &= \frac{1}{g(\delta)} \frac{dg(\delta)}{d\delta} \frac{d\delta}{d\boldsymbol{\beta}} \\
 &= \frac{g'(\delta)}{g(\delta)} (-2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) \\
 &= V_{g(\delta)}(-2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}).
 \end{aligned} \tag{A.2}$$

Igualando (A.2) a zero, temos que

$$U(\boldsymbol{\beta}) = 0 \rightarrow V_{g(\delta)}(-2\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}) = 0.$$

Essa identidade é verdadeira apenas quando o termo $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}$ se torna zero. Então,

$$\begin{aligned}
\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon} &= 0 \\
\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= 0 \\
\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \\
(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y},
\end{aligned} \tag{A.3}$$

que é o estimador apresentado em (2.9).

Em relação a equação de ϕ , primeiro derivamos δ em relação a ϕ_j , tomando a forma

$$\begin{aligned}
\frac{d\delta}{d\phi_j} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \frac{d(\boldsymbol{\Sigma}^{-1})}{d\phi_j} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \left(-\boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \boldsymbol{\Sigma}^{-1} \right) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= -\boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon},
\end{aligned} \tag{A.4}$$

e temos então que

$$\begin{aligned}
\frac{dl(\boldsymbol{\theta}, \mathbf{y})}{d\phi_j} &= -\frac{1}{2} \frac{d(\log |\boldsymbol{\Sigma}|)}{d\phi_j} + V_{g(\delta)} \frac{d\delta}{d\phi_j} \\
&= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{d \log \boldsymbol{\Sigma}}{d\phi_j} \right) - V_{g(\delta)} \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \frac{d\boldsymbol{\Sigma}}{d\phi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon},
\end{aligned} \tag{A.5}$$

que é a equação apresentada em (2.10).

Apêndice B: Reparametrização da matriz escala Σ

A seguir serão apresentados os cálculos necessários para a obtenção da função de log-verossimilhança concentrada (2.14) e da função Q concentrada (2.22). Sejam $\omega = \phi_1/\phi_2$, $\mathbf{V} = \omega\mathbf{I} + \mathbf{R}(\phi_3)$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi_1, \phi_2, \phi_3)$ e $\boldsymbol{\theta}^* = (\boldsymbol{\beta}', \omega, \phi_2, \phi_3)$.

B.1 Aplicada a função de log-verossimilhança

Podemos escrever a matriz de covariâncias como $\Sigma = \phi_2\mathbf{V}$ e então substituir na função de log-verossimilhança em (2.4). Obtemos

$$\begin{aligned} l(\boldsymbol{\theta}^*, \mathbf{y}) &= -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\phi_2\mathbf{V}| - \frac{\delta^*}{2\phi_2} \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \phi_2 - \frac{1}{2} \log |\mathbf{V}| - \frac{\delta^*}{2\phi_2}, \end{aligned} \quad (\text{B.1})$$

em que $\delta^* = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

A função escore para $\boldsymbol{\beta}$ é expressa por

$$\begin{aligned} U(\boldsymbol{\beta}) &= \frac{dl(\boldsymbol{\theta}^*, \mathbf{y})}{d\boldsymbol{\beta}} \\ &= -\frac{1}{2\phi_2} \frac{d\delta^*}{d\boldsymbol{\beta}} \\ &= \frac{1}{\phi_2} \mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\epsilon}. \end{aligned} \quad (\text{B.2})$$

Ao igualar (B.2) a zero, as contas que seguem são equivalentes àquelas da expressão em (A.3).

O estimador de $\boldsymbol{\beta}$ fica dado por

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

Para encontrar o estimador de ϕ_2 , derivamos (B.1) em relação a ϕ_2 e obtemos

$$\frac{dl(\boldsymbol{\theta}^*, \mathbf{y})}{d\phi_2} = -\frac{n}{2\phi_2} + \frac{1}{2\phi_2^2}\delta^*. \quad (\text{B.3})$$

Igualando (B.3) a zero,

$$\begin{aligned}\frac{n}{2\phi_2} &= \frac{\delta^*}{2\phi_2^2} \\ 2\phi_2^2 n - 2\phi_2 \delta^* &= 0 \\ 2\phi_2(\phi_2 n - \delta^*) &= 0 \\ \hat{\phi}_2(\mathbf{V}) &= \frac{\delta^*}{n}.\end{aligned}$$

Substituindo $\hat{\phi}_2(\mathbf{V})$ e $\hat{\beta}(\mathbf{V})$ em (B.1), temos a função de log-verossimilhança concentrada dada por

$$\begin{aligned}l(\boldsymbol{\theta}^*, \mathbf{y}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2\frac{\delta^*}{n}} \delta^* \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2} \log |\mathbf{V}| - \frac{n}{2}\end{aligned}\quad (\text{B.4})$$

Para se estimar ω e ϕ_3 , maximizamos a função de log-verossimilhança concentrada em relação aos dois parâmetros. Podemos desconsiderar os termos constantes em relação a ω e ϕ_3 em (B.4), obtendo então

$$l(\omega, \phi_3, \mathbf{y}) = -\frac{n}{2} \log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2} \log |\mathbf{V}|.$$

B.2 Aplicada a função Q

Seja $E_U = E(U|\mathbf{Y}, \boldsymbol{\theta}^{*(n)})$. Podemos escrever a matriz de covariâncias como $\boldsymbol{\Sigma} = \phi_2 \mathbf{V}$ e então substituir na função Q em (2.16). Obtemos

$$Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(n)}) = -\frac{n}{2} \log \phi_2 - \frac{1}{2} \log |\mathbf{V}| - \frac{\delta^*}{2\phi_2} E_U. \quad (\text{B.5})$$

em que $\delta^* = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

A função escore para $\boldsymbol{\beta}$ é expressa por

$$\begin{aligned}U(\boldsymbol{\beta}) &= \frac{dQ(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(n)})}{d\boldsymbol{\beta}} \\ &= -\frac{E_U d\delta^*}{2\phi_2 d\boldsymbol{\beta}} \\ &= \frac{E_U}{\phi_2} \mathbf{X}' \mathbf{V}^{-1} \boldsymbol{\epsilon}.\end{aligned}\quad (\text{B.6})$$

Ao igualar (B.6) a zero, as contas que seguem são equivalentes àsquelas da expressão em (A.3).
O estimador de β fica dado por

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

Para encontrar o estimador de ϕ_2 , derivamos (B.5) em relação a ϕ_2 e obtemos

$$\frac{dQ(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(n)})}{d\phi_2} = -\frac{n}{2\phi_2} + \frac{\delta^*}{2\phi_2^2}E_U. \quad (\text{B.7})$$

Igualando (B.7) a zero,

$$\begin{aligned} \frac{n}{2\phi_2} &= \frac{\delta^*}{2\phi_2^2}E_U \\ 2\phi_2^2n - 2\phi_2\delta^*E_U &= 0 \\ 2\phi_2(\phi_2n - \delta^*E_U) &= 0 \\ \hat{\phi}_2(\mathbf{V}) &= \frac{\delta^*}{n}E_U. \end{aligned}$$

Substituindo $\hat{\phi}_2(\mathbf{V})$ e $\hat{\beta}(\mathbf{V})$ em (B.5), temos a função Q concentrada dada por

$$\begin{aligned} Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(n)}) &= -\frac{n}{2}\log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2}\log |\mathbf{V}| - \frac{n}{\delta^*E_U}\frac{\delta^*}{2}E_U, \\ &= -\frac{n}{2}\log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2}\log |\mathbf{V}| - \frac{n}{2}. \end{aligned} \quad (\text{B.8})$$

Para se estimar ω e ϕ_3 , maximizamos a função Q concentrada em relação aos dois parâmetros. Podemos desconsiderar os termos constantes em relação a ω e ϕ_3 em (B.8), obtendo então

$$Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(n)}) = -\frac{n}{2}\log \hat{\phi}_2(\mathbf{V}) - \frac{1}{2}\log |\mathbf{V}|.$$

Apêndice C: Valor esperado da distribuição condicional de U

A seguir serão apresentados os cálculos necessários para a obtenção da expressão de $E(U|\mathbf{Y})$ para as distribuições t-Student e slash.

C.1 Distribuição t-Student

A variável aleatória U possui distribuição $Gama(\nu/2, \nu/2)$. A função densidade de probabilidade de $U|\mathbf{Y}$ será proporcional a

$$\begin{aligned} f_{U|\mathbf{Y}}(u) &\propto f_{\mathbf{Y}|U}(\mathbf{y})f_U(u) \\ &\propto u^{\frac{n}{2}} \exp\left\{-\frac{u\delta}{2}\right\} u^{\frac{\nu}{2}-1} \exp\left\{-\frac{u\nu}{2}\right\} \\ &\propto u^{\frac{n}{2}+\frac{\nu}{2}-1} \exp\left\{-\frac{1}{2}(\nu + \delta)u\right\}, \end{aligned}$$

em que a última expressão é o núcleo de uma distribuição $Gama((n + \nu)/2, (\nu + \delta)/2)$. Então,

$$E(U|\mathbf{Y}) = \frac{\nu + n}{\nu + \delta}.$$

C.2 Distribuição slash

A variável aleatória U possui distribuição $Beta(\nu, 1)$. A função densidade de probabilidade de U é

$$f_U(u) = \nu u^{\nu-1}, \quad u \in [0, 1] \text{ e } \nu > 0.$$

A função densidade de probabilidade de $U|\mathbf{Y}$ será

$$\begin{aligned} f_{U|\mathbf{Y}}(u) &= \frac{f_{\mathbf{Y}|U}(\mathbf{y})f_U(u)}{f_{\mathbf{Y}}(\mathbf{y})} \\ &= \frac{u^{n/2}(2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\{-(u\delta)/2\}\nu u^{\nu-1}}{\nu(2\pi)^{-n/2}|\Sigma|^{-1/2}G(n/2 + \nu, \delta/2)} \\ &= \frac{u^{n/2+\nu-1}\exp\{-(u\delta)/2\}}{G(n/2 + \nu, \delta/2)}, \quad u \in [0, 1] \text{ e } \nu > 0. \end{aligned} \tag{C.1}$$

Então, o termo $E(U|\mathbf{Y})$ é expresso por

$$\begin{aligned} E(U|\mathbf{Y}) &= \int_0^1 u f_{U|\mathbf{Y}}(u) du \\ &= \frac{\int_0^1 u^{n/2+\nu} \exp\{-(u\delta)/2\} du}{G(n/2 + \nu, \delta/2)} \\ &= \frac{G(n/2 + \nu + 1, \delta/2)}{G(n/2 + \nu, \delta/2)}. \end{aligned}$$

Apêndice D: Maximização da função Q

Primeiro vamos maximizar a função Q , dada em (2.16), em relação a β . Derivando a função Q em relação a β , obtemos

$$\frac{dQ(\theta|\theta^{(n)})}{d\beta} = -\frac{1}{2}E(U|Y, \theta^{(n)})\frac{d\delta^2}{d\beta}, \quad (\text{D.1})$$

e $d\delta^2/d\beta$ possui a forma em (A.3). Note que igualar (D.1) a zero para encontrar pontos de máximo é equivalente a igualar $d\delta^2/d\beta$ a zero. As contas são equivalentes às aquelas apresentadas no Apêndice A, obtendo

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

que é o estimador mostrado em (2.17).

Para obter a solução (2.18), maximizamos (2.16) em relação a ϕ_j . Para facilitar a notação, considere $\omega_1 = E(U|\mathbf{Y}, \theta^{(n)})$ e $\epsilon = (\mathbf{Y} - \mathbf{X}\beta)$.

$$\begin{aligned} \frac{dQ(\theta|\theta^{(n)})}{d\phi_j} &= \frac{d}{d\phi_j} \left(-\frac{1}{2}\omega_1\epsilon'\Sigma^{-1}\epsilon \right) - \frac{d}{d\phi_j} \left(\frac{1}{2}\log|\Sigma| \right) \\ &= -\frac{1}{2}\omega_1\epsilon' \left(\frac{d}{d\phi_j}\Sigma^{-1} \right) \epsilon - \frac{1}{2} \left(\frac{d}{d\phi_j}\log|\Sigma| \right) \\ &= -\frac{1}{2}\omega_1\epsilon' \left(-\Sigma^{-1}\frac{d\Sigma}{d\phi_j}\Sigma^{-1} \right) \epsilon - \frac{1}{2}\text{tr} \left(\Sigma^{-1}\frac{d\Sigma}{d\phi_j} \right) \\ &= \frac{1}{2}\omega_1\epsilon' \left(\Sigma^{-1}\frac{d\Sigma}{d\phi_j}\Sigma^{-1} \right) \epsilon - \frac{1}{2}\text{tr} \left(\Sigma^{-1}\frac{d\Sigma}{d\phi_j} \right) \end{aligned} \quad (\text{D.2})$$

Igualando (D.2) a zero, temos

$$\begin{aligned} \frac{dQ(\theta|\theta^{(n)})}{d\phi_j} &= 0 \\ &\Downarrow \\ \text{tr} \left(\Sigma^{-1}\frac{d\Sigma}{d\phi_j} \right) &= \omega_1\epsilon' \left(-\Sigma^{-1}\frac{d\Sigma}{d\phi_j}\Sigma^{-1} \right) \epsilon, \end{aligned}$$

que é a solução apresentada em (2.18).

Apêndice E: Derivadas da matriz Σ

Considere a matriz escala na forma paramétrica dada em (1.6), $\Sigma = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}(\phi_3)$. A seguir apresentamos as derivadas $\frac{d\Sigma}{d\phi_i}$.

$$\frac{d\Sigma}{d\phi_1} = \mathbf{I}_n; \quad \frac{d\Sigma}{d\phi_2} = \mathbf{R}(\phi_3); \quad \frac{d\Sigma}{d\phi_3} = \phi_2 \frac{d\mathbf{R}(\phi_3)}{d\phi_3} = \phi_2 \left[\left(\frac{dr_{ij}(\phi_3)}{d\phi_3} \right) \right];$$

$$\frac{d^2\Sigma}{d\phi_1 d\phi_1} = \frac{d^2\Sigma}{d\phi_1 d\phi_2} = \frac{d^2\Sigma}{d\phi_1 d\phi_3} = \frac{d^2\Sigma}{d\phi_2 d\phi_1} = \frac{d^2\Sigma}{d\phi_2 d\phi_2} = \frac{d^2\Sigma}{d\phi_3 d\phi_1} = \mathbf{0}_n;$$

$$\frac{d^2\Sigma}{d\phi_2 d\phi_3} = \frac{d^2\Sigma}{d\phi_3 d\phi_2} = \left[\left(\frac{dr_{ij}(\phi_3)}{d\phi_3} \right) \right]; \quad \frac{d^2\Sigma}{d\phi_3 d\phi_3} = \left[\left(\frac{d^2 r_{ij}(\phi_3)}{(d\phi_3)^2} \right) \right];$$

A função de correlação do modelo exponencial é $\mathbf{R}(\phi_3) = e^{-h/\phi_3}$, $h \geq 0$. Temos que

$$\frac{dr_{ij}(\phi_3)}{d\phi_3} = e^{-h/\phi_3} \frac{h}{\phi_3^2};$$

$$\frac{d^2 r_{ij}(\phi_3)}{(d\phi_3)^2} = e^{-h/\phi_3} \left[-2 \frac{h}{\phi_3^3} + \left(\frac{h}{\phi_3^2} \right)^2 \right].$$

A função de correlação do modelo Gaussiano é $\mathbf{R}(\phi_3) = e^{-(h/\phi_3)^2}$, $h \geq 0$. Temos que

$$\frac{dr_{ij}(\phi_3)}{d\phi_3} = e^{-(h/\phi_3)^2} \frac{2h}{\phi_3^3};$$

$$\frac{d^2 r_{ij}(\phi_3)}{(d\phi_3)^2} = e^{-(h/\phi_3)^2} \left[-6 \frac{h^2}{\phi_3^4} + \left(\frac{2h}{\phi_3^3} \right)^2 \right].$$

Apêndice F: Derivadas da matriz Δ

Considerando $\delta_\omega = \epsilon'_\omega \Sigma^{-1} \epsilon_\omega$, $\epsilon_\omega = \mathbf{Y} + \boldsymbol{\omega} - \mathbf{X}\boldsymbol{\beta}$ e a perturbação aditiva $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\omega}$, temos as seguintes quantidades

$$\frac{d\delta_\omega}{d\boldsymbol{\omega}'} = \frac{d\delta_\omega}{d\epsilon'_\omega} \frac{d\epsilon_\omega}{d\boldsymbol{\omega}'} = 2\epsilon'_\omega \Sigma^{-1};$$

$$\frac{dl_\theta}{d\boldsymbol{\omega}'} = \frac{dl_\theta}{d\delta'_\omega} \frac{d\delta_\omega}{d\boldsymbol{\omega}'} = 2V_{g(\delta_\omega)} \epsilon'_\omega \Sigma^{-1}.$$

Assim, Δ_β é expressa por

$$\begin{aligned} \Delta_\beta &= \frac{d^2 l_\theta}{d\boldsymbol{\beta} d\boldsymbol{\omega}'} \\ &= \frac{d}{d\boldsymbol{\beta}} [2V_{g(\delta_\omega)} \epsilon'_\omega \Sigma^{-1}] \\ &= 2V_{g(\delta_\omega)} \frac{d\epsilon'_\omega}{d\boldsymbol{\beta}} \Sigma^{-1} + 2 \frac{dV_{g(\delta_\omega)}}{d\boldsymbol{\beta}} \epsilon'_\omega \Sigma^{-1} \\ &= 2 \left[V_{g(\delta_\omega)} (-\mathbf{X}') \Sigma^{-1} + 2V'_{g(\delta_\omega)} (-\mathbf{X}') \Sigma^{-1} \epsilon_\omega \epsilon'_\omega \Sigma^{-1} \right] \\ &= -2\mathbf{X}' \Sigma^{-1} \left[V_{g(\delta_\omega)} \Sigma + 2V'_{g(\delta_\omega)} \Sigma^{-1} \epsilon_\omega \epsilon'_\omega \right] \Sigma^{-1} \end{aligned}$$

em que $\frac{dV_{g(\delta_\omega)}}{d\boldsymbol{\beta}} = \frac{dV_{g(\delta_\omega)}}{d\delta_\omega} \left(\frac{d\delta_\omega}{d\epsilon'_\omega} \frac{d\epsilon_\omega}{d\boldsymbol{\beta}'} \right)' = -2V'_{g(\delta_\omega)} \mathbf{X}' \Sigma^{-1} \epsilon_\omega$.

A componente Δ_{ϕ_i} é expressa por

$$\begin{aligned} \Delta_{\phi_i} &= \frac{d^2 l_\theta}{d\phi_i d\boldsymbol{\omega}'} \\ &= \frac{d}{d\phi_i} [2V_{g(\delta_\omega)} \epsilon'_\omega \Sigma^{-1}] \\ &= 2V_{g(\delta_\omega)} \epsilon'_\omega \frac{d\Sigma^{-1}}{d\phi_i} + 2 \frac{dV_{g(\delta_\omega)}}{d\phi_i} \epsilon'_\omega \Sigma^{-1} \\ &= -2V_{g(\delta_\omega)} \epsilon'_\omega d\Sigma^{-1} \frac{d\Sigma}{d\phi_i} d\Sigma^{-1} - 2V'_{g(\delta_\omega)} \epsilon'_\omega d\Sigma^{-1} \frac{d\Sigma}{d\phi_i} d\Sigma^{-1} \epsilon'_\omega d\Sigma^{-1} \\ &= -2\epsilon' \Sigma^{-1} \frac{d\Sigma}{d\phi_i} \Sigma^{-1} \left(V'_{g(\delta_\omega)} \epsilon \epsilon' \Sigma^{-1} + V_{g(\delta_\omega)} \mathbf{I}_n \right) \end{aligned}$$

em que $\frac{dV_{g(\delta_\omega)}}{d\phi_i} = \frac{dV_{g(\delta_\omega)}}{d\delta_\omega} \frac{d\delta'_\omega}{d\phi_i} = -V'_{g(\delta_\omega)} \epsilon'_\omega d\Sigma^{-1} \frac{d\Sigma}{d\phi_i} d\Sigma^{-1}$.