MÉTODOS PARA INCORPORAÇÃO DA DEPENDÊNCIA ESPACIAL NA ANÁLISE DE DADOS EXPERIMENTAIS

ANA ALICE PILON

Licenciada em Matemática

Orientadora: Prof² Dr² SÔNIA MARIA DE STEFANO PIEDADE

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, para obtenção do título de Mestre em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

PIRACICABA São Paulo - Brasil Agosto - 2004

Dados Internacionais de Catalogação na Publicação (CIP) DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Pilon, Ana Alice

Método para incorporação da dependência espacial na análise de dados experimentais / Ana Alice Pilon. - - Piracicaba, 2004. 128 p.

Dissertação (mestrado) - Escola Superior de Agricultura Luiz de Queiroz, 2004. Bibliografia.

1. Análise espacial 2. Estimação paramétrica 3. Geoestatística 4. Métodos MCMC 5. Verossimilhança I. Título

CDD 519.5

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte - O autor"

DEDICATÓRIA

A

Deus

Aos meus pais
João Pilon e
Joana Pimentel Pilon,
pelo amor incondicional
e confiança, que norteiam
minha vida e me fizeram progredir.

Aos meus irmãos, sobrinhas, cunhadas, amiga Jucelene e tia Silvia, o carinho e incentivo.

AGRADECIMENTOS

À Prof^a Dr^a Sônia Maria De Stefano Piedade a orientação, amizade e crédito em meu trabalho.

Ao Prof^o Dr^o Paulo Justiano Ribeiro Junior do Departamento de Estatística da Universidade Federal do Paraná a paciência, os ensinamentos, as sugestões, a amizade e apoio, que foram fundamentais para a realização deste trabalho.

À Prof
ª Roseli Aparecida Leandro a amizade, o apoio, discussões e sugestões ao meu trabalho.

Ao Prof^o Dr^o Décio Barbin ao carinho e sugestões em meu trabalho.

Aos amigos do mestrado e doutorado, em especial à Juliana Garcia Céspedes, Antônio Willians Moita, Adriano Ferreti Borgato, Carlos Henrique Garcia, Maria Cristina Neves de Oliveira, Melissa Oda e Denise Viola, e às amigas Natali Fortinolli, Luciane Chiodi, Claudia Agnolini e Fabiana Vizzaccaro o apoio, companherismo e compreensão.

Aos professores e funcionários do Departamento de Ciências Exatas da ESALQ/USP, em especial a $Prof^{\underline{a}}$ $Dr^{\underline{a}}$ Clarice Garcia Borges Demétrio, que me propiciaram condições para a realização deste trabalho.

À CAPES o apoio financeiro.

Aos meus professores de graduação da Universidade Estadual Paulista, em especial o Profº Drº Antônio Carlos Simões Pião, o incentivo.

 ${\bf E}$ a todos que de forma direta ou indireta contribuíram para a realização deste trabalho.

SUMÁRIO

	Página
LISTA	DE FIGURAS viii
LISTA	DE TABELAS
RESUN	1O
	ARY
1	INTRODUÇÃO 1
2	REVISÃO DE LITERATURA
2.1	Modelos Geoestatísticos
2.1.1	Funções de Correlação
2.1.2	Semivariograma e Covariância
2.1.2.1	Características dos parâmetros que descrevem um semivariograma 15
2.1.2.2	Modelos teóricos de semivariogramas
2.1.2.3	Alguns métodos de estimação dos parâmetros do semivariograma 17
2.2	Inferência Bayesiana
2.2.1	Monte Carlo com Cadeia de Markov
2.3	Modelo Markoviano
2.3.1	Escolha das distribuições a priori
2.3.2	A distribuição a priori baseada em campos aleatórios Markovianos 24
3	MATERIAL E MÉTODOS
3.1	Material
3.2	Métodos
3.2.1	Modelo de Gauss-Markov Normal
3.2.2	Modelagem de Papadakis (PP)

		vi
3.2.3	Modelagem geoestatística	35
3.2.3.1	Método de estimação do Variograma seguido de Mínimos Quadrados	
	Generalizados	36
3.2.3.2	Estimação de Máxima Verossimilhança	41
3.2.3.3	Estimação de Máxima Verossimilhança Restrita	43
3.2.4	A modelagem com estrutura Markoviana	44
3.2.5	Experimento I	46
3.2.6	Experimento II	47
3.2.7	Simulação I	48
3.2.8	Simulação II	49
3.2.9	Simulação III	50
3.2.10	Medidas utilizadas na comparação dos métodos de estimação dos con-	
	trastes entre os tratamentos	50
3.2.10.1	Diferença absoluta	51
3.2.10.2	Medida de dispersão	52
3.2.10.3	Erros Quadráticos calculados analiticamente	52
3.2.10.4	Erros Quadráticos obtidos por simulação de Monte Carlo	52
3.2.10.5	Intervalo de confiança e Intervalo de cobertura	53
4	RESULTADOS E DISCUSSÃO	54
4.1	Resultados do Experimento I	56
4.1.1	Ajuste do modelo geoestatístico	56
4.1.2	Ajuste do modelo Markoviano	58
4.1.3	Comparações entre os modelos ajustados	59
4.2	Resultados do Experimento II	61
4.2.1	Ajuste do modelo geoestatístico	61
4.2.2	Ajuste do modelo Markoviano	63
4.2.3	Comparações entre os modelos ajustados	64
4.3	Resultados da Simulação I	65
4.3.1	Simulação Ia: com baixa dependência espacial	68

		vii
4.3.2	Simulação Ib: média dependência espacial	70
4.3.3	Simulação Ic: com alta dependência espacial	72
4.3.4	Análise freqüentista versus Bayesiana - Simulação I	74
4.3.4.1	Análise freqüentista versus Bayesiana - Simulação Ia 	7 5
4.3.4.2	Análise freqüentista versus Bayesiana - Simulação Ib $\ \ldots \ \ldots \ \ldots$	76
4.3.4.3	Análise freqüentista versus Bayesiana - Simulação Ic $\ \ldots \ \ldots \ \ldots$	77
4.4	Resultados da Simulação II	78
4.4.1	Simulação IIa	80
4.4.2	Simulação IIb	82
4.4.3	Simulação IIc	84
4.4.4	Análise freqüentista versus Bayesiana - Simulação II $\ \ldots \ \ldots \ \ldots$	86
4.4.4.1	Análise freqüentista versus Bayesiana - Simulação IIa $\ \ldots \ \ldots \ \ldots$	87
4.4.4.2	Análise freqüentista versus Bayesiana - Simulação IIb $\ \ \ldots \ \ldots \ \ \ldots$	88
4.4.4.3	Análise freqüentista versus Bayesiana - Simulação IIc	89
4.5	Resultados da Simulação III	90
5	CONCLUSÕES	94
ANEX	KOS	95
REFE	ERÊNCIAS BIBLIOGRÁFICAS	100
APÊN	NDICES	106

LISTA DE FIGURAS

	Pág	ina
1	Semivariograma apresentando variância nugget diferente de zero	15
2	Variograma empírico, em que cada ponto representa o valor da semi-	
	variância de $N(u)$ pares de parcelas	38
3	Ajuste do variograma teórico	39
4	Distribuição a priori gama para o hiperparâmetro λ_S	45
5	Estrutura de vizinhança de segunda ordem	45
6	Dados sem os efeitos dos tratamentos	56
7	Dados com a adição dos efeitos dos tratamentos	56
8	Ajustes dos variogramas teóricos aos valores do variograma empírico	57
9	Ajuste dos variogramas teóricos aos valores do variograma empírico	62

LISTA DE TABELAS

	Pag	ına
1	Distribuições consideradas para os parâmetros e hiperparâmetros do mo-	
	delo Markoviano	45
2	Efeitos atribuídos aos tratamentos-Experimento I	46
3	Efeitos atribuídos aos tratamentos-Experimento II	47
4	Estimativas dos parâmetros de covariância, $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2)$, obtidas pelos	
	métodos de estimação de parâmetros do modelo geoestatístico utilizando	
	diferentes modelos de funções de correlação, soma dos mínimos quadrados	
	ponderados (W_p) , logaritmo da função de verossimilhança (Log) e medida	
	de dispersão (D)	58
5	Valores verdadeiros dos contrastes entre os tratamentos (VERD.), esti-	
	mativas desses contrastes obtidas por cada método adotado e medida de	
	dispersão (D)	59
6	Diferenças absolutas entre as estimativas dos contrastes e seus valores	
	verdadeiros	60
7	Erros quadráticos das estimativas dos contrastes calculados pelos métodos	
	de estimação de parâmetros, tratados segundo a abordagem freqüentista	6 0
8	Estimativas dos parâmetros de covariância, $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2)$, obtidas pelos	
	métodos de estimação de parâmetros do modelo geoestatístico utilizando	
	diferentes modelos de funções de correlação, soma dos mínimos quadrados	
	ponderados (W_p) , logaritmo da função de verossimilhança (Log) e medida	
	de dispersão (D)	63

		x
9	Valores verdadeiros dos contrastes entre os tratamentos (VERD.), esti-	
	mativas desses contrastes obtidas pelos métodos adotados e medida de	
	dispersão (D)	64
10	Diferenças absolutas entre os contrastes verdadeiros e a média de seus	
	valores estimados	65
11	Erros quadráticos das estimativas dos contrastes calculados pelos métodos	
	de estimação de parâmetros, tratados segundo a abordagem frequentista	65
12	Médias das estimativas dos contrastes resultantes da simulação com baixa	
	dependência espacial	68
13	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com baixa dependência espacial	68
14	Médias e desvios padrões da medida de dispersão (D), resultantes da	
	simulação com baixa dependência espacial	68
15	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente, da simulação com baixa dependência espacial	69
16	Erros quadráticos das estimativas dos contrastes obtidos por simulação	
	de Monte Carlo da simulação com baixa dependência espacial	69
17	Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados	
	por (54), resultantes da simulação com baixa dependência espacial	69
18	Médias das estimativas dos contrastes resultantes da simulação com média	
	dependência espacial	70
19	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com média dependência espacial	70
20	Médias e desvios padrões da medida de dispersão (D), resultantes da	
	simulação com média dependência espacial	70
21	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente, da simulação com média dependência espacial	71
22	Erros quadráticos das estimativas dos contrastes obtidos por simulação	
	de Monte Carlo, simulação com média dependência espacial	71

.

23	Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados	
	por (54), resultantes da simulação com média dependência espacial	71
24	Médias das estimativas dos contrastes resultante da simulação com alta	
	dependência espacial	72
25	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com alta dependência espacial	72
26	Médias e desvios padrões da medida de dispersão (D), resultantes da	
	simulação com alta dependência espacial	72
27	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente, da simulação com alta dependência espacial	73
28	Erros quadráticos das estimativas dos contrastes obtidos por simulação	
	de Monte Carlo da simulação com alta dependência espacial	73
29	Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados	
	por (54), resultantes da simulação com alta dependência espacial	73
30	Estimativas dos contrastes resultantes da simulação com baixa de-	
	pendência espacial	75
31	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com baixa dependência espacial	75
32	Medidas de dispersão (D), resultantes da simulação com baixa de-	
	pendência espacial	75
33	Estimativas dos contrastes resultantes da simulação com média de-	
	pendência espacial	76
34	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com média dependência espacial	76
35	Medidas de dispersão (D), resultantes da simulação com média de-	
	pendência espacial	76
36	Estimativas dos contrastes resultantes da simulação com alta dependência	
	espacial	77

37	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com alta dependência espacial	77
38	Medidas de dispersão (D), resultantes da simulação de dados com alta	
	dependência espacial	77
39	Médias das estimativas dos contrastes resultantes da simulação com baixa	
	dependência espacial	80
40	Diferenças absolutas entre os valores verdadeiros e as estimativas da si-	
	mulação com baixa dependência espacial	80
41	Médias e desvios padrões da medida de dispersão (D), resultantes da	
	simulação com baixa dependência espacial	80
42	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente, da simulação com baixa dependência espacial	81
43	Erros quadráticos das estimativas dos contrastes obtidos por simulação	
	de Monte Carlo da simulação com baixa dependência espacial	81
44	Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados	
	por (54), resultantes da simulação com baixa dependência espacial	81
45	Médias das estimativas dos contrastes resultantes da simulação com média	
	dependência espacial	82
46	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com média dependência espacial	82
47	Médias e desvios padrões da medida de dispersão (D), resultantes da	
	simulação com média dependência espacial	82
48	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente, da simulação com média dependência espacial	83
49	Erros quadráticos das estimativas dos contrastes obtidos por simulação	
	de Monte Carlo na simulação com média dependência espacial	83
50	Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados	
	por (54), resultantes da simulação com média dependência espacial	83

		xiii
51	Médias das estimativas dos contrastes resultantes da simulação com alta	
	dependência espacial	84
52	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, resultantes da simulação com alta dependência espacial	84
53	Médias e desvios padrões da medida de dispersão (D), resultantes da	
	simulação com alta dependência espacial	84
54	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente, da simulação com alta dependência espacial	85
55	Erros quadráticos das estimativas dos contrastes obtidos por simulação	
	de Monte Carlo na simulação com alta dependência espacial	85
56	Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados	
	por (54), resultantes da simulação com alta dependência espacial	85
57	Estimativas dos contrastes obtidas da simulação com baixo grau de de-	
	pendência espacial	87
58	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, obtidas da simulação com baixo grau de dependência espacial	87
59	Medidas de dispersão (D), obtidas da simulação com baixo grau de de-	
	pendência espacial	87
60	Estimativas dos contrastes obtidas da simulação com médio grau de de-	
	pendência espacial	88
61	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, obtidas da simulação com médio grau de dependência espacial	88
62	Medidas de dispersão (D), obtidas da simulação com médio grau de de-	
	pendência espacial	88
63	Estimativas dos contrastes obtidas da simulação com alto grau de de-	
	pendência espacial	89
64	Diferenças absolutas entre os valores verdadeiros e as estimativas dos	
	contrastes, obtidas da simulação com alto grau de dependência espacial.	89

		xiv
65	Medidas de dispersão (D), obtidas da simulação com alto grau de de-	
	pendência espacial	89
66	Médias das estimativas dos contrastes entre os tratamentos	91
67	Diferenças absolutas entre os valores verdadeiros e a média das estimativas	
	dos contrastes	91
68	Médias e desvios padrões da medida de dispersão (D)	91
69	Médias dos erros quadráticos das estimativas dos contrastes calculados	
	analiticamente	92
70	Erros quadráticos das estimativas dos contrastes calculados por simulação	
	de Monte Carlo	92
71	Intervalos de cobertura de cada contraste e do vetor dos contrastes $$	92

MÉTODOS PARA INCORPORAÇÃO DA DEPENDÊNCIA ESPACIAL NA ANÁLISE DE DADOS EXPERIMENTAIS

Autora: ANA ALICE PILON

Orientadora: Prof^a Dr^a SÔNIA MARIA DE STEFANO PIEDADE

RESUMO

A influência da heterogeneidade do solo na estimativa das diferenças entre os tratamentos é uma preocupação para os pesquisadores de campo. Na busca de alternativas para controlar esse fator e aumentar a precisão das comparações entre os tratamentos, o presente trabalho comparou a análise clássica ANOVA e alguns modelos espaciais que consideram a dependência no espaço entre as parcelas, levando em consideração suas posições geográficas na área experimental. Os modelos espaciais considerados no estudo foram: o modelo proposto por Papadakis, que consiste na adição de uma covariável construída a partir de observações em parcelas vizinhas, o modelo geoestatístico que consiste na adição de um efeito aleatório com correlação, dada por uma função da distância entre parcelas, e o modelo Markoviano que adiciona um efeito aleatório obtido a partir de observações em parcelas vizinhas. Os modelos foram ajustados a dados provenientes de dois ensaios em branco

obtidos da literatura, aos quais foram adicionados efeitos arbitrários de tratamentos segundo delineamento inteiramente casualizado. Os efeitos dos tratamentos foram arbitrariamente adicionados a fim de simular um experimento real com o propósito de demonstração, para comparar a análise clássica (ANOVA) e os modelos espaciais. Nestas análises foi verificado que para os dados do ensaio em branco com 25 parcelas o modelo Markoviano forneceu maior precisão às estimativas dos contrastes entre os tratamentos, com grande diferenca em relação aos resultados obtidos pelos demais modelos. Para o ensaio em branco, com 225 parcelas, o modelo Markoviano também foi o mais eficiente, porém, seus resultados foram próximos aos obtidos pelos outros modelos. O modelo geoestatístico junto aos métodos de estimação de máxima verossimilhanca e máxima verossimilhanca restrita foi o segundo melhor entre os modelos espaciais. Para a investigação dos resultados obtidos nos ajustes dos modelos aos dados dos ensaios em branco, foram realizadas simulações com alta, média e baixa dependência espacial entre as parcelas em condições experimentais similares aos dois conjuntos de dados investigados. Os resultados indicaram que os modelos espaciais oferecem uma maior precisão às estimativas dos contrastes entre as médias dos tratamentos, quando comparados à análise clássica, e o modelo geoestatístico junto aos métodos de estimação de parâmetros de máxima verossimilhança e de máxima verossimilhança restrita, foi na maioria das vezes o mais eficiente. O modelo Markoviano, em alguns casos, apresentou resultados mais precisos do que o geoestatístico, porém não foi suficientemente explorado nesse estudo sendo ainda mais trabalhoso do ponto de vista computacional. A fim de investigar se os efeitos dos tratamentos podem ser confundidos com a variação local, foi realizada uma simulação com 5000 diferentes aleatorizações dos efeitos de tratamentos aos dados do ensaio em branco com 25 parcelas, cujos resultados indicaram que os modelos espaciais oferecem estimativas mais precisas aos contrastes entre os tratamentos em comparação à análise clássica ANOVA (i.i.d.).

METHODS TO INCORPORATION OF SPATIAL DEPENDENCY IN ANALYSIS OF EXPERIMENTS DATAS.

Author: ANA ALICE PILON

Adviser: Prof² Dr² SÔNIA MARIA DE STEFANO PIEDADE

SUMMARY

Soil heterogeneity's influence on the estimation of the treatment differences is one of the researcher's concerns in this field. The aim of this work is to find alternatives methods to control this problem. This work compares classical ANOVA to some other spatial models which consider the spatial dependence between plots according to their geographical position on the experimental area. Three spatial models were considered: firstly, the model proposed by Papadakis, which adds a covariate derived from observations in neighbours' plots. Secondly, a geostatistical model which includes a random effect with correlation given by a function of the distance between plots, and finally, a Markov model which includes a random effect derived from observations in neighbours' plots. The models were fit to two set of real data, derived from uniformity trials with treatments arbitrary effects added randomly. Treatments arbitrary effects were added to the real data to simulate a real

experiment for the purposes demonstration, and to evaluate the classical ANOVA and spatial methods. In these analysis, the Markov model produced more precise estimates of the contrast between treatments, within a dataset of 25 plots. The difference between the models decreased as the dataset size increased, and for a dataset consisting of 225 plots the results were very similar. To investigate the results obtained from the models, data with three different levels of spatial dependance between plots were simulated, reproducing the experimental conditions of the two uniformity trials. The results indicated that spatial models are more precise in the estimation of the treatment effects when compared to the classical ANOVA, and in particular most of the times the geostatistical model together with maximum likelihood and restricted maximum likelihood estimation was the most efficient. The Markov model provided more precise results in some cases, but it was not deeply investigated in this study because it is too computationally demanding. In order to exclude a possible confounding effect due to a local variation, a simulation was carried with 5000 randomizations, and the results indicated that spatial models provided more precise estimates of the treatment contrasts when compare to the classical ANOVA.

1 INTRODUÇÃO

O maior desafio quando se conduz um experimento é o de comparar os tratamentos com a maior precisão possível, para se ter segurança nas inferências a serem feitas a partir dos resultados alcançados.

A precisão de um experimento está totalmente ligada a pequenas variações nas unidades experimentais, antes de serem aplicados os tratamentos, ou induzidas involuntariamente, durante a execução do experimento, que causam heterogeneidade entre as parcelas, também conhecida como variação do acaso, variação ambiental ou, simplesmente, erro experimental (Storck et al., 2000).

Quando o erro experimental é muito grande, os testes estatísticos usados para comparar as médias de tratamentos podem ser influenciados, impedindo que algumas diferenças entre médias de tratamentos sejam detectadas. Por exemplo Souza (1997) menciona que para se continuar obtendo sucesso em programas de melhoramento genético, o desafio é cada vez maior, pois as diferenças entre os materiais genéticos são gradativamente menores. Isso mostra que a experimentação agrícola requer a utilização de técnicas experimentais cada vez mais refinadas. Como técnica experimental estão envolvidos diversos fatores, tais como: escolha criteriosa do delineamento, locais de instalação dos experimentos, número de repetições, tamanho e forma da parcela e uma perfeita condução dos experimentos. Todos esses fatores irão atuar de forma direta ou indireta na precisão experimental.

A análise clássica baseia-se em três princípios básicos, repetição, casualização e controle local, estabelecidos por Fisher ¹ (1935).

¹Mencionado por Ver Hoef & Cressie (1993).

Conforme Barbin (2003), o controle local consiste na subdivisão da área experimental heterogênea, em áreas menores e homogêneas e em cada uma deve-se colocar todos os tratamentos, de preferência em igual número, dessa forma caracterizam-se os chamados blocos completos. Brownie et al. (1993), afirmam que a finalidade da blocagem é reduzir a variação residual e assim aumentar a precisão experimental, pois ela garante que os tratamentos sejam avaliados em condições ambientais similares dentro de um mesmo bloco.

O conceito de repetição, segundo Barbin (2003), incide em se terem várias parcelas com o mesmo tratamento, com o objetivo de confirmar a resposta dos indivíduos a um determinado tratamento. De acordo com Ramalho et al. (2000), o uso de um número apropriado de repetições contribui para aumentar a precisão das estimativas dos parâmetros e, ao mesmo tempo, para aumentar o poder dos testes estatísticos.

De acordo com Mead & Curnow (1983), a casualização deve ser usada na alocação dos tratamentos às parcelas e tem por finalidade propiciar a todos os tratamentos a mesma probabilidade de serem designados a qualquer uma das unidades experimentais. Com isso, evita-se que qualquer tratamento seja sistematicamente favorecido e garante que os erros sejam independentes, conseqüentemente, a análise de variância é válida. Como a casualização pode reduzir a correlação espacial entre parcelas adjacentes (Grondona & Cressie, 1991), as posições das unidades experimentais no mapa de campo são normalmente ignoradas na abordagem tradicional.

Com o desenvolvimento computacional, outras alternativas tornaramse disponíveis aos experimentadores, como por exemplo os métodos de análise espacial de experimentos, que em certas situações podem proporcionar resultados satisfatórios. Esses métodos se baseiam na análise de vizinhança, ou seja, nas parcelas vizinhas ou modelam a dependência espacial em função da distância entre as parcelas, visando o controle da heterogeneidade espacial.

Conforme Costa (2003), as formas de análises baseadas em modelagem espacial são mais apropriadas para experimentos em que é detectada dependência

espacial entre os erros experimentais, pois, nestes casos, as eficiências dos estimadores dos contrastes dos tratamentos não dependerão apenas da variação residual, mas também das posições das parcelas.

Em muitas aplicações estatísticas existem problemas inerentemente espaciais, em que os estudos são freqüentemente observáveis ao invés de delineados, tornando-se impossível a aplicação dos procedimentos exigidos à uma análise clássica de qualidade. Existem ainda outras barreiras como, custos elevados, dificuldades para executar repetições, heterogeneidade das áreas, ocorrências de erros de medidas em experimentos realizados no campo, que são freqüentemente enfrentadas por algumas áreas como geologia, ciências do solo, epidemiologia, florestal entre outras (Ver Hoef & Cressie, 1993).

Cressie (1993) mostra os efeitos da dependência espacial em experimentos clássicos. E argumenta que a independência é uma suposição que facilita a teoria da estatística matemática, contudo, os modelos que envolvem dependência são freqüentemente mais realísticos. Pois, mesmo que realizado o princípio da casualização, a noção de que dados mais próximos, no tempo ou no espaço, sejam correlacionados é natural. O autor assegura, que a detecção da estrutura de autocorrelação e o uso dessa informação na análise estatística, garantem estimativas mais eficientes dos contrastes entre as médias dos tratamentos, por outro lado, a desconsideração dessa estrutura pode impedir que diferenças reais sejam detectadas.

Conforme Ribeiro Junior (1995), a dependência espacial não é uma inconveniência estatística, mas um verdadeiro benefício que pode informar sobre locais não amostrados a partir de dados tomados em posições próximas aos pontos desejados.

Um método muito utilizado em modelagem espacial é o dos Vizinhos - Próximos, o qual analisa experimentos de campo examinando a dependência espacial usando informações observadas nas parcelas vizinhas (Papadakis, 1937; Besag & Kempton, 1986).

Neste estudo, serão abordadas três estratégias de análise espacial, a

análise de vizinhança segundo Papadakis (1937), a análise geoestatística e a análise Markoviana. Cada tipo de análise possui uma particularidade na forma de considerar a dependência espacial.

Na análise de vizinhança segundo Papadakis (1937), a dependência espacial é modelada por uma análise de covariância, utilizando-se como covariável as médias dos resíduos entre parcelas vizinhas. A análise geoestatística considera a distância entre as unidades experimentais para modelar a dependência espacial através de um componente aleatório de erros dependentes. E finalmente, a análise Markoviana que assim como o modelo proposto por Papadakis (1937), baseia-se numa análise de vizinhança, mas modela a dependência espacial através de um componente de erros aleatórios dependentes, com distribuição a priori dada por um modelo CAR (Condicionalmente Auto Regressivo), essa análise será tratada segundo a teoria Bayesiana.

O presente trabalho pretende fazer comparações entre modelos e métodos de estimação de parâmetros que consideram a presença e a importância da informação espacial na análise de dados experimentais. Os modelos em questão são: i) a análise de vizinhança proposta por Papadakis; ii) o modelo geoestatístico e iii) o modelo de estrutura Markoviana. Serão comparados três métodos de estimação do modelo geoestatístico, o método do variograma seguido de mínimos quadrados generalizados, a estimação de máxima verossimilhança e a estimação de máxima verossimilhança restrita. O modelo Markoviano será tratado segundo a teoria Bayesiana e seus parâmetros serão estimados utilizando-se amostras construídas por métodos Monte Carlo com Cadeia de Markov (MCMC).

2 REVISÃO DE LITERATURA

A preocupação com a influência da heterogeneidade do solo na estimativa das diferenças entre os tratamentos na experimentação de campo, além dos blocos, exigiu o surgimento de novas abordagens no planejamento e na análise de experimentos.

A distribuição dos resíduos de uma análise estatística, fornece indicações de probabilidades sobre as estimativas dos efeitos de tratamentos ou combinações de efeitos de tratamentos ($\mu + \tau_1$ ou $\tau_1 - \tau_2$). Para uma análise de variância clássica (ANOVA), todo erro experimental é assumido normalmente distribuído, com esperança zero, e variância constante σ^2 e a aleatoriedade em um delineamento experimental permite considerar os erros independentes. A distribuição aleatória dos tratamentos produz erros experimentais aleatórios, que possuem pequenas autocorrelações e assim permitem o uso da teoria clássica, Grondona & Cressie (1991), como confirmado por muitas simulações (Besag & Kempton, 1986; Baird & Mead, 1991; Zimmerman & Harville, 1991;). Porém, se ignoradas todas as possíveis aleatoriedades, admitindo-se somente o que ocorreu, a variação natural nas parcelas faz com que os erros experimentais sejam autocorrelacionados. O problema então é como modelar a autocorrelação e subseqüentemente usar essa informação para obtenção da melhor estimativa paramétrica.

Existe uma vasta literatura sobre alternativas de análises estatísticas que consideram a posição espacial das amostras, entre elas podem ser citadas, por exemplo, a análise geoestatística utilizando estimadores de parâmetros como o variograma seguido de mínimos quadrados generalizados (Cressie (1993); Ver Hoef

& Cressie (1993)), estimação de máxima verossimilhança (Cressie (1993); Zimmerman & Harville (1991)), estimação de máxima verossimilhança restrita (Cressie 1993)), a análise de vizinhança proposta por Papadakis e o método de médias móveis (Papadakis (1997); Vivaldi (1990); Wilkinson et al. (1983); Costa (2003), Cargnelutti Filho (2002)), além dos modelos baseados em campos aleatórios de Markov (Besag & Higdon (1999)).

A maioria dos métodos espaciais foram desenvolvidos extrapolando-se os conceitos de séries temporais e geoestatística para a experimentação agrícola, os quais foram a base para a construção dos modelos de tendência no solo e dos erros correlacionados entre as parcelas vizinhas que são as principais fontes de heterogeneidade do solo (Souza, 1997).

Papadakis (1937) foi o pioneiro a utilizar o método do vizinho mais próximo (Nearest-Neighbor Method-NN) com a finalidade de reduzir o erro experimental. O conhecido método de Papadakis, consiste em corrigir a produção de cada parcela pelo efeito médio do tratamento aplicado à parcela, e então, a média dos valores corrigidos das parcelas vizinhas é utilizada como covariável, seguindose com uma análise de covariância. Procura-se dessa forma, eliminar grande parte das tendências existentes no solo, através do desempenho das parcelas vizinhas, com uma conseqüente diminuição no erro experimental. Algumas pesquisas teóricas foram conduzidas por Atkinson (1969), Bartlett (1938, 1978), Wilkinson et al. (1983). Bartlett (1938, 1978) mostrou em seus trabalhos que o método é aproximado, contudo, pode ser satisfatório quando o número de parcelas por bloco é grande.

Wilkinson et al. (1983) a partir de dados simulados mostraram que o método de Papadakis é tendencioso e propuseram uma alternativa para análise de vizinhança denominada de método blocos móveis, o qual consiste na correção de cada parcela pela média das parcelas vizinhas, sendo que o bloco caminha ou se move até o fim do experimento. O método de análise e eficiência do modelo, contudo, está associado a um balanceamento com relação aos vizinhos, exigindo assim, um delineamento sistemático ou semi-sistemático.

Costa (2003) comparou três métodos espaciais em busca de melhorar a eficiência dos experimentos de avaliação de progênies de feijão e milho. Os resultados obtidos indicaram que os métodos de Médias Móveis e Papadakis em muitos casos melhoraram o modelo original de Gauss Markov normal justificado pelo delineamento aplicado, embora a Análise Espacial que modela covariâncias residuais não o tenha melhorado.

Também a partir da análise de experimentos simulados, Vivaldi (1990) constatou que o método de Papadakis foi o mais eficiente na redução do erro experimental, nas condições estudadas. Segundo o autor, os métodos de médias móveis e blocos móveis ainda necessitam de estudos suplementares, através da modificação dos pesos dados às parcelas ou pela obtenção do número de parcelas envolvidas na criação de covariáveis.

Para se usar o método de Papadakis é necessário que seja determinado uma estrutura de vizinhança. Cargnelutti Filho (2003) avaliou diferentes estruturas de vizinhança no cálculo do índice de produtividade usado como covariável pelo método de Papadakis. Os resultados mostraram que a vizinhança mais próxima, ou seja, a unidade experimental de referência e as quatro laterais, é a forma mais eficiente na redução do erro experimental.

Segundo Eiseberg (1996), nos experimentos agrícolas, as medidas realizadas sobre variáveis, como por exemplo, produtividade de grãos, estão sujeitas a muitas influências que podem aumentar o ruído contido na informação, entre as quais podem ser citadas: variabilidade do solo não expressa no estágio de planejamento, ataques de insetos em direções preferenciais, desenvolvimento de doenças a partir de focos ou reboleiras, entre outras. Assim, mesmo que se tomem os cuidados experimentais rotineiros, variáveis não controladas podem aparecer e afetar o ensaio a ponto dos efeitos de tratamentos serem tomados com um baixo grau de confiança. De acordo com Cressie (1993), métodos espaciais podem recuperar parte da eficiência perdida, quando um método, não espacial, ineficiente foi utilizado, ou um experimento bem planejado foi prejudicado por alguma variável não esperada.

Uma maneira de resolver esses problemas é utilizar uma aproximação espacial, cuja estratégia consiste em estimar e modelar a correlação a partir dos resíduos de uma análise clássica e usar a teoria dos mínimos quadrados generalizados para estimar os efeitos dos tratamentos, esse método é conhecido como variograma seguido de mínimos quadrados generalizados.

De acordo com o trabalho de Zimmerman & Harville (1991), a maioria dos métodos de análise de vizinhança procuram remover os efeitos ambientais de forma indireta, ou seja, a partir do uso de covariáveis obtidas pela diferença ou pelos resíduos das parcelas vizinhas. Estes autores propuseram um método, onde os efeitos de tendência e as correlações entre os erros foram incluídos diretamente no modelo, e utilizaram como método de estimação dos parâmetros, a máxima verossimilhança. Os resultados indicaram significante redução na estimativa da variância.

O método de máxima verossimilhança consiste em maximizar a função de densidade de probabilidades, em relação aos efeitos fixos e aos componentes de variâncias dos efeitos aleatórios do modelo. Esse método é iterativo e fornece sempre estimativas positivas de componentes de variância. Existem alguns problemas no uso deste estimador devido ao fato de não considerar a perda de graus de liberdade resultantes da estimação dos efeitos fixos do modelo. Uma outra alternativa que pode solucionar esse problema é o método de máxima verossimilhança restrita, no qual, combinações lineares especiais dos dados, chamados de incrementos generalizados, são usados no lugar das observações originais. Estes incrementos generalizados filtram as tendências e estimam somente os parâmetros de covariância, levando a um estimador não viesado.

Por isso, nos ensaios de campo na agricultura e na pesquisa com melhoramento animal, como método de seleção, a estimação de máxima verossimilhança restrita dos parâmetros de covariância espacial, tem sido preferida à de máxima verossimilhança (Cressie & Lahiri, 1996; Kitanidis & Shen, 1996).

Ainda, foram realizadas por Cressie (1993) comparações entre alguns métodos espaciais e experimentos clássicos. O autor apresentou vários enfoques de

dados espaciais e diferentes métodos de estimação de parâmetros, como por exemplo, o da máxima verossimilhança e do variograma seguido de mínimos quadrados generalizados. Afirmou que os delineamentos em blocos completos casualizados, muito utilizados em pesquisas com fertilidade do solo, são preferidos devido à facilidade de construção, porém são ineficientes para avaliar a variabilidade espacial, principalmente quando a estrutura de correlação é maior ou igual a 0,5.

Ver Hoef & Cressie (1993) compararam a análise espacial através do método do variograma seguido de mínimos quadrados generalizados e a análise clássica ANOVA em um experimento inteiramente casualizado. Os dados consistem no número de diferentes espécies de plantas em uma determinada área e os autores atribuíram valores aos efeitos dos tratamentos ao invés de aplicá-los diretamente na área, procederam dessa forma, para que fosse possível conhecer os verdadeiros valores dos contrastes entre os tratamentos. Os resultados obtidos indicaram que com o método espacial a estimativa da variância teve uma redução de 58% em relação à análise clássica.

Grondona & Cressie (1991), em um experimento em blocos casualizados, utilizaram como estimador dos parâmetros um semivariograma robusto (Cressie & Hawkins, 1980) e também obtiveram como resultados uma redução de 75 % na estimativa da variância em relação à análise clássica.

Outros métodos espaciais foram estudados por: Cullins & Gleeson (1989; 1991), Baird & Mead (1991), Besag & Kempton (1986).

O enfoque Bayesiano em experimentos de campo na agricultura tem recebido pouca atenção, embora, exista uma vasta literatura freqüentista como mencionado por Besag & Higdon (1999). Os autores também afirmaram que a adaptação do paradigma Bayesiano simplifica a interpretação dos resultados, especialmente no que se diz respeito a posição e seleção de tratamentos ou variedades. Ressaltaram que, atualmente, alguns modelos complexos podem ser analisados com certa facilidade através dos algoritmos Monte Carlo Cadeia de Markov. Essas conclusões foram baseadas nos resultados de algumas análises em que foram avaliadas produções de di-

ferentes variedades de trigo, morango e milho. Para isso, foram utilizadas as análises Bayesiana: Gaussiana e t-Hierárquica, essa última proposta com objetivo de acomodar os possíveis outliers. E algumas análises freqüentistas como: blocos aleatórios completos, tendência quadrática, blocos incompletos com recuperação interblocos usando a estimação de parâmetros de máxima verossimilhança restrita e a análise geoestatística também usando a estimação de parâmetros de máxima verossimilhança restrita.

2.1 Modelos Geoestatísticos

formato básico para dados geoestatísticos univariados é (x_i,y_i) , i=1,...,n, sendo que x_i identifica a posição espacial tipicamente em um plano bidimensional e y_i é uma medição escalar tomada na posição x_i . A variável de medição y_i pode, em princípio, estar localizada em qualquer lugar numa região de estudo A. No contexto de delineamento experimental, x_i , representa a posição central de uma unidade experimental. Usualmente, é assumido que o delineamento de amostragem para as posições x_i é estocasticamente independente do processo que gera as medições y_i e que a região A é um subconjunto fixo de pontos em \Re^2 . A forma básica de um modelo geoestatístico é um processo estocástico de valor real $\{Y(x):x\in A\}$ o qual é tipicamente considerado como uma realização parcial de um processo estocástico $\{Y(x):x\in\Re^2\}.$ Freqüentemente, o processo de medição Y_i pode ser registrado como uma versão ruidosa de uma variável aleatória subjacente $S(x_i)$, o valor na posição x_i do processo de interesse $\{S(x):x\in A\subset \Re^2\}$, sendo S(x) chamado de sinal. Esse modelo básico pode ser estendido para um modelo com dois componentes: um processo estocástico S(x), e um modelo estatístico para medições, $y=(y_1,...,y_n)^T$ condicional a $\{S(x):x\in A\subset\Re^2\}$. Um processo estocástico é considerado Gaussiano, se a distribuição conjunta de $S(x_1),...,S(x_n)$ é multivariada Gaussiana para qualquer inteiro n, e um conjunto de posições x_i . O processo é estacionário se, a esperança e a variância de S(x) são as mesmas para todo x, e a correlação entre $S(x_i)$ e $S(x_j)$ depende unicamente da distância entre as posições x_i e x_j . O processo é estacionário e isotrópico se, adicionalmente, esta correlação depende unicamente de $u = ||x_i - x_j||$, a distância euclidiana entre x_i e x_j , para qualquer par de inteiros $0 < i, j \le n$. Segundo Diggle & Ribeiro Jr. (2000), o modelo Gaussiano, estacionário e isotrópico para um conjunto de dados (x_i, x_j) , i, j = 1, ..., n, é definido pelas seguintes suposições:

- $\{S(x): x \in A \subset \mathbb{R}^2\}$ é um processo Gaussiano estacionário com média μ , variância σ^2 , função de covariância $cov\{S(x_i), S(x_j)\} = \gamma(x_i, x_j) = \sigma^2 \rho(u)$, e função de correlação $\rho(u) = corr\{S(x_i), S(x_j)\}$, em que $u = ||x_i x_j||$;
- Condicionalmente a $\{S(x): x \in \mathbb{R}^2\}$, os y_i são realizações mutuamente independentes dos Y_i , normalmente distribuídos com média condicional $E[Y_i|S(.)] = S(x_i)$ e variância condicional τ^2 .

Uma forma equivalente de expressar este modelo é dada por

$$Y_i = S(x_i) + Z_i, \quad i = 1, ..., n,$$
 (1)

sendo que Z_i são erros aleatórios mutuamente independentes distribuídos normalmente com média zero e variância τ^2 .

Para especificar esse modelo, são necessários apenas os momentos de primeira e segunda ordem de Y(x). O momento de primeira ordem de uma função aleatória Y(x) é a esperança matemática, definida como:

$$E[Y(x)] = \mu(x). \tag{2}$$

Além deste, existem três momentos de segunda ordem considerados em geoestatística, definidos por Samper & Ramirez (1990) e Braga (1990) os quais são apresentados como:

a) a variância de Y(x) quando existe, pode ser definida como o momento de segunda ordem em relação à média $\mu(x)$.

$$\sigma^2 = \text{Var}[Y(x)] = E[Y(x) - \mu(x)]^2;$$
 (3)

b) covariância - sejam duas variáveis aleatórias $Y(x_i)$ e $Y(x_j)$, nos pontos x_i e x_j , respectivamente, então elas têm covariância como função das locações x_i e x_j e, consequentemente têm variância finita. Assim a expressão da covariância é dada por:

$$C(x_i, x_j) = E[[Y(x_i) - E(Y(x_i))][Y(x_j) - E(Y(x_j))]],$$
(4)

A covariância, dada por (4), deve ser positiva definida para assegurar a existência de todas as distribuições acumuladas. Esta restrição assegura que para qualquer inteiro m, qualquer conjunto de posições x_i e qualquer constante real a_i , a combinação linear $\sum a_i S(x_i)$ terá uma variância não-negativa.

c) semivariograma $\gamma(x_i, x_j)$

$$\gamma(x_i, x_j) = \frac{1}{2} E[(Y(x_i) - Y(x_j))^2]$$
 (5)

O variograma é, portanto, apresentado por $2\gamma(x_i, x_j)$, mas existem autores que usam indistintamente ambas as formas para referir-se à função $\gamma(x_i, x_j)$.

Tanto a variância quanto o semivariograma são sempre positivos enquanto que a covariância (4), pode apresentar valores negativos (Samper & Ramirez, 1990).

Uma outra classe de modelos muito útil na prática, são os processos gaussianos para os quais a média é variável, que podem depender de covariáveis e erros aleatórios, porém, com estrutura de covariância estacionária. Para tais situações, $Y(x) - \mu(x)$ é um processo estacionário Gaussiano com média zero.

2.1.1 Funções de Correlação

É, usualmente, exigido que o modelo de função de correlação $\rho(u)$ incorpore as características dadas por:

i) $\rho(u)$ é monótona não crescente em u (as correlações entre duas medidas decrescem com o aumento da distância entre as duas locações amostrais correspondentes),

- ii) $\rho(u) \longrightarrow 0 \text{ com } u \longrightarrow \infty \text{ e}$
- iii) pelo menos um parâmetro no modelo controla a taxa com que $\rho(u)$ tende a zero.

Alguns exemplos de famílias paramétricas que reúnem essas características e que podem ser assumidas como uma função de correlação, apresentadas por Diggle & Ribeiro Jr. (2000):

1. A Família Exponencial Poder, definida por

$$\rho(\mu) = exp\left\{-\left(\frac{u}{\phi}\right)^{\kappa}\right\},\tag{6}$$

com $\phi > 0$ e $0 < \kappa \le 2$.

O parâmetro ϕ determina a taxa para qual a correlação tende a zero ao se aumentar u. O parâmetro, κ , é chamado de ordem do modelo e determina a suavidade analítica do sinal S(x). A função de correlação exponencial é correspondente ao caso em que $\kappa=1$. O caso que $\kappa=2$ é denominado algumas vezes função de correlação Gaussiana. Essa família freqüentemente ajusta com razoável qualidade, estruturas de correlações empíricas de dados espaciais.

2. A Família Matérn, cuja definição é dada por

$$\rho(u;\phi,\kappa) = \left\{2^{\kappa-1}\Gamma(k)\right\}^{-1} \left(\frac{u}{\phi}\right)^{\kappa} K_{\kappa} \left(\frac{u}{\phi}\right), \tag{7}$$

em que $K_{\kappa}(.)$ denota a função Bessel modificada de terceiro tipo de ordem κ . Os parâmetros ϕ e κ são como descritos nas famílias anteriores. Essa família é a mais recomendada devido a sua flexibilidade.

2.1.2 Semivariograma e Covariância

O semivariograma, dado por (5), é uma função matemática definida para representar o nível de dependência entre duas unidades experimentais e, para exemplificar, pode ser o grau de continuidade das propriedades da fertilidade do solo. Ele é usado para modelar dois valores correlacionados no espaço ou no tempo.

Num estudo em que se deseja averiguar a dependência espacial de amostras georeferenciadas por meio do variograma ou semivariograma têm-se três tipos de situações variográficas, isto é, o variograma observado, o variograma verdadeiro e o variograma teórico (Guerra, 1988).

O variograma observado ou experimental é a primeira informação gráfica que se obtém sobre os dados. É proveniente do conjunto de dados da pesquisa realizada, originados de um processo de amostragem sobre coordenadas geográficas, antes de qualquer ajuste de modelo. Esse variograma é dado pela expressão:

$$\hat{\gamma}(u) = \frac{\sum_{M(u)} (Y(x_i) - Y(x_{i'}))^2}{2N(u)},$$
(8)

em que,

- i) $u = ||x_i x_j||$ é a distância euclidiana entre x_i e x_j .
- ii) $Y(x_i)$ e $Y(x_j)$ são as medidas de interesse nas posições x_i e x_j .
- iii) $M(u) \equiv \{[x_i = (l,c), x_{i'} = (l',c')] : \sqrt{(l-l')^2 + (c-c')^2} = u\}$, o conjunto de todos os pares de posições que estão separadas a uma distância euclidiana u, freqüentemente denominada "lag" e
- iv) N(u) é o número total de pares de posições em M(u).

É comum serem considerados pertencentes a M(u) os pares de parcelas $(x_i, x_{i'})$ tal que, $||x_i - x_{i'}|| \approx u$.

O variograma verdadeiro é aquele que representa a situação real de um micronutriente no solo do campo experimental, ou do depósito de algum minério e é sempre desconhecido (Kitanidis, 1997). Assim, na análise estrutural o objetivo é estudar qual é o variograma teórico que melhor se ajusta ao variograma experimental, de forma que a partir do variograma teórico, realizam-se inferências sobre o variorama verdadeiro (Guerra, 1988).

O método de ajuste de variogramas teóricos a variogramas experimentais é denominado de método das aproximações sucessivas, pois o processo de ajuste finaliza quando as discrepâncias entre os valores experimentais e teóricos forem mínimas (Oliveira, 2003).

2.1.2.1 Características dos parâmetros que descrevem um semivariograma

A construção do semivariograma para identificar a dependência espacial origina-se nas diferenças entre duas unidades experimentais georeferenciadas no espaço, separadas por uma distância u.

Na presença de dependência espacial , em geral, duas parcelas próximas são mais semelhantes que duas mais afastadas permitindo afirmar que $\gamma(u)$ aumenta à medida que u aumenta até estabilizar. Este ponto, em que os dados deixam de apresentar dependência espacial, é chamado de patamar ou sill e, no semivariograma, a distância, a partir da origem até o patamar é chamada de alcance ou range denotado neste trabalho por ϕ .

Pela definição $\gamma(u)$ para u=0 deverá ser zero, mas na prática observase que há casos em que u=0 e $\gamma(u)$ aproxima-se de um valor positivo denominado efeito pepita, ou efeito nugget simbolizado por τ^2 .

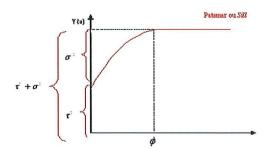


Figura 1 - Semivariograma apresentando variância nugget diferente de zero

Na Figura 1 podem-se identificar os parâmetros do semivariograma, com efeito pepita, da seguinte forma: τ^2 é o efeito pepita ou nugget, σ^2 é a variância

ou sill, $\sigma^2 + \tau^2$ a variância total e patamar também conhecida como sill total e ϕ que é o alcance ou range.

Alguns autores afirmam que o fato de ocorrer efeito pepita, pode estar associado a algum erro de medição (Vieira, 1996; Guerra, 1988).

2.1.2.2 Modelos teóricos de semivariogramas

a) Modelo Esférico ou Modelo de Matheron

Este modelo apresenta um crescimento rápido na origem, alcança um patamar a uma distância u finita e é representado por:

$$\gamma(u) = \begin{cases} \tau^2 + \sigma^2 \left(\frac{3u}{2\phi} - \frac{1}{2} \left(\frac{u}{\phi} \right)^3 \right) & \text{se } 0 < u \le \phi, \\ \tau^2 + \sigma^2 & \text{se } u > \phi, \\ 0 & \text{se } u = 0, \end{cases}$$
(9)

com $\tau^2 \ge 0$, $\sigma^2 \ge 0$, $\phi \ge 0$ e $u \ge 0$.

b) Modelo Exponencial ou de Formery

Este modelo apresenta um comportamento linear na origem, alcança seu patamar somente assintoticamente e é representado por:

$$\gamma(u) = \begin{cases} \tau^2 + \sigma^2 \left[1 - exp\left(-\frac{u}{\phi} \right) \right] & \text{se } 0 < u \le \phi, \\ \tau^2 + \sigma^2 & \text{se } u > \phi, \\ 0 & \text{se } u = 0. \end{cases}$$
 (10)

c) Modelo Matérn

Diggle & Ribeiro Jr. (2000) comentam sobre as diferentes estruturas de covariância utilizadas no estudo espacial e sugerem uma nova estrutura da função de correlação denominada família Matérn, descrita na equação (7). Essa função apresenta um parâmetro κ chamado de ordem do modelo Matérn o qual determina a suavização analítica do sinal S(x) num modelo Gaussiano.

2.1.2.3 Alguns métodos de estimação dos parâmetros do semivariograma

a) Método de estimação empírico

O método de ajuste de modelo de semivariograma mais comum é o ajuste a "sentimento". Por meio deste escolhe-se um modelo e seus parâmetros em que é sobreposto aos pontos estimados, e a partir de uma nuvem de pontos obtém-se a estimativa dos parâmetros do semivariograma por meio de uma avaliação visual (Ribeiro Junior, 1995). Como sugerido pelo autor, o processo deve ser repetido até que se obtenha um modelo adequado, e recomendou que seja utilizado o ajuste empírico como valor inicial para estimar os parâmetros do modelo de interesse por outro método em que as estimativas dos parâmetros sejam mais precisas. Menciona também, que este ajuste subjetivo é questionado e será cada vez menos utilizado.

b) Mínimos Quadrados Ordinários

Alguns métodos de ajuste que minimizam as somas de quadrados foram também utilizados, como o do mínimos quadrados ordinários e, considerado insensível para o propósito de ajuste, entretanto é um dos que permite a implementação de forma simplificada (Ribeiro Junior, 1995). Para o método de quadrados mínimos ordinário tem-se a seguinte expressão:

$$L(\tau^2, \sigma^2, \phi) = \sum_{M(u)} \{\hat{\gamma}(u) - \gamma(u; (\tau^2, \sigma^2, \phi))\}^2, \tag{11}$$

- i) $u \in M(u)$ estão descritos na equação (8);
- ii) $\hat{\gamma}(u)$ é o valor da semivariância obtida a partir do semivariograma empírico para as parcelas que estão separadas à distância u e
- iii) $\gamma(u;(\tau^2,\sigma^2,\phi))$ corresponde ao valor das semivariâncias esperadas a partir do semivariograma teórico para as parcelas que estão separadas pela distância u.

McBratney & Webster (1986) assumiram que os resíduos obtidos pela equação (11) são normalmente distribuídos e independentes, e todas as semivariâncias estimadas com a mesma variância. Segundo os autores, estas pressuposições são criticadas, e ainda, através deste método não se leva em consideração que as semivariâncias entre dois pontos sejam correlacionadas. Essa correlação sendo ignorada, os resultados podem ser comprometidos quando se aumenta o número de observação (Samper & Ramirez (1990)). Esse problema foi contornado quando McBratney & Webster (1986) utilizaram o método de mínimos quadrados generalizados.

c) Mínimos Quadrados Generalizados

McBratney & Webster (1986), com o método de mínimos quadrados generalizados, obtiveram estimativas eficientes assintoticamente, mas com a exigência de inversão de uma matriz de grandes dimensões. A expressão utilizada para o ajuste dos parâmetros é dada por:

$$G((\tau^2, \sigma^2, \phi); \hat{\gamma}) = \{\hat{\gamma} - \gamma(\tau^2, \sigma^2, \phi)\}' \sum_{i=1}^{n-1} \{\hat{\gamma} - \gamma(\tau^2, \sigma^2, \phi)\}$$
(12)

- i) $\hat{\gamma}$ representa o vetor de estimativas de semivariâncias $n_u \times 1$, em que n_u é o número de distâncias u;
- ii) $\gamma(\tau^2, \sigma^2, \phi)$ é o vetor, $n_u \times 1$, das semivariâncias estimadas que depende do vetor das estimativas dos parâmetros;
- iii) \sum é a matriz de variâncias e covariâncias de dimensões $n_u \times n_u$ entre as estimativas de $\gamma(u; \tau^2, \sigma^2, \phi)$.

d) Mínimos Quadrados Ponderados

Ainda McBratney & Webster (1986) consideraram o método de mínimos quadrados ponderados mais real e computacionalmente eficiente e o modelo é ajustado minimizando a seguinte equação:

$$W(\tau^{2}, \sigma^{2}, \phi) = \sum_{M(u)} \frac{\{\hat{\gamma}(u) - \gamma(u; (\tau^{2}, \sigma^{2}, \phi))\}^{2}}{\text{var}[\hat{\gamma}(u)]},$$
(13)

em que, $var[\hat{\gamma}(u)]$ é a variância da estimativa da semivariância.

Depois Barry et al². (1997) propopuseram o método de mínimos quadrados n-ponderados, que pesa os resíduos somente de acordo com os números de parcelas que estão separadas pela distância u, tal método consiste na minimização da expressão:

$$W_p(\tau^2, \sigma^2, \phi) = \sum_{M(u)} N(u) \{ \hat{\gamma}(u) - \gamma(u; (\tau^2, \sigma^2, \phi)) \}^2, \tag{14}$$

em que, u, M(u), $\hat{\gamma}(u)$ e $\gamma(u; (\tau^2, \sigma^2, \phi))$, já foram definidos pelas equações (8) e (11).

2.2 Inferência Bayesiana

Na perspectiva Bayesiana, os parâmetros desconhecidos são considerados aleatórios e todas as conclusões estão baseadas na distribuição condicional dos parâmetros sobre os dados observados, a distribuição a posteriori.

Segundo Gilks et al. (1997), seja y os dados observados e $\theta \in \Re^k$ os parâmetros do modelo. A inferência formal requer a distribuição conjunta das quantidades aleatórias, $\pi(y,\theta)$, a qual é composta por duas partes: a distribuição a priori $\pi(\theta)$ e a verossimilhança $L(\theta|y)$. Observados os dados y, o teorema de Bayes é usado para determinar, condicionalmente aos dados, a distribuição de θ , ou seja,

$$\pi(\theta|y) = \frac{\pi(\theta)L(\theta|y)}{\int \pi(\theta)L(\theta|y)d\theta}.$$
 (15)

²Mencionado por Diggle & Ribeiro (2000).

Essa distribuição é chamada de distribuição a posteriori de θ , e pode ser escrita também como,

$$\pi(\theta|y) \propto \pi(\theta)L(\theta|y).$$

Quantidades como momentos, quantis, entre outras, são expressas em termos de esperanças a posteriori de funções de θ . A esperança a posteriori de uma função $f(\theta)$ é dada por,

$$E[f(\theta)] = \int f(\theta)\pi(\theta|y) = \int f(\theta) \frac{\pi(\theta)L(\theta|y)d\theta}{\int \pi(\theta)L(\theta|y)d\theta}.$$
 (16)

Porém essa integração, geralmente é analiticamente intratável, principalmente nos casos em que a dimensão de θ é alta. O uso de métodos computacionalmente intensivos tais como MCMC (Monte Carlo com cadeia de Markov), fornece uma alternativa para a obtenção de esperanças da distribuição a posteriori conjunta ou marginal.

2.2.1 Monte Carlo com Cadeia de Markov

A integração Monte Carlo avalia $E[f(\theta)]$ apresentada em (16), gerando amostras $\{X_t, t=1,...,n\}$ de $\pi(.|y)$ e aproximando

$$E[f(\theta)] \approx \frac{1}{n} \sum_{t=1}^{n} f(X_t). \tag{17}$$

Dessa forma, a média populacional de $f(\theta)$ é estimada pela média amostral. Quando as amostras $\{X_t, t=1,...,n\}$ são independentes, a aproximação pode ser feita tão precisa quanto desejado, aumentando-se o tamanho amostral n. Nesse caso, n está sob o controle do analista, isto não é o tamanho fixo de uma amostra de dados.

Geralmente, não é possível gerar amostras $\{X_t, t=1,...,n\}$ independentes a partir de $\pi(.|y)$, pois $\pi(.|y)$ pode ser diferente das distribuições padrões. Contudo $\{X_t, t=1,...,n\}$ não precisa ser necessariamente independente.

As $\{X_t, t=1,...,n\}$ podem ser geradas por qualquer processo que gera amostras através do suporte de $\pi(.|y)$ numa proporção correta. Uma forma de

se fazer isso, é construir essa amostra através de uma cadeia de Markov tendo $\pi(.|y)$ como sua distribuição estacionária.

Suponha a geração de uma seqüência de variáveis aleatórias, $\{X_0, X_1, X_2, ...\}$, tal que para cada tempo $t \geq 0$, o próximo estado X_{t+1} é amostrado da distribuição $P(X_{t+1}|X_t)$ que depende somente do estado atual da cadeia X_t . Isto é, dado X_t o próximo estado X_{t+1} não depende mais da história da cadeia $\{X_0, X_1, ..., X_{t-1}\}$. Esta seqüência é chamada uma cadeia de Markov, e P(.|.) é chamado Kernel de transição . É assumido que a cadeia é homogênea, isto é, P(.|.) não depende de t. Como o estado X_0 afeta X_t ? Esta questão refere-se à distribuição de X_t dado o estado X_0 , a qual é denotada por $P^{(t)}(X_t|X_0)$. Aqui, não estão sendo dadas as variáveis $X_1, X_2, ..., X_{t-1}$. Assim, X_t depende diretamente de X_0 . Sujeito a condições de regularidade, a cadeia irá gradualmente "esquecendo" seu estado inicial e $P^{(t)}(.|X_0)$ eventualmente convergirá para uma única distribuição estacionária (ou invariante), a qual não depende de t ou X_0 . Por enquanto, essa distribuição é denotada por $\phi(.)$. Assim com o crescimento de t as amostras $\{X_t, t=1,...,n\}$ se parecerão mais e mais com amostras dependentes de $\phi(.)$.

Portanto, depois de um período de convergência de m iterações, conhecido como burn-in, as amostras $\{X_t, t=m+1,...,n\}$ serão aproximadamente amostras dependentes de $\phi(.)$. A cadeia de Markov gerada pode ser usada para estimar a esperança E[f(X)] em que X tem distribuição $\phi(.)$. As Amostras burn-in são descartadas para o cálculo do estimador dado por:

$$\overline{f} = \frac{1}{n-m} \sum_{t=m+1}^{n} f(X_t).$$
 (18)

Esta é a denominada média ergódica.

A equação (18) mostra como a cadeia de Markov pode ser usada para estimar E[f(X)], em que a esperança é tomada sobre a distribuição estacionária $\phi(.)$. Mas primeiro é necessário descobrir como construir uma cadeia de Markov para garantir que a distribuição $\phi(.)$ seja precisamente a distribuição de interesse $\pi(\theta|y)$ e dessa forma poder estimar o momento desejado, ou seja, $E[f(\theta)]$. Para

isso existem algoritmos, como por exemplo, o proposto por Hastings (1970), o qual é uma generalização do método proposto por Metropolis et al. (1953), conhecido como algoritmo *Metropolis-Hastings*, e o algoritmo Gibbs Sampling, como caso particular do algoritmo de *Metropolis-Hastings*

Além das n iterações e m burn-in, os algoritmos utilizados para construir a cadeia de Markov requerem que seja estabelecido o valor do "espaçamento" que garante a "independência" das amostras $\{X_t, t=1,...,n\}$, considerando-as de "t em t" também conhecido por thin.

2.3 Modelo Markoviano

O modelo Markoviano é adequado aos denominados dados de área, que são representados por um vetor $(y_1, y_2, ..., y_n)$ e cada um dos valores y_i está associado a uma área i do mapa dividido em n áreas.³ Sua principal característica é que representa uma agregação de valores que encontram-se dispersos dentro de cada uma das áreas. Ou seja, y_i , não se refere a nenhuma posição específica dentro da área i mas, na verdade, está associado a toda a área i em questão.

O modelo Modelo Markoviano é dado por

$$y_{ik} = \alpha_k + S_i + Z_i, \quad i = 1, ..., n \quad e \quad k = 1, ..., p.$$
 (19)

Ou seja, supondo a área experimental dividida em n parcelas regulares ou não, a resposta da i-ésima parcela que recebeu o k-ésimo tratamento, y_{ik} , depende do valor da média do tratamento aplicado a ela, α_k , de um efeito espacial S_i , que representa índices ambientais não medidos e que depende da posição geográfica x_i ocupada pela parcela i, e ainda de um erro aleatório e independente, Z_i , normalmente distribuído, com média 0 e variância τ^2 , que ocorre devido a erros de medida.

O Modelo Markoviano é assim denominado, pois o componente espa-

³No contexto de delineamento experimental, as áreas referem-se as parcelas e o mapa refere-se a área experimetal.

cial, S_i , tem como distribuição de probabilidade um campo aleatório de Markov. Seja, $S=(S_1,...,S_n)$, o vetor de parâmetros para o qual uma distribuição de probabilidade com um caráter espacial deve ser definida. Supondo estabelecida uma estrutura de vizinhança espacial, denota-se por ∂_i o conjunto de vizinhos da área i e o vetor $(S_1...,S_{i-1},S_{i+1},...,S_k)$ por S_{-i} . A distribuição de S possui a propriedade Markoviana, e é chamada de campo aleatório de Markov, se, para todo i, i = 1,...,n, sua distribuição conjunta satisfaz a propriedade dada por

$$f(S_i|S_{-i}) = f(S_i|\{S_j : j \in \partial_i\}, \tag{20}$$

isto é, S possui uma distribuição de Markov se a distribuição de S_i condicionada em todo o restante do mapa for a mesma que a distribuição de S_i condicionada apenas nos valores de áreas vizinhas de i. Quem é definido como vizinho de i é, em princípio, deixado a critério do usuário desses modelos. Duas áreas, j e k, ambas pertencentes a ∂_i , poderão ter contribuições ou pesos diferentes na determinação do valor de S_i .

Conforme a abordagem Bayesiana, condicionalmente a $S = \{S_1, ..., S_n\}$ e $\alpha = (\alpha_1, ..., \alpha_p)$, os y_{ik} são realizações mutuamente independentes dos Y_{ik} , sendo que matricialmente,

$$Y|S, \alpha \sim MVN(F\alpha + WS, \tau^2 I_n),$$
 (21)

em que,

- i) F é a matriz de delineamento;
- ii) α é um vetor com as médias dos tratamentos;
- iii) W uma matriz $n \times n$ com a estrutura de vizinhança considerada;
- iv) S o vetor dos efeitos ambientais não medidos e
- v) τ^2 a variância do erro aleatório independente, equivalente a $\frac{1}{\lambda_y}$, em que λ_y é o parâmetro de precisão dos dados.

Esse modelo será tratado segundo a abordagem Bayesiana. Portanto, a distribuição a priori para o vetor paramétrico $\theta = (\alpha_1, ..., \alpha_p, S)$ deve ser especificada. Geralmente é assumido a priori que esses parâmetros são independentes.

2.3.1 Escolha das distribuições a priori

Em termos de modelagem, as distribuições a priori mais simples atribuídas às médias dos tratamentos α_k , são a uniforme $(-\infty, +\infty)$ e a $N(0, I_m/\lambda_\alpha)$, com $\lambda_\alpha \sim \Gamma(a,b)$, suficientemente dispersa. Besag & Higdon (1999), sugerem a substituição da distribuição a priori Gaussiana por uma distribuição t-Student, como uma alternativa para a acomodação de outliers na análise Bayesiana de experimentos agrícolas de campo.

A distribuição a priori escolhida para o componente, $S = (S_1, ..., S_n)'$, que tem uma estrutura espacial é, tipicamente um campo aleatório de Markov Gaussiano. Esta distribuição além de ter uma estrutura de vizinhança espacial, depende de um parâmetro desconhecido adicional λ_S , inversamente relacionado à variabilidade dos Si's e que mede o grau de similaridade espacial entre estes componentes do modelo.

2.3.2 A distribuição a priori baseada em campos aleatórios Markovianos

O modelo mais popular a priori para os efeitos aleatórios espacialmente estruturados, $S=(S_1,...,S_n)$, num modelo hierárquico Bayesiano é aquele em que a distribuição conjunta é uma distribuição normal multivariada. O valor do índice ambiental na parcela i, S_i , i=1,...,n, pode ser visto como uma auto-regressão na média dos índices ambientais, $S_{j's}$, dos vizinhos j's da parcela i. Assumindo que existe uma tendência μ , matricialmente, esse processo é dado por,

$$S - \mu = \gamma W(S - \mu) + e_i, \qquad e_i \sim N_n(0_n, \sigma^2 I_n), \tag{22}$$

em que a matriz W é uma matriz de pesos representando a estrutura de vizinhança, por exemplo, $W_{ij} = 1$, se as parcelas i e j são adjacentes ($i \sim j$) e $W_{ij} = 0$, caso

contrário, e γ é o coeficiente linear dessa auto-regressão. O modelo (22) é conhecido na literatura como modelo espacial auto-regressivo.

A distribuição condicional do processo S, é conhecida como Modelo Auto-regressivo Condicional, também conhecido como modelo CAR (Besag, 1974), muito usado na literatura como conhecimento a priori dos efeitos aleatórios espacialmente estruturados, dado pela expressão,

$$(S_i|S_j, j \neq i) \sim N\left(\mu_i + \sum_j \gamma C_{ij}(S_j - \mu_j), \sigma_i^2\right), \tag{23}$$

em que C_{ij} são os elementos da matriz C, de dimensão $n \times n$, que representa uma matriz de pesos e reflete a dependência espacial entre as áreas i's e j's. Essa matriz é tal que, $I_n - C$, é simétrica e positiva definida⁴. Equivalentemente,

$$S_i - \mu_i = \sum_{j=1}^n \gamma C_{ij} (S_j - \mu_j) + \epsilon_i,$$
 (24)

em que $\epsilon_i \sim N(0, \nu_i^2)$, para i = 1, ..., n.

Segundo Besag (1974), a especificação condicional auto-regressiva (23), implica que, a distribuição conjunta de S é dada por

$$S \sim N(\mu, \sigma_S^2 \Sigma),$$
 (25)

com $\mu = (\mu_1, \mu_2, ..., \mu_n)$ e $\Sigma = (I - \gamma C)^{-1}M$, em que,

- i) $M = diag(\nu_1^2, ..., \nu_n^2)$, $n \times n$, com elementos M_{ii} proporcionais à σ_i^2 , variância condicional de $S_i | S_j$. Ou seja, $\sigma_i^2 = \sigma_S^2 M_{ii}$. E se ν^2 for o mesmo para qualquer $i, i = 1, ..., n, \sigma_i^2 = \sigma^2$ para qualquer i, i = 1, ..., n;
- ii) C como já dito, é uma matriz de pesos, com elementos C_{ij} refletindo a associação espacial entre áreas i e j e
- iii) γ controla a magnitude da correlação espacial.

 $^{^4}C$ é uma matriz definida positiva, se para qualquer vetor x, exceto x nulo, a forma quadrática x'Cx, é maior do que zero. Essa propriedade também assegura que a matriz C é não singular, ou seja, tem determinante diferente de zero e conseqüentemente é invertível.

A escolha da matriz Σ é importante, já que ela indica como a associação espacial entra no modelo. Σ será simétrica se e somente se $c_{ij}m_{jj}=c_{ji}m_{ii}$, i,j=1,...,I. Para assegurar que Σ seja positiva definida, o parâmetro γ deve pertencer ao intervalo $(\gamma_{min},\gamma_{max})$, em que $1/\gamma_{min}<0<1/\gamma_{max}$ são o menor e maior autovalores de $M^{-1/2}CM^{1/2}$. Na prática, freqüentemente é esperada uma dependência espacial positiva, e desse modo é comum tomar γ no intervalo $(0,\gamma_{max})$.

A maioria das abordagens, incorporando um modelo condicional para dependência espacial, iniciam especificando-se um conjunto de pesos espaciais para o uso em (23). Esses pesos estão associados a um conjunto de parcelas vizinhas que contribuem com um peso positivo para a esperança de S_i , com $c_{ij}=0$ para as demais regiões e $c_{ii}=0$. Há um grande número de escolhas para M, C e γ que satisfazem à condição de que Σ seja uma matriz de variâncias, simétrica e positiva-definida.

Um modelo de campos de Markov muito usado para prioris espaciais é a auto-regressão condicional intrínseca gaussiana (Besag et al. 1991; Besag & Kooperberg, 1995). Seja ∂_i o conjunto dos j's vizinhos da área i e n_i o número de elementos desse conjunto, a forma do modelo é dada pela expressão

$$S_i|S_j \sim N\left(\frac{\sum_{j \in \partial_i} w_{ij} S_j}{\sum_{j \in \partial_i} w_{ij}}, \frac{\sigma_S^2}{\sum_{j \in \partial_i} w_{ij}}\right). \tag{26}$$

Fazendo-se $w_{ij}=1,$ o que é equivalente a $c_{ij}=\frac{w_{ij}}{\sum_j w_{ij}},$ para $j\in\partial_i,$ tem-se o modelo

$$\pi(S_i|S_j) \sim N\left(\frac{\sum_{j \in \partial_i} S_j}{n_i}, \frac{\sigma_S^2}{n_i} = \frac{1}{\lambda_S n_i}\right),$$
 (27)

em que n_i é o número de vizinhos da parcela i. Sob esse modelo, o efeito espacial na área i é suavizado para um efeito médio definido pelo conjunto de áreas na vizinhança da área i, com variância inversamente proporcional ao número de vizinhos. A comparação com o modelo (23) mostra que o modelo (27) corresponde às seguintes

escolhas:

$$\mu=0; \quad \gamma=\gamma_{max}=1; \quad M_{ii}=rac{1}{n_i}; \quad c_{ij}=\left\{ egin{array}{l} rac{1}{n_i}, & ext{se as áreas i e j são vizinhas} \ 0, & ext{em caso contrário.} \end{array}
ight.$$

Esta especificação não leva a uma matriz Σ definida positiva. Para verificar isto, basta notar que as somas das linhas de I-C são todas identicamente iguais a zero, já que a i-ésima linha possui uma única entrada igual a 1 e as n_i entradas iguais a $-1/n_i$. Assim, Σ não tem posto completo e não é invertível. Devido a isso, o modelo (27) define uma distribuição a priori imprópria para os efeitos aleatórios espacialmente estruturados S em (19). A variância $\sigma_S^2 = \frac{1}{\lambda_S}$ em (27), a qual somente é interpretável condicionalmente, não é mais proporcional à variância marginal, já que o modelo é não estacionário (é possível ter uma média arbitrária para cada S_i) e, portanto, a distribuição conjunta de S não existe. Segundo Wakefield et al. (2000), uma grande vantagem do modelo não estacionário (27), é que a forma da dependência espacial pode variar através da região de estudo.

A forma de uma distribuição conjunta que implicaria nas distribuições condicionais em (26) é dada por

$$\pi(S_1, ..., S_n) \propto \lambda_S^{\frac{n}{2}} exp\left(-\frac{1}{2\lambda_S} \sum_{j \in \partial_i} w_{ij} (S_i - S_j)^2\right). \tag{28}$$

É possível notar que a distribuição conjunta (28) é uma distribuição imprópria, pois seu lado direito é invariante por translação de S por meio de uma mesma constante c adicionada a todos os seus elementos, isto é, $f(S_1,...,S_n) = f(S_1-c,...,S_n-c)$ e, portanto, a integral em R^n da função (29) não é finita. Ela pode ser transformada numa *priori* própria adicionando alguma restrição linear que os parâmetros deveriam satisfazer.

Embora não seja necessária uma priori própria para obter uma a posteriori própria, este procedimento, chamado de centering on the fly por Besag & Kooperberg (1995), dá aos algoritmos de Monte Carlo Cadeia de Markov um desempenho melhor. Os autores propuseram uma reparametrização, incluindo no modelo um intercepto separado, μ , representando uma média geral, e a restrição $\sum_i S_i = 0$,

para assegurar que o modelo seja identificável. A distribuição a priori para esse termo de intercepto é totalmente não informativa, ou seja, uma uniforme variando entre $-\infty$ e $+\infty$.

Outras escolhas para a matriz de variância-covariância podem ser feitas, por exemplo, Cressie & Chan (1989) propuseram uma escolha para a matriz de um processo CAR (auto-regressão condicional), a qual pode ser usada como uma distribuição a priori para os efeitos aleatórios espacialmente estruturados do modelo (19) com as seguintes escolhas para M, C e γ ,

$$\begin{split} \gamma &\in (\gamma_{max}, \gamma_{min}); \quad M_{ii} = \frac{1}{E_i}, \text{o valor esperado da i-\'esima \'area e} \\ c_{ij} &= \left\{ \begin{array}{l} \sqrt{\frac{E_j}{E_i}}, & \text{se as \'areas i e j s\~ao vizinhas} \\ 0, & \text{em caso contr\'ario.} \end{array} \right. \end{split}$$

Ao contrário da auto-regressão intrínseca (27) que seleciona $\gamma = \gamma_{max} = 1$, indicando uma forte suposição a priori em relação à dependência espacial, o modelo proposto por Cressie & Chan (1989) permite aos dados sugerir prováveis valores para γ dentre aqueles para os quais Σ é positiva-definida.

Quanto à escolha da vizinhança, vários autores (Clayton & Kaldor, 1987; Besag et al., 1991; Waller et al., 1997, dentre outros) têm considerado as áreas⁵ i e j como vizinhas se elas compartilham um limite comum. Essa escolha é razoável no caso de todas as áreas serem de tamanho similar e arranjadas num padrão regular, como existem contextos em que isso não ocorre, como por exemplo, em mapeamento de doenças, em que as regiões podem variar muito tanto em área geográfica quanto em tamanho da população, outros esquemas de vizinhanças são propostos (Conlon & Waller, 1998; Cliff & Ord, 1981, dentre outros). Por exemplo, Cressie & Chan (1989) consideram uma estrutura de vizinhança que depende da distância entre centróides de áreas.

Cada uma das distribuições a priori dos componentes $\alpha = (\alpha_1, ..., \alpha_p)$, $S = (S_1, ..., S_n)$ e a distribuição condicional dos dados, dependem dos parâmetros

⁵Em um delineamento experimental, essa área refere-se a uma parcela.

do inverso da variância (precisão), λ_{α} , λ_{S} e λ_{y} . Estes parâmetros são chamados de hiperparâmetros e são modelados por uma distribuição chamada hiperpriori, já que ela se refere aos parâmetros da distribuição a priori dos parâmetros da função de verossimilhança. Deve-se tomar cuidado na especificação destas hiperprioris, pois, por exemplo, o uso de prioris impróprias uniformes pode resultar em distribuições a posteriori impróprias.

Tipicamente tem sido escolhida uma distribuição Gama $\Gamma(\epsilon,\epsilon)$ para o inverso da variância (precisão) dos efeitos aleatórios, com ϵ sendo um valor pequeno tal como 10^{-2} ou 10^{-3} para ambos os hiperparâmetros. Kellsall & Wakefiel (1999) sugerem o uso de uma hiperpriori Gama (0,5;0,005), já que, em muitos contextos esta dará um intervalo plausível para o inverso da variância. Segundo Ruiz Cárdenas (2002), a média da distribuição gama representa a suposição a priori dos valores dos parâmetros de precisão λ_{α} , λ_{S} e λ_{y} , e sua variância reflete a incerteza em relação a esta crença a priori. Portanto, as estimativas Bayesianas dos parâmetros de interesse no modelo poderiam ser sensíveis à escolha dos parâmetros desta hiperpriori.

3 MATERIAL E MÉTODOS

3.1 Material

A motivação para esse trabalho surgiu a partir de um conjunto de dados apresentado por Ver Hoef & Cressie (1993), provenientes de um ensaio em branco no qual foram adicionados efeitos de tratamentos arbitrários. Os dados representam contagens de diferentes espécies de plantas antes da adição dos tratamentos de queimadas nos meses de maio, agosto, novembro e fevereiro. Esse experimento foi organizado em uma área florestal subdividida em 25 parcelas de mesmo tamanho dispostas em 5×5 .

Também foi material do presente estudo, o conjunto de dados proveniente de um ensaio em branco apresentado por Mercer & Hall⁷ (1911), que representam produção de grãos de milho em libras⁸. O ensaio em branco foi realizado em uma área de $4046,84 \ m^2$ subdividida em 500 parcelas de mesmo tamanho dispostas em 25 parcelas de leste para oeste e 20 parcelas de norte para sul.

Além dos ensaios em branco, foram usados conjuntos de dados a partir de duas simulações de experimentos subdivididas em simulações com baixo, médio e alto níveis de dependência espacial. Na primeira simulação foram gerados 5000 experimentos com 25 parcelas e na segunda 3000 experimentos com 225 parcelas para cada nível de dependência espacial.

⁶Experimento sem aplicação de tratamento.

⁷Mencionado por Cressie (1993).

⁸Unidade de medida equivalente a 453,592 gramas.

3.2 Métodos

Os modelos e métodos considerados nesse estudo foram os seguintes.

- i) a metodologia tradicional de unidades experimentais independentes por meio do modelo linear de Gauss-Markov Normal;
- ii) uma análise de covariância proposta por Papadakis;
- iii) a análise espacial por meio do modelo geoestatístico e três métodos para estimação de seus parâmetros: o método do variograma para estimação dos parâmetros de correlação seguido de mínimos quadrados generalizados para estimação dos efeitos dos tratamentos; o método de estimação de máxima verossimilhança e o método de estimação de máxima verossimilhança restrita, e
- iv) a análise espacial por meio do modelo Markoviano, baseado em prioris espaciais definidas como campos aleatórios de Markov, que foi tratado segundo a abordagem Bayesiana utilizando-se amostras construídas por métodos Monte Carlo com Cadeia de Markov (MCMC).

As comparações entre os modelos e métodos abordados nesse estudo foram realizadas através de cinco etapas:

- i) Experimento I: Análise de dados apresentados por Ver Hoef & Cressie (1993), provenientes de um ensaio em branco com 25 parcelas, no qual aos valores observados, foram adicionados efeitos arbitrários de tratamentos.
- ii) Experimento II: Análise de dados, provenientes de um ensaio em branco de 500 parcelas apresentado por Mercer & Hall (1911), no qual aos valores observados foram adicionados efeitos arbitrários de tratamentos.

- iii) Simulação I: Análise de dados gerados por simulações conforme experimentos inteiramente casualizados com 25 parcelas e diferentes níveis de dependência espacial.
- iv) Simulação II: Análise de dados simulados gerados por simulações conforme experimentos inteiramente casualizados de 225 parcelas e diferentes níveis de dependência espacial.
- v) Simulação III: Análise de 5000 conjuntos de dados gerados a partir da simulação de 5000 diferentes aleatorizações de tratamentos, cujos efeitos arbitrários foram adicionados aos dados do ensaio em branco apresentado por Ver Hoef & Cressie (1993).

3.2.1 Modelo de Gauss-Markov Normal

Para a realização das análises de variância de acordo com o modelo de Gauss-Markov Normal, foi utilizado o modelo linear de um delineamento inteiramente casualizado dado pela expressão

$$y_{ik} = \alpha_k + e_i, \qquad i = 1, ..., n \quad e \quad k = 1, ..., p,$$
 (29)

em que,

- i) y_{ik} é o valor observado na i-ésima parcela que recebeu o k-ésimo tratamento;
- ii) α_k é a média do k-ésimo tratamento aplicado à i-ésima parcela e
- iii) e_i é o erro experimental associado à observação y_{ik} .

A pressuposição feita com relação ao erro experimental foi o que definiu esse modelo. Considerou-se que os erros são variáveis aleatórias independentes e têm distribuição normal com média zero e variância constante σ_a^2 . Dessa forma, no modelo Gauss-Markov Normal, os erros são apresentados, matricialmente, por

$$e \sim N(0, I_n \sigma_a^2),$$

em que I_n é a matriz identidade $n \times n$.

As estimativas dos contrastes são dadas por:

$$\hat{c} = D(F'F)^{-}F'y,\tag{30}$$

em que F é a matriz do delineamento e D é a matriz dos coeficientes dos contrastes. E a variância desses contrastes é dada por

$$v\hat{a}r(\hat{c}) = \sigma_a^2 D(F'F)^- D', \tag{31}$$

em que $\sigma_a^2 = \frac{r'r}{n - rank(F)}$, sendo $r = y - F\hat{\alpha}$, o vetor de resíduos.

3.2.2 Modelagem de Papadakis (PP)

Nessa modelagem foi realizada uma análise de covariância, em que o valor observado em cada parcela foi corrigido pela média do tratamento e pelo efeito de uma covariável espacial adicionada ao modelo, que consiste na média dos resíduos das parcelas vizinhas, denominada aqui como índice ambiental. Para isso, foi estimado o resíduo de cada unidade experimental da seguinte forma:

$$r_{ik} = y_{ik} - \overline{y}_k,$$

em que, r_{ik} é o resíduo da i-ésima parcela que recebeu o k-ésimo tratamento, y_{ik} é o valor observado na i-ésima parcela que recebeu o k-ésimo tratamento e \overline{y}_k é a média do k-ésimo tratamento.

Os valores r_{ik} foram posicionados de acordo com as posições das parcelas recebendo a notação $r_{(l,c)}$. Cada parcela i é representada pelas coordenadas (l,c), em que l=1,...,L, é a coordenada que representa a l-ésima parte da área experimental dividida em L linhas e, c=1,...,C representa a c-ésima parte da área experimental dividida em C colunas.

O índice ambiental (Z), considerando a parcela i representada pela posição (l,c), foi calculado em cada parcela da seguinte forma:

$$Z_{(l,c)} = \frac{r_{(l,c)} + r_{(l+1,c)} + r_{(l-1,c)} + r_{(l,c+1)} + r_{(l,c-1)}}{5}.$$
 (32)

Ou seja, foi considerada como estrutura de vizinhança da parcela i, quatro parcelas, uma acima, uma abaixo, uma à esquerda e uma à direita. Para as parcelas das extremidades foram consideradas as parcelas vizinhas disponíveis, dentre as quatro citadas.

Então foi realizada a análise de covariância em esquema inteiramente casualizado, de acordo com o modelo estatístico dado por

$$y_{ik} = \alpha_k + \beta Z_i + e_i, \tag{33}$$

em que

- i) y_{ik} , α_k e e_i , já foram descritos antes conforme equação (29).
- ii) β é o coeficiente de regressão entre a variável resposta e a covariável índice ambiental.
- iii) Z_i é a covariável, índice ambiental, medida na *i*-ésima parcela.

As estimativas dos contrastes entre as médias dos tratamentos são dadas por

$$\hat{c} = D(F'F)^{-}F'yb, \tag{34}$$

em que, $b = \{Z'[I_n - F(F'F)^-F']Z\}^-Z'[I_n - F(F'F)^-F']y$, F a matriz de delineamento, D a matriz dos coeficientes dos contrastes, Z o vetor das covariáveis e I_n a matriz identidade $n \times n$.

E a matriz de variâncias desses contrastes é expressa como

$$v\hat{a}r(\hat{c}) = D[(F'F)^{-} + (F'F)^{-}F'Z(R'_{z}R_{z})^{-1}Z'F(F'F)^{-}]D'\hat{\sigma}_{p}^{2},$$
(35)

em que, $R_z = (I_n - F(F'F)^- F')Z$, é o vetor de resíduos do ajustamento da covariável e $\hat{\sigma}_p^2 = \frac{y'y - y'F(F'F)^- F'y - y'R_z(R'_zR_z)^{-1}R'_zy}{[n - rank(F) - rank(Z)]}$. Os erros quadráticos das estimativas dos contrastes consistem nos valores das raízes dos elementos da diagonal de $v\hat{a}r(\hat{c})$.

3.2.3 Modelagem geoestatística

Diferentemente do modelo de Papadakis que busca controlar a heterogeneidade da área acrescentando uma covariável ao modelo, a modelagem geoestatística modela a dependência espacial através de um componente de erro aleatório. O modelo geoestatístico assume que os dados, $y=(y_1,...,y_n)$, seguem o modelo linear Gaussiano dado por

$$y(x_{ik}) = \mu(x_{ik}) + S(x_{ik}) + z_i, \qquad i = 1, ..., n \text{ e } k = 1, ..., p,$$
 (36)

em que,

- i) x_{ik} é a posição espacial da *i*-ésima parcela que recebeu o k-ésimo tratamento.
- ii) $y(x_{ik})$ é o valor observado na *i*-ésima parcela que recebeu o k-ésimo tratamento.
- iii) $\mu(x_{ik})$ é determinada a partir de um modelo de regressão linear dado por

$$\mu(x_{ik}) = \sum_{k=1}^{p} f_k(x_i)\alpha_k, \tag{37}$$

em que, $f_k(.)$ são observações espacialmente relacionadas às variáveis explanatórias e $(f_1(x_i),...,f_p(x_i))$ é a *i*-ésima linha da matriz de delineamento.

- iv) S(.) é um processo estacionário Gaussiano, com média zero, variância σ^2 e função de correlação $\rho(u;\phi)$. Ou seja, $S \sim NMV(0,\sigma^2R(\phi))$, em que $R(\phi)$ é a matriz de covariância $n \times n$ com (i,j)-ésimo elemento, $\rho(u_{ij})$, sendo $u_{ij} = ||x_i x_j||$ a distância euclidiana entre x_i e x_j .
- v) Z_i são variáveis aleatórias, independentes e normalmente distribuídas com média zero e variância τ^2 .

Esse modelo pode ser escrito na forma matricial e expresso por

$$Y = F\alpha + S + \varepsilon \tag{38}$$

em que, $Y = (y(x_1), ..., y(x_n))$ é o vetor dos valores da variável resposta nas posições $(x_1, ..., x_n)$; α é o vetor das médias desconhecidas dos tratamentos; S é um processo estocástico e, ε é o vetor dos erros independentes e normalmente distribuídos.

Assim, para Y é assumida a distribuição normal multivariada, isto é, $Y \sim NMV(F\alpha, \sigma^2 R(\phi) + \tau^2 I_n)$, em que I_n é uma matriz identidade $n \times n$, $R(\phi)$ é a matriz de covariância de dimensão $n \times n$, cujos elementos são fornecidos pela função de correlação $\rho(u, \phi)$.

O modelo geoestatístico decompõe o erro, $e \sim N(0, I_n \sigma_a^2)$, do modelo de Gauss-Markov Normal (29) em dois componentes de erros, um de variáveis dependentes, $S \sim NMV(0, \sigma^2 R(\phi))$, que indicam índices ambientais não medidos e outro de variáveis independentes, $Z \sim N(0, I_n \tau^2)$, que ocorrem devido a erros de medidas. Como visto na revisão de literatura, existem várias escolhas para a função de correlação $\rho(u)$, nesse trabalho foram avaliadas três funções de correlação: a função exponencial e as funções Matérn, com $\kappa = 1$ e $\kappa = 2$ descritas no item 3.1.1.

As estimativas dos parâmetros do modelo geoestatístico foram obtidas pelos métodos de estimação de parâmetros: variograma seguido de mínimos quadrados generalizados, estimação de máxima verossimilhança e de máxima verossimilhança restrita.

3.2.3.1 Método de estimação do Variograma seguido de Mínimos Quadrados Generalizados

No método de estimação de parâmetros do variograma seguido de mínimos quadrados generalizados, os parâmetros de correlação σ^2 , τ^2 e ϕ foram estimados por um semivariograma e as médias dos tratamentos, $\alpha = (\alpha_1, ..., \alpha_p)$, foram estimadas pela teoria dos mínimos quadrados generalizados, a qual utiliza uma matriz de covariância constituída dos valores do semivariograma.

Um raciocínio intuitivo que justifica esse método é o seguinte: Devido ao delineamento inteiramente casualizado, existe a possibilidade de um mesmo tra-

tamento, k, ser alocado em várias parcelas próximas. Para contornar esse problema em vez de se estimarem α_k como a simples média de todas as parcelas designadas ao tratamento k, esse método utiliza uma média ponderada de modo a atribuir um peso menor àquelas parcelas que se encontram juntas, pois, provavelmente possuem valores mais similares e dessa forma, possuem pequenas informações extras além daquelas que estão sozinhas na mesma região. Por outro lado, uma parcela com o tratamento, k, isolada das outras com o mesmo tratamento designado, terá um peso alto porque representa uma região maior. Tais pesos, conhecidos também por pesos ótimos, são obtidos, através dos valores do semivariograma.

O método do variograma seguido de mínimos quadrados generalizados seguiu um processo iterativo, cujo primeiro passo consistiu na realização de uma análise de variância clássica ANOVA (i.i.d.), e uma vez obtidas as estimativas das médias dos tratamentos, denotadas $\hat{\alpha_k}$, os resíduos r_{ik} foram calculados da seguinte forma

$$r_{ik} = y_{ik} - \hat{\alpha_k},\tag{39}$$

em que r_{ik} e y_{ik} correspondem, respectivamente, ao resíduo e ao valor observado da i-ésima parcela que recebeu o k-ésimo tratamento.

O segundo passo desse processo iterativo consiste no ajustamento do semivariograma empírico, descrito pela equação (7), aos resíduos, r_{ik} , da análise clássica ANOVA (i.i.d.). Da seguinte forma:

$$\hat{\gamma}(u) = \frac{\sum_{M(u)} (R(x_i) - R(x_{i'}))^2}{2N(u)},\tag{40}$$

em que,

- i) $x_i = (l, c)$ é a posição espacial da i-ésima parcela.
- ii) $u = ||x_i x_j||$ é a distância euclidiana entre x_i e x_j ;
- iii) $R(x_i)$ e $R(x_{i'})$ são os resíduos da análise de variância clássica (ANOVA) das i-ésima e da i'-ésima parcela;

- iv) $M(u) \equiv \{[x_i = (l,c), x_{i'} = (l',c')] : \sqrt{(l-l')^2 + (c-c')^2} = u\}$, o conjunto de todos os pares de posições de parcelas que estão separadas a uma distância euclidiana u, freqüentemente denominada "lag" e
- v) N(u) é o número total de pares de parcelas em M(u).

A Figura 2 mostra um semivariograma empírico ajustado aos resíduos de um conjunto de dados arbitrário.

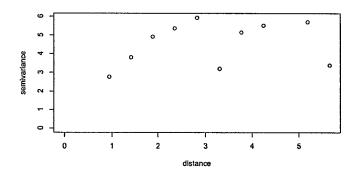


Figura 2 - Variograma empírico, em que cada ponto representa o valor da semivariância de N(u) pares de parcelas

Para que as estimativas das médias dos tratamentos sejam obtidas através do método de mínimos quadrados generalizados é necessário a construção da matriz de covariância entre todos os pares de parcelas. Entretanto, os valores empíricos calculados para $\hat{\gamma}(u)$ não devem ser usados na matriz Γ do semivariograma, porque Γ deve satisfazer propriedades positivas definidas, as quais asseguram que para qualquer inteiro m, qualquer conjunto de posições x_i e qualquer constante real a_i , a combinação linear $\sum_{i=1}^m a_i \gamma(u) \geq 0$ (Cressie, 1993).

Para a garantia das propriedades positivas definidas, no terceiro passo do processo foi ajustado um semivariograma teórico aos valores do semivariograma empírico. Nesse trabalho foram utilizadas três funções de correlação que definiram três diferentes modelos de semivariogramas teóricos: o exponencial, o modelo Matérn com $\kappa=1$ e o Matérn com $\kappa=2$, descritos no item 3.1.2.2. Para exemplificar foi tomado o semivariograma exponencial dado abaixo:

$$\gamma(u; (\tau^2, \sigma^2, \phi)) = \begin{cases} \tau^2 + \sigma^2 \left[1 - exp\left(-\frac{u}{\phi} \right) \right] & \text{se } 0 < u \le \phi, \\ \tau^2 + \sigma^2 & \text{se } u > \phi, \\ 0 & \text{se } u = 0. \end{cases}$$

$$(41)$$

Os parâmetros σ^2 , ϕ e τ^2 do semivariograma exponencial, equação (41), foram estimados pelo método de mínimos quadrados n-ponderados, descrito no item 3.1.2.3 conforme equação (14).

No quarto passo as estimativas $\hat{\sigma}^2$, $\hat{\tau}^2$ e $\hat{\phi}$ foram inseridas no modelo do semivariograma teórico, equação (41) e assim construiu-se a matriz de covariância, Γ de dimensão $n \times n$, cujo (i,j)-ésimo elemento corresponde a $\gamma(u_{ij}; (\hat{\tau}^2, \hat{\sigma}^2, \hat{\phi}))$.

A Figura 3 ilustra o ajuste de um variograma teórico aos valores do variograma empírico conforme Figura 2.

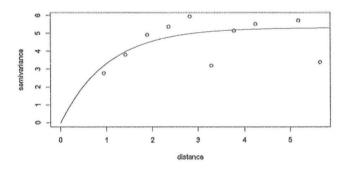


Figura 3 - Ajuste do variograma teórico

Como último passo, foram aplicados os estimadores de mínimos quadrados generalizados (GLS - Generalized Least Squares), os quais necessitam da matriz Γ obtida no quarto passo do processo para a obtenção das estimativas dos

contrastes entre os tratamentos e de seus respectivos erros padrões. As estimativas das médias dos tratamentos são dadas por

$$\hat{\alpha}_g = (F'\Gamma^{-1}F)^- F'\Gamma^{-1}y,\tag{42}$$

em que, F é a matriz de delineamento, Γ^{-1} é a inversa generalizada da matriz dos valores do semivariograma entre todos os possíveis pares de locações contendo os dados y. A estimativa da variância das estimativas $\hat{\alpha}_g$ é dada por

$$v\hat{a}r(\hat{\alpha}_g) = (F'\Gamma^{-1}F)^{-}. \tag{43}$$

As estimativas dos contrastes entre os tratamentos foram obtidas por

$$\hat{c} = D\hat{\alpha_g},\tag{44}$$

em que, $\hat{c} = (\hat{c}_1, ..., \hat{c}_m)$ é o vetor das estimativas dos contrastes entre os tratamentos e D é a matriz dos coeficientes desses contrastes.

A matriz de covariância dos contrastes é estimada por

$$v\hat{a}r(\hat{c}) = Dv\hat{a}r(\hat{\alpha}_{\sigma})D',\tag{45}$$

e seus erros quadráticos são as raízes dos elementos da diagonal de $var(\hat{c})$.

Resumo dos passos do método de estimação de parâmetros do variograma seguido de mínimos quadrados generalizados:

- 1- São calculados os resíduos da análise clássica (ANOVA) de acordo com a equação (39).
- 2- É construído um semivariograma empírico, dado em (40), a partir dos resíduos calculados no passo 1.
- 3- É ajustado um semivariograma teórico, como por exemplo o modelo exponencial apresentado por (41), aos valores do semivariograma empírico. Depois, são obtidas as estimativas dos parâmetros desse semivariograma, $\hat{\sigma}^2$, $\hat{\tau}^2$ e $\hat{\phi}$, pelo método de mínimos quadrados n-ponderados conforme a equação (14).

- 4- Nesse passo os valores das estimativas dos parâmetros de correlação são inseridas no modelo do semivariograma teórico e é construída a matriz de covariância, Γ de dimensão $n \times n$, entre todos os pares de parcelas, sendo que o (i, j)-ésimo elemento dessa matriz, corresponde ao valor, $\gamma(u_{ij}; \hat{\tau}^2, \hat{\sigma}^2, \hat{\phi})$, do semivariograma teórico, dado por (41), para o par de parcelas (i, j).
- 5- Enfim, as médias dos tratamentos são estimadas pela teoria de mínimos quadrados generalizados como apresentado na equação (42) e os contrastes entre os tratamentos são estimados conforme a equação (44).

3.2.3.2 Estimação de Máxima Verossimilhança

Considerando o modelo apresentado em (36) e as estruturas de covariância exponencial e família Matérn e assumindo que $y = (y_1, ..., y_n)$ segue uma distribuição multivariada Gaussiana tem-se que a função de verossimilhança é expressa por:

$$L(\alpha, \theta; y) = (2\pi)^{-\frac{n}{2}} |G(\theta)|^{-\frac{1}{2}} exp[-\frac{1}{2} (y - F\alpha)' \{G(\theta)\}^{-1} (y - F\alpha)], \tag{46}$$

em que y é o vetor de observações experimentais; F é a matriz do delineamento; α é o vetor de parâmetros desconhecidos associados aos efeitos fixos; θ são os parâmetros da matriz de covariância, $G(\theta) = \sigma^2 R(\phi) + \tau^2 I_n$, com $\theta = (\tau^2, \sigma^2, \phi)$.

Como pode ser visto, a função de verossimilhança é uma função dos parâmetros desconhecidos e não dos dados, a estimação de máxima verossimilhança consiste em encontrar os valores $(\hat{\alpha}, \hat{\theta})$ de (α, θ) mais prováveis de terem produzido os valores observados y.

Desta forma, procura-se o vetor (α, θ) que satisfaça a seguinte expressão:

$$L[(\hat{\alpha}, \hat{\theta}); y] = f[(y; \hat{\alpha}, \hat{\theta})] = \max\{f(y; \alpha, \theta)\} \ \alpha, \theta \in \Theta\}.$$

Então, o logaritmo da função de verossimilhança para (α, θ) é dado

por

$$\ell(\alpha, \theta) = -\frac{1}{2} \{ n \log(2\pi) + \log|G(\theta)| + (y - F\alpha)' \{ G(\theta) \}^{-1} (y - F\alpha) \}, \tag{47}$$

Por conveniência computacional, é possível a extração de um fator escalar a partir de $G(\theta)$, como por exemplo:

$$G(\theta) = \sigma^2 \{ \nu^2 + R(\phi) \} = \sigma^2 V, \tag{48}$$

em que $\nu^2 = \frac{\tau^2}{\sigma^2}$. Obtendo-se o logaritmo da função de verossimilhança $l(\alpha, \theta', \sigma^2; y)$, em que $\theta' = (\nu^2, \phi)$, dado por:

$$l(\alpha, \theta', \sigma^2; y) = -\frac{1}{2} \{ n \log(2\pi) + n \log(\sigma) + \log|V| + (y - F\alpha)' \{V\}^{-1} (y - F\alpha) \}$$
 (49)

Diferenciando-se parcialmente $l(\alpha, \theta', \sigma^2; y)$ em relação a α e σ^2 e igualando-se as derivadas a zero, obtém-se

$$\hat{\alpha}(\theta) = [F'(V)^{-1}F]^{-1}F'(V)^{-1}y,\tag{50}$$

que coincide com a estimativa de mínimos quadrados generalizados de α e matriz de covariância $v\hat{a}r(\hat{\alpha}) = \hat{\sigma}^2(F'V^{-1}F)^{-1}$.

$$\hat{\sigma}^2 = n^{-1}(y - F\hat{\alpha})'V^{-1}(y - F\hat{\alpha}).$$

As estimativas de θ' são obtidas substituindo-se $\hat{\alpha}$ e $\hat{\sigma}^2$ no logaritmo da função de verossimilhança apresentada na equação (49). Dessa forma, a maximização do logaritmo da função de verossimilhança só depende de dois parâmetros desconhecidos, ν^2 e ϕ , simplificando muito as rotinas de maximizações. Nesse trabalho, tais estimativas foram obtidas por um procedimento numérico por meio da subrotina geoR (Ribeiro Junior & Diggle, 2004), que encontra-se disponível no programa estatístico R.

Desta forma, as estimativas entre os contraste dos tratamentos são obtidas por

$$\hat{c} = D\hat{\alpha}$$

em que, D é a matriz dos coeficientes dos contrastes e $\hat{\alpha}$ o vetor das estimativas das médias dos tratamentos.

A matriz de covariância dessas estimativas é dada por

$$v\hat{a}r(\hat{c}) = Dv\hat{a}r(\hat{\alpha})D'. \tag{51}$$

Em que, $\hat{\alpha}$ e $v\hat{a}r(\hat{\alpha})$ são dadas conforme equação (50).

3.2.3.3 Estimação de Máxima Verossimilhança Restrita

Segundo Diggle & Ribeiro Jr. (2000), o método de estimação de máxima verossimilhança restrita ou residual (REML), é considerado uma melhoria do método de estimação de máxima verissimilhança por ser menos tendencioso para amostras pequenas.

Tal método foi utilizado inicialmente para estimar os componentes de variância na análise de variância para os efeitos fixos e aleatórios independentemente, ou seja, a função de verossimilhança é fatorada em duas partes, sendo uma delas totalmente livre dos efeitos fixos. Esta fatoração é obtida por meio de uma transformação linear nos dados originais (Patterson & Thompson, 1971).

Os estimadores de máxima verossimilhança restrita para os dados $Y \sim NMV(F\alpha, G(\theta))$, consistem na aplicação do método da máxima verossimilhança aos dados transformados por $Y^* = PY$, uma combinação linear em que Y^* não depende de α . Ou seja, pela suposição Gaussiana, $Y^* \sim NMV(0, P'G(\theta)P)$.

A matriz P é construída sem o conhecimento dos valores verdadeiros de α e θ , como por exemplo, o projetor ortogonal P expresso por:

$$P = I - F(F'F)^{-1}F'$$

O logaritmo da função da verossimilhança restrita para θ pode ser calculado por maximizar a seguinte expressão:

$$\ell^*(\theta) = -0.5\{\log|G(\theta)| + \log|F'\{G(\theta)\}^{-1}F| + (y - F\widetilde{\alpha})'\{G(\theta)\}^{-1}(y - F\widetilde{\alpha})\}, \quad (52)$$

em que $\widetilde{\alpha} = \hat{\alpha}(\theta)$.

Note que $\ell^*(\theta)$ depende de F, e consequentemente de uma especificação correta para o modelo de $\mu(x)$. Portanto dado o modelo para $\mu(.)$, o método emprega as mesmas objetividades da estimação de máxima verossimilhança. Segundo Diggle & Ribeiro Jr. (2000), esse método é mais sensível quanto à má especificação do modelo de média, equação (37), do que o da máxima verossimilhança.

De forma análoga ao estimador de máxima verossimilhança foram consideradas as estruturas de covariância exponencial e família Matérn.

3.2.4 A modelagem com estrutura Markoviana

Foi seguida a abordagem Bayesiana para a estimação dos parâmetros do modelo Markoviano dado pela equação

$$y_{ik} = \mu + \delta_k + S_i + e_i, \quad i = 1, ..., n \quad e \quad k = 1, ..., p.$$
 (53)

Foram atribuídas distribuições para μ , $\delta = (\delta_1, ..., \delta_k)$ e $S = (S_1, ..., S_n)$ que representam o conhecimento a priori a respeito desses parâmetros.

Nessa modelagem, o parâmetro α_k do modelo (19) foi substituído por $\mu + \delta_k$ no modelo (53), porque com visto na revisão de literatura, é preciso considerar o intercepto, μ , que representa a média geral do modelo, quando se utiliza um componente espacial com uma distribuição a priori baseada em uma auto-regressão intrínseca, dada pela equação (27). A priori para esse parâmetro consiste em uma distribuição uniforme variando em $(-\infty, +\infty)$.

A escolha feita para a distribuição do hiperparâmetro λ_S , se baseia em uma distribuição gama, não muito informativa, conforme pode ser visto na Figura 4, refletindo a incerteza sobre esse parâmetro.

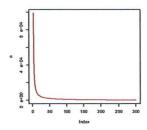


Figura 4 - Distribuição a priori gama para o hiperparâmetro λ_S

Tabela 1. Distribuições consideradas para os parâmetros e hiperparâmetros do modelo Markoviano

μ	δ	S	λ_S	λ_y
$\mathrm{Uni}(-\infty;+\infty)$	N(0; 0,001)	$\mathrm{CAR}(\lambda_S)$	Gama(0,001;0,001)	Gama(0,001;0,001)

 $CAR(\lambda_S)$ corresponde a uma auto-regressão condicional intrínseca Gaussiana como definida na equação (27), com parâmetro de precisão λ_S .

A média e a variância da distribuição, Gama(0,001; 0,001), atribuída ao componente espacial do modelo Markoviano, são (m=1; var=1000).

O esquema de vizinhança adotado na especificação da auto-regressão condicional Gaussiana, descrita na equação (27), e que constitue as distribuições a priori para os efeitos aleatórios espacialmente estruturados S, do modelo (19), será o de segunda ordem como descrito em Besag (1974). Nesse sistema, as parcelas vizinhas são aquelas que compartilham um limite em comum.



Figura 5 - Estrutura de vizinhança de segunda ordem

Os modelos foram implementados no programa Winbugs (Spiegelhalter, 2002). Na análise dos dados foram consideradas somente uma cadeia do amostrador Gibbs, em que foram gerados 200.000 valores para cada parâmetro do modelo Markoviano, dado em (53), dos quais foram desprezados os 10.000 primeiros e selecionado 1 a cada 20 dos 190.000 valores restantes. A convergência das estimativas foi verificada através da análise gráfica dos valores gerados e usando os diagnósticos de convergência, implementado no módulo CODA do programa R.

3.2.5 Experimento I

Para o Experimento I foram utilizados dados apresentados por Ver Hoef & Cressie (1993) provenientes de um ensaio em branco, no qual aos valores observados foram adicionados aleatoriamente efeitos arbitrários de tratamentos de queimadas. Esse ensaio foi realizado em uma área florestal dividida em 5×5 parcelas de mesmo tamanho, em que se observou o número de algumas espécies de plantas sem aplicação de tratamentos. Posteriormente, foram adicionados a seus valores observados efeitos arbitrários de tratamentos de queimadas: controle, queimadas no meses de maio, agosto, novembro e fevereiro, como se esses realmente tivessem sido aplicados às parcelas. Os tratamentos e os efeitos a eles atribuídos estão apresentados na Tabela 2.

Tabela 2. Efeitos atribuídos aos tratamentos-Experimento I

Tratamento	Ver Hoef & Cressie (1993)	
controle	0	
queimada maio	-3	
queimada agosto	-5	
queimada novembro	+6	
queimada fevereiro	+6	

Estabelecidos os valores para os efeitos dos tratamentos foram cons-

truídos os seguintes constrastes:

$$c_{1} = \frac{\alpha_{2} + \alpha_{3}}{2} - \alpha_{1}$$

$$c_{2} = \frac{\alpha_{4} + \alpha_{5}}{2} - \alpha_{1}$$

$$c_{3} = \frac{\alpha_{4} + \alpha_{5}}{2} - \frac{\alpha_{2} + \alpha_{3}}{2}$$

$$c_{4} = \alpha_{2} - \alpha_{3}$$

$$c_{5} = \alpha_{4} - \alpha_{5}.$$

$$(54)$$

As estimativas desses contrastes foram obtidas através dos métodos de estimação de parâmetros abordados nesse estudo e serviram como base para comparálos uma vez que os verdadeiros valores dos contrastes entre os tratamentos eram conhecidos.

3.2.6 Experimento II

No Experimento II foi utilizado um ensaio em branco apresentado por Mercer & Hall (1911) com 500 parcelas de mesmo tamanho, dispostas na área em 25 parcelas de leste para oeste e 20 de norte para sul. Os dados são de produção de milho, dada em libras, aos quais também foram adicionados efeitos arbitrários de tratamentos conforme Tabela 3.

Tabela 3. Efeitos atribuídos aos tratamentos-Experimento II

Tratamento	Mercer & Hall (1911)	
controle	0	
Trat 1	-0,6	
Trat 2	-0,9	
Trat 3	+1,2	
Trat 4	+1,2	

Estabelecidos os valores para os efeitos dos tratamentos, foram cons-

truídos os contrastes apresentados em (54).

Para a escolha desses contrastes, partiu-se do princípio de que o primeiro tratamento trata-se do controle, o segundo e o terceiro tratamentos possuem uma afinidade em comum assim como o terceiro e quarto.

3.2.7 Simulação I

A Simulação I foi realizada para uma investigação mais profunda dos resultados obtidos no Experimento I seguindo as mesmas características experimentais. Essa simulação foi subdividida em três níveis distintos de dependência espacial entre as parcelas:

- i) Simulação Ia: com baixa dependência espacial;
- ii) Simulação Ib: com média dependência espacial e
- iii) Simulação Ic: com alta dependência espacial.

Os dados dos experimentos simulados foram gerados a partir do modelo dado pela expressão

$$y(x_{ik}) = \alpha_k + S(x_i) + e_i, \quad i = 1, ..., n \text{ e } k = 1, ..., p,$$
 (55)

em que,

- i) $y(x_{ik})$ é o valor observado na i-ésima parcela que recebeu o k-ésimo tratamento;
- ii) α_k é a média do k-ésimo tratamento, e são valores fixos;
- iii) $x_i = (l, c)$ é a posição espacial da *i*-ésima parcela.
- iv) $S(x_{ik})$ é o componente espacial, que foi gerado aleatoriamente a partir da função grf, do pacote geoR (Diggle & Ribeiro Jr. (2004)) do programa R (Spiegelhalter, (2002)), que gera efeitos aleatórios espaciais dados os parâmetros de covariância σ^2 e ϕ .

v) e_{ik} são os erros aleatórios, normalmente distribuídos com média 0 e variância τ^2 , que também foram gerados aleatoriamente.

Para a Simulação I foram estabelecidos os seguintes valores aos elementos que compõe o modelo (55).

n=25,
$$l=1,...,5$$
 e $c=1,...,5$;
$$\alpha_1=25;\ \alpha_2=22;\ \alpha_3=20;\alpha_4=31;\ \alpha_5=31;$$

$$\sigma^2=4;\ \tau^2=1 \text{ e } \phi=0,5;\ 1,5 \text{ e } 2,5.$$

A cada valor de ϕ ($\phi=0,5;\ 1,5$ e 2,5) foram gerados 5000 conjuntos de 25 dados determinando a Simulação Ia: com baixa dependência espacial, a Simulação Ib: com média dependência espacial e a Simulação Ic: com alta dependência espacial, respectivamente.

O modelo Markoviano foi ajustado a somente um conjunto de dados escolhido aleatoriamente em cada simulação, pois, o programa computacional utilizado para ajustar este modelo não é adequado para a automatização dos procedimentos.

3.2.8 Simulação II

A Simulação II foi realizada para a investigação dos resultados obtidos no Experimento II, seguindo suas características experimentais. Essa simulação também foi subdividida em três níveis distintos de dependência espacial entre as parcelas:

- i) Simulação IIa: com baixa dependência espacial;
- ii) Simulação IIb: com média dependência espacial e
- iii) Simulação IIc: com alta dependência espacial.

Os dados da Simulação II também foram gerados a partir do modelo dado pela equação (55), em que foram estabelecidos os seguintes valores aos elementos que compõe o modelo.

n=225,
$$l=1,...,15$$
 e $c=1,...,15$; $\alpha_1=25; \ \alpha_2=22; \ \alpha_3=20; \alpha_4=31; \ \alpha_5=31;$ $\sigma^2=4; \ \tau^2=1$ e $\phi=2; \ 4$ e 6.

A cada valor de ϕ (ϕ = 2; 4 e 6), foram gerados 3000 conjuntos de 225 dados determinando a Simulação Ia: com baixa dependência espacial, a Simulação Ib: com média dependência espacial e a Simulação Ic: com alta dependência espacial entre as parcelas, respectivamente.

O modelo Markoviano foi ajustado a somente um conjunto de dados escolhido aleatoriamente em cada simulação, procedimento justificado na Simulação I.

3.2.9 Simulação III

A Simulação III foi realizada para investigar os resultados obtidos no Experimento I e verificar se os efeitos dos tratamentos realmente podem ser confundidos com a variação espacial da área experimental. A aleatorização utilizada por Ver Hoef & Cressie (1993) foi uma particular entre várias que poderiam ter ocorrido. Então, foram realizadas 5000 diferentes aleatorizações dos efeitos dos tratamentos apresentados na Tabela 2, cujos efeitos foram adicionados aos valores observados no ensaio em branco apresentado por Ver Hoef & Cressie (1993). Foram ajustados os modelos tratados segundo a análise freqüêntista: i) modelo clássico ANOVA (i.i.d), ii) modelo proposto por Papadakis e iii) modelo geoestatístico, para obtenção das estimativas dos contrastes apresentados por (54).

3.2.10 Medidas utilizadas na comparação dos métodos de estimação dos contrastes entre os tratamentos

É importante ressaltar que nas Simulação I e III foram gerados 5000 conjuntos de dados e na Simulação II foram gerados 3000. Portanto, foram geradas amostras de tamanho 3000 e 5000 para as medidas que se seguem.

- i) estimativas dos contrastes entre os tratamentos;
- ii) medida de dispersão;
- iii) erros quadráticos e
- iii) intervalo de confiança.

A fim de resumir os resultados alcançados nas simulações foram calculadas a partir dessas amostras as seguintes medidas.

- i) média das estimativas dos contrastes entre os tratamentos;
- ii) diferença absoluta entre o valor verdadeiro e a média das estimativas dos contrastes entre os tratamentos;
- iii) média das medidas de dispersão e desvio padrão;
- iv) média dos erros quadráticos;
- v) erros quadráticos obtidos por simulação de Monte Carlo e
- iv) intervalo de cobertura.

3.2.10.1 Diferença absoluta

Para os Experimentos I e II foram calculadas as diferenças absolutas entre as estimativas dos contrastes, $\hat{c_j}$, e seus valores verdadeiros, c_j , dadas por

$$dif = |c_i - \hat{c}_i| \quad j = 1, ..., 5.$$
 (56)

Para as simulações como foram gerados 3000 e 5000 conjuntos de dados, a diferença absoluta calculada foi entre a média das estimativas dos contrastes e seus valores verdadeiros dada por

$$dif = |c_j - \overline{\hat{c}_j}| \quad j = 1, ..., 5, \tag{57}$$

em que $\overline{\hat{c}_j}$ é a média aritmética dos 3000 ou 5000 valores estimados para os contrastes entre os tratamentos.

3.2.10.2 Medida de dispersão

A medida de dispersão utilizada para comparar os métodos de estimação de parâmetros abordados nesse estudo consistiu na variabilidade das estimativas dos contrastes apresentados na equação (54) em relação aos seus valores verdadeiros, dada por

$$D = \sum_{j=1}^{m} (\hat{c}_j - c_j)^2, \tag{58}$$

em que, $\hat{c_j}$ representa o valor do contraste estimado, c_j o valor verdadeiro do contraste e m o número de contrastes.

Nas simulações como foram gerados 3.000 ou 5.000 conjuntos de dados a cada nível de dependência espacial, foi necessário utilizar uma medida de dispersão média, que representasse seus resultados resumidamente, expressa por

$$\overline{D} = \frac{\sum_{i=1}^{w} D_i}{w},$$

em que, w é o número de conjuntos de dados gerados na simulação e D_i é a medida de dispersão calculada em cada conjunto de dados gerado.

3.2.10.3 Erros Quadráticos calculados analiticamente

Os erros quadráticos calculados analiticamente consistiram na raiz quadrada dos elementos das diagonais das matrizes de covariância das estimativas dos contrastes entre os tratamentos, segundo cada método de estimação de parâmetros, tais matrizes estão expressas em (31), (35), (45) e (51). Nas simulações foram calculadas as médias dos erros quadráticos para apresentação dos resultados.

3.2.10.4 Erros Quadráticos obtidos por simulação de Monte Carlo

Os erros quadráticos obtidos por simulação de Monte Carlo consistem nos desvios padrões das amostras das estimativas dos contrastes entre os tratamentos, dados por (54), geradas nas simulações.

3.2.10.5 Intervalo de confiança e Intervalo de cobertura

A cada conjunto de dados simulados foram construídos intervalos de confiança ao nível de 5% de significância. Para a análise clássica (ANOVA) e a análise de covariância proposta por Papadakis, o intervalo de confiança foi expresso por

$$IC(c_j) = \hat{c_j} \pm t_{\alpha} \sqrt{v \hat{a} r(\hat{c_j})}, \qquad j = 1, ..., 5,$$
 (59)

em que, \hat{c}_j é o valor estimado ao j-ésimo contraste e t_α é o t de Student tabelado a um nível α de significância. Para os métodos de estimação dos parâmetros do modelo geoestatístico a distribuição-t foi substituída pela distribuição normal, desta forma o intervalo de confiança foi dado por

$$IC(c_j) = \hat{c_j} \pm z_{\alpha} \sqrt{\hat{var}(\hat{c_j})}, \qquad j = 1, ..., 5,$$
 (60)

em que, z_{α} é o z normal tabelado a um nível α de significância.

Nas simulações foram calculados os intervalos de cobertura, que consistem na porcentagem dos intervalos de confiança gerados, conterem o verdadeiro valor do contraste, já que se trata de um valor conhecido. O intervalo de cobertura foi calculado para cada contraste, c_j : j=1,..., 5, separadamente a um nível de significância de 5% e também para o vetor composto pelos cinco contrastes, $c = (c_1, ..., c_5)$, com um nível de confiança de 77, 38%.

4 RESULTADOS E DISCUSSÃO

Para verificar se a modelagem da estrutura de correlação entre as parcelas aumentam a precisão das comparações entre os tratamentos, foram comparados à análise clássica ANOVA (i.i.d.), o modelo espacial proposto por Papadakis (1937), o modelo geoestatístico e o modelo Markoviano. É válido ressaltar as características de cada modelo abordado nesse estudo, descritos a seguir.

- i) Análise clássica ANOVA (i.i.d.): os erros experimentais são tomados como independentes no modelo dado por (29).
- ii) Modelo de Papadakis (PP): trata-se de uma análise de covariância, em que a covariável é de natureza espacial pois consiste na média dos resíduos de parcelas vizinhas, ou seja, a correlação é modelada por um componente fixo do modelo conforme (33).
- iii) Modelo geoestatístico: a dependência espacial é modelada pela adição de um componente de erros aleatórios dependentes, cuja correlação é modelada como função da distância entre os centros das parcelas conforme modelo (36). Seus parâmetros foram estimados por: i) variograma seguido de mínimos quadrados generalizados; ii) estimação de verossimilhança; iii) estimação de máxima verossimilhança restrita.
- iv) Modelo Markoviano: a dependência espacial é modelada por um componente de erros aleatórios dependente, com distribuição a priori dada por modelo CAR (Condicionalmente Auto Regressivo), em que o condicionamento é uma ob-

servação das parcelas vizinhas conforme (53). Seus parâmetros foram estimados pelos métodos Monte Carlo Cadeia de Markov (MCMC).

Nas tabelas de resultados apresentadas nesse capítulo adota-se a seguinte convenção de notação:

- i) ANOVA: Análise de varância (i.i.d.).
- ii) PP: Análise de covariância segundo Papadakis.
- iii) VAR-MQG: Variograma seguido de mínimos quadrados generalizados no modelo geoestatístico.
- iv) MV: Estimação de máxima verossimilhança no modelo geoestatístico.
- v) MVR: Estimação de máxima verossimilhança restrita no modelo geoestatístico.
- vi) MCMC: Monte Carlo Cadeia de Markov no modelo Markoviano.

A eficiência de um método de estimação dos contrastes entre os tratamentos em relação a outro foi avaliada, principalmente, por meio da medida de dispersão (D) dada por (58) e do intervalo de cobertura (ICB) conforme 4.2.9.2.

O trabalho foi composto por cinco etapas: Experimento I, Experimento II, Simulação I, Simulação II e Simulação III.

Nos experimentos I e II, descritos nas seções 4.2.5 e 4.2.6 da metodologia, os modelos foram ajustados a dois conjuntos de dados obtidos da literatura, um com 25 e o outro com 500 observações, provenientes de ensaios em branco, nos quais, aos seus valores observados foram adicionados efeitos arbitrários de tratamentos. Dessa forma foi possível obter os verdadeiros valores dos contrastes entre os tratamentos e compará-los aos seus valores estimados a partir dos métodos de estimação de parâmetros abordados nesse estudo.

A fim de se investigar em maior profundidade os resultados obtidos nos Experimentos I e II foram realizadas as Simulações I e II, seguindo as mesmas caraterísticas experimentais.

A Simulação III foi realizada com o objetivo de averiguar se as diferenças entre os tratamentos dependem da alocação dos tratamentos no campo. Para isto foram feitas 5000 diferentes aleatorizações de tratamentos em um mesmo experimento, conforme descrito na seção 4.2.8 da metodologia.

4.1 Resultados do Experimento I

Na Figura 6 estão apresentados os dados do ensaio em branco apresentado por Ver Hoef & Cressie (1993) que refletem a variação espacial natural da área experimental sem os efeitos dos tratamentos. A Figura 7 apresenta os dados desse ensaio em branco, aos quais foram adicionados efeitos arbitrários de tratamentos apresentados na Tabela 2, procedimento que permitiu o conhecimento dos verdadeiros valores dos contrastes entre os tratamentos apresentados em 54.

32	26	24	24	24
26	25	22	22	23
26	21	21	20	24
26	23	26	22	25
25	23	24	24	27

327	¹ 26	² 21	221	⁴ 30
532	⁵ 31	¹ 22	⁴ 28	4 29
529	¹ 21	³ 16	426	⁴ 30
³ 21	1 23	5 32	² 19	3 20
⁵ 31	² 20	³ 19	¹ 24	² 24

Figura 6 - Dados sem os efeitos dos tratamentos

Figura 7 - Dados com a adição dos efeitos dos tratamentos

4.1.1 Ajuste do modelo geoestatístico

Nos diferentes métodos de estimação de parâmetros do ajuste do modelo geoestatístico aos dados do Experimento I, foram comparadas três funções de correlação: a função exponencial, a Matérn com $\kappa=1$ e a Matérn com $\kappa=2$.

A Figura 8 ilustra os três semivariogramas teóricos ajustados aos valores do semivariograma empírico construído a partir dos resíduos da análise clássica, utilizados na estimação dos parâmetros do modelo geoestatístico pelo método do variograma seguido de mínimos quadrados generalizados. O comportamento das curvas apontam para existência de correlação entre os resíduos. Os valores dos resíduos de parcelas mais próximas apresentam menos variabilidade do que os das mais distantes.

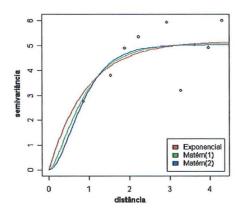


Figura 8 - Ajustes dos variogramas teóricos aos valores do variograma empírico

Na Tabela 4 estão apresentadas as estimativas dos parâmetros de covariância $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2)$ obtidas segundo os métodos do variograma seguido de mínimos quadrados generalizados, estimação de máxima verossimilhança e estimação de máxima verossimilhança restrita, além da soma dos mínimos quadrados ponderados (W_p) dada por (14), da medida de Dispersão (D) dada por (58) e do valor do logaritmo da função de verossimilhança (Log) utilizadas para avaliar a função de correlação mais apropriada. Como pode ser visto, para o método de estimação de parâmetros do variograma seguido de mínimos quadrados o modelo de função de correlação adotada foi a exponencial, que apresentou a menor medida de dispersão (D). De acordo com o logaritmo da função de verossimilhança (Log) e a medida de dispersão (D), para o método de estimação de parâmetros de máxima verossimilhança foi considerada a função Matérn com $\kappa = 2$ e para o método de estimação

de máxima verossimilhança restrita a Matérn com $\kappa = 1$.

Tabela 4. Estimativas dos parâmetros de covariância, $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2)$, obtidas pelos métodos de estimação de parâmetros do modelo geoestatístico utilizando diferentes modelos de funções de correlação, soma dos mínimos quadrados ponderados (W_p) , logaritmo da função de verossimilhança (Log) e medida de dispersão (D)

	Função de correlação	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{ au}^2$	W_p/Log^*	D
VAR-MQG	Exponencial	5,1908	0,9412	0	185,8432	9,1240
	$Matérn(\kappa = 1)$	5,0874	0,5876	0	175,3634	10,3131
	$Matérn(\kappa = 2)$	5,0299	0,3895	0	170,4875	11,3448
MV	Exponencial	6,0818	1,7868	0	-51,34	8,1293
	$Matérn(\kappa = 1)$	6,4570	1,0048	0	-50,81	7,3974
	$\mathbf{Mat\acute{e}rn}(\kappa=2)$	6,0575	0,5537	0	-50,64	6,9690
MVR	Exponencial	13,5526	3,7343	0,0000	-42,22	7,7187
	$\mathbf{Mat\acute{e}rn}(\kappa=1)$	10,5815	1,5027	0,5520	-42,08	7,4205
	$Matérn(\kappa = 2)$	10,0224	1,0626	1,4930	-42,10	7,6396

^{*}Considerando-se Wp para o método VAR-MQG e Log para os métodos MV e MVR.

4.1.2 Ajuste do modelo Markoviano

A partir do ajuste do modelo Markoviano foi possível obter a distribuição marginal a posteriori da média geral μ , dos efeitos dos tratamentos, $\delta = (\delta_1, ..., \delta_5)$ e dos efeitos aleatórios espacialmente estruturados S_i , i=1,...,n. As estimativas a posteriori das médias dos efeitos de tratamentos, δ_k , foram as medidas assumidas para a construção dos contrastes entre os tratamentos das equações 54, substituindo-se α_k por δ_k já que a média geral se anula. Através da análise gráfica dos valores gerados para μ e $\delta = (\delta_1, ..., \delta_5)$ e usando os diagnósticos de convergência,

implementados no módulo CODA do programa R, foi verificado que as condições para a convergência dos parâmetros μ e $\delta = (\delta_1, ..., \delta_5)$ foram satisfeitas. Os traços e as densidades da distribuição marginal a posteriori da média geral e efeitos dos tratamentos estão apresentados no ANEXO A e a autocorrelação das amostras geradas para os parâmetros μ e $\delta = (\delta_1, ..., \delta_5)$ do modelo Markoviano estão apresentadas no ANEXO B.

4.1.3 Comparações entre os modelos ajustados

A Tabela 5 compara as estimativas dos contrastes entre as médias dos tratamentos obtidas pelos métodos de estimação de parâmetros aos seus valores verdadeiros, segundo o valor da medida de dispersão (D) e a Tabela 6, mostra as diferenças entre as médias das estimativas dos contrastes e seus verdadeiros valores. Na Tabela 7 estão apresentados os erros quadráticos referentes às estimativas dos contrastes, obtidas pelos modelos tratados segundo a abordagem freqüentista.

Tabela 5. Valores verdadeiros dos contrastes entre os tratamentos (VERD.), estimativas desses contrastes obtidas por cada método adotado e medida de dispersão (D)

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5	D
VERD.	-4,0000	6,0000	10,0000	2,0000	0,0000	-
ANOVA	-2,4000	6,6000	9,0000	0,4000	-2,4000	12,2400
PP	-2,9802	6,6400	9,6211	0,7012	-2,2090	8,1608
VAR-MQG	-2,7695	6,6715	9,4411	0,4386	-2,0997	9,1240
MV	-2,9108	7,0976	10,0084	0,7837	-1,7603	6,9690
MVR	-2,9790	6,9332	9,9121	0,5945	-1,8773	7,4205
MCMC	-3,8414	5,3636	9,2050	2,5750	-1,2910	3,0595

Tabela 6. Diferenças absolutas entre as estimativas dos contrastes e seus valores verdadeiros

	$ c_1-\hat{c}_1 $	$ c_2-\hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	1,6000	0,6000	1,0000	1,6000	2,4000
PP	1,0198	0,6400	0,3789	1,2988	2,2090
VAR-MQG	1,2304	0,6715	0,5589	1,5613	2,0997
MV	1,0892	1,0976	0,0084	1,2163	1,7603
MVR	1,0210	0,9332	0,0879	1,4054	1,8773
MCMC	0,1586	0,6364	0,7950	0,5750	1,2910

Tabela 7. Erros quadráticos das estimativas dos contrastes calculados pelos métodos de estimação de parâmetros, tratados segundo a abordagem freqüentista

	$eq(\hat{c}_1)$	$eq(\hat{c}_2)$	$eq(\hat{c}_3)$	$eq(\hat{c}_4)$	$eq(\hat{c}_5)$
ANOVA	1,2915	1,2915	1,0545	1,4913	1,4913
PP	0,6761	0,6715	0,5547	0,7765	0,7758
VAR-MQG	0,9875	1,1040	0,8907	1,1823	1,5905
MV	0,6982	0,8871	0,7407	0,9254	1,4737
MVR	0,8224	1,0240	0,8243	1,0176	1,6522

Os modelos espaciais forneceram melhores estimativas aos contrastes entre os tratamentos do que a análise de variância ANOVA (i.i.d.). O modelo Markoviano foi o mais apropriado aos dados do Experimento I, com grande diferença em relação aos demais modelos. O método de estimação de máxima verossimilhança do modelo geoestatístico foi o segundo melhor, cujos resultados ficaram próximos aos obtidos pelos métodos de estimação de parâmetros de máxima verossimilhança restrita, aos do variograma seguido de mínimos quadrados generalizados e aos da análise de covariância de acordo com Papadakis (1937).

Esse estudo confirmou o que Ver Hoef & Cressie (1993) já haviam

observado quando compararam a análise clássica ANOVA ao método de estimação de parâmetros do variograma seguido de mínimos quadrados generalizados. Como mostra a Figura 7, o tratamento 4 foi agrupado no canto superior direito da área em que se encontram as unidades experimentais e o tratamento 5 no canto superior esquerdo dessa área. Os métodos espaciais consideram a aleatorização "pobre" para o contraste c_5 , apresentado em (54), pois os efeitos dos tratamentos podem ser confundidos com a variação local. Por outro lado, para o contraste c_4 , os tratamentos 2 e 3 estão bem dispersos por todas as partes, ou seja, "bem" aleatorizados (Hulbert, 1984). Conseqüentemente, os erros quadráticos segundo os métodos espaciais são menores, exceto para o proposto por Papadakis conforme Tabela 7. Isto ocorre porque a análise clássica ANOVA não reconhece as características singulares desta aleatorização particular, assim não difere os erros quadráticos desses dois contrastes. Também é possível observar que o contraste c_5 está mais afastado do valor verdadeiro do que o c_4 , característica que foi refletida nos erros quadráticos estimados pela análise clássica.

4.2 Resultados do Experimento II

O Experimento II foi realizado para verificar se a performance dos modelos espaciais em conjuntos de dados com maior número de parcelas continuava a mesma que a verificada no experimento de 25 unidades experimentais. Os dados utilizados no Experimento II foram apresentados por Mercer & Hall (1911) e são provenientes de um ensaio em branco com 500 parcelas. Tais dados refletem a variação espacial natural da área experimental. Aos dados desse ensaio em branco também foram adicionados efeitos de tratamentos arbitrários, apresentados na Tabela 3, distribuídos aleatoriamente entre as parcelas.

4.2.1 Ajuste do modelo geoestatístico

Foram comparadas as mesmas funções de correlação utilizadas no ajuste do modelo geoestatístico aos dados do Experimento I, ou seja, a função expo-

nencial, a Matérn com $\kappa = 1$ e a Matérn com $\kappa = 2$.

A Figura 9 ilustra os três modelos de semivariogramas teóricos de acordo com as funções de correlação, ajustados aos valores do semivariograma empírico que foram utilizados pelo método de estimação de parâmetros do variograma seguido de mínimos quadrados generalizados.

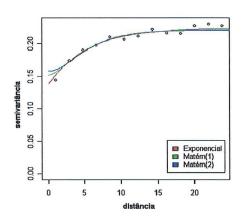


Figura 9 - Ajuste dos variogramas teóricos aos valores do variograma empírico

Conforme a Tabela 8, a função de correlação que forneceu a menor soma dos mínimos quadrados ponderados (W_p) no ajuste do semivariograma foi a exponencial, porém, não foi a que forneceu a menor medida de dispersão (D). Entre essas duas medidas é preferível considerar a medida de dispersão, assim, a função de correlação adotada foi a Matérn com $\kappa=2$. De acordo com o valor máximo do logaritmo da função de verossimilhança (Log) e a medida de dispersão (D), para os métodos de estimação de parâmetros de máxima verossimilhança e máxima verossimilhança restrita a função de correlação que apresentou melhor desempenho foi a exponencial.

Tabela 8. Estimativas dos parâmetros de covariância, $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2)$, obtidas pelos métodos de estimação de parâmetros do modelo geoestatístico utilizando diferentes modelos de funções de correlação, soma dos mínimos quadrados ponderados (W_p) , logaritmo da função de verossimilhança (Log) e medida de dispersão (D)

	Função de correlação	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{ au}^2$	W_p/Log^*	D
VAR-MQG	Exponencial	0,0857	5,5706	0,1383	2,0924	0,0089
	$Matérn(\kappa = 1)$	0,0705	3,7733	0,1514	2,9790	0,0088
	$Matérn(\kappa = 2)$	0,0627	2,5779	0,1579	2,4983	0,0088
MV	Exponencial	0,2054	1,0165	0,0000	-248,4000	0,0059
	$\mathrm{Mat\'ern}(\kappa=1)$	0,2013	0,6064	0,0026	-248,3000	0,0070
	$Matérn(\kappa = 2)$	0,1651	0,4461	0,0380	-248,8000	0,0072
MVR	Exponencial	0,2085	1,0265	0,0000	-247,70	0,0058
	$Matérn(\kappa = 1)$	0,1980	0,6235	0,0085	-247,80	0,0068
	$Matérn(\kappa = 2)$	0,1637	0,4551	0,0419	-248,40	0,0071

^{*}Considerando-se Wp para o método VAR-MQG e Log para os métodos MV e MVR.

4.2.2 Ajuste do modelo Markoviano

Ajustando o modelo Markoviano foram obtidas as estimativas a posteriori das médias de δ_k e então foram construídos os contrastes entre os tratamentos apresentados em (54). Através da análise gráfica dos valores gerados para μ e $\delta = (\delta_1, ..., \delta_5)$ e usando os diagnósticos de convergência, implementados no módulo CODA do programa R, foi verificado que as condições para a convergência dos parâmetros μ e $\delta = (\delta_1, ..., \delta_5)$ foram satisfeitas. O traço e a densidade da distribuição marginal a posteriori da média geral e dos efeitos dos tratamentos estão apresentados no ANEXO C, e a autocorrelação das amostras geradas para os parâmetros μ e $\delta = (\delta_1, ..., \delta_5)$ do modelo Markoviano estão apresentadas no ANEXO D.

4.2.3 Comparações entre os modelos ajustados

A Tabela 9 compara as estimativas dos contrastes entre os tratamentos aos seus verdadeiros valores segundo a medida de dispersão (D). A Tabela 10 mostra as diferenças absolutas entre as médias das estimativas dos contrastes e seus verdadeiros valores. Na Tabela 11 estão apresentados os erros quadráticos referentes às estimativas dos contrastes obtidas pelos métodos de estimação de parâmetros das análises freqüentistas.

Tabela 9. Valores verdadeiros dos contrastes entre os tratamentos (VERD.), estimativas desses contrastes obtidas pelos métodos adotados e medida de dispersão (D)

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5	D
VERD.	-0,7500	1,2000	1,9500	0,3000	0,0000	-
ANOVA	-0,8676	1,1575	2,0251	0,2980	-0,0359	0,0226
PP	-0,8131	1,1761	1,9892	0,2973	-0,0065	0,0061
VAR-MQG	-0,8202	1,1776	1,9978	0,3184	-0,0273	0,0088
MV	-0,8121	1,1719	1,9840	0,3049	0,0074	0,0059
MVR	-0,8119	1,1719	1,9838	0,3048	0,0076	0,0058
MCMC	-0,8063	1,1726	1,9789	0,3172	-0,0230	0,0056

De modo geral, não ocorreram grandes diferenças entre os resultados obtidos nos ajustes dos diferentes modelos. Apesar do modelo Markoviano mais uma vez ter proporcionado a menor medida de dispersão, (D), a diferença em relação aos resultados obtidos pelos demais métodos de estimação de parâmetros foi menor do que as verificadas no Experimento I. O método de estimação de parâmetros de máxima verossimilhança restrita do modelo geoestatístico ocupou a segunda posição entre os melhores e o variograma seguido de mínimos quadrados generalizados foi o menos preciso entre os métodos espaciais de estimação de parâmetros.

Tabela 10. Diferenças absolutas entre os contrastes verdadeiros e a média de seus valores estimados

	$ c_1-\hat{c}_1 $	$ c_2-\hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	0,1176	0,0424	0,0751	0,0020	0,0359
PP	0,0631	0,0239	0,0392	0,0027	0,0065
VAR-MQG	0,0702	0,0224	0,0478	0,0184	0,0273
MV	0,0621	0,0281	0,0340	0,0049	0,0074
MVR	0,0620	0,0281	0,0338	0,0048	0,0076
MCMC	0,0563	0,0274	0,0289	0,0172	0,0230

Tabela 11. Erros quadráticos das estimativas dos contrastes calculados pelos métodos de estimação de parâmetros, tratados segundo a abordagem freqüentista

	$eq(\hat{c}_1)$	$eq(\hat{c}_2)$	$eq(\hat{c}_3)$	$eq(\hat{c}_4)$	$eq(\hat{c}_5)$
ANOVA	0,0560	0,0560	0,0458	0,0647	0,0647
PP	0,0345	0,0344	0,0281	0,0398	0,0398
VAR-MQG	0,0507	0,0507	0,0416	0,0581	0,0578
MV	0,0459	0,0451	0,0340	0,0049	0,0074
MVR	0,0461	0,0453	0,0370	0,0515	0,0518

4.3 Resultados da Simulação I

Para verificar as performances dos modelos ajustados aos dados do Experimento I, foram realizadas três simulações de 25 parcelas, supondo-as dispersas em uma área quadrada, com 5×5 parcelas, conforme modelo (55), referidas aqui como: i)Simulação Ia: com baixa dependência espacial; ii) Simulação Ib: com média dependência espacial e iii) Simulação Ic: com alta dependência espacial. O que diferenciou essas simulações foram os valores considerados ao parâmetro ϕ , que mede a dependência espacial entre as unidades experimentais, que consistiram em: 0,5;

1,5 e 2,5; respectivamente. Os valores adotados aos parâmetros de variâncias foram: $\sigma^2 = 5$ e $\tau^2 = 1$. Em cada simulação foram gerados 5000 conjuntos de dados e portanto 5.000 valores para as seguintes medidas:

- i) Estimativas dos contrastes;
- ii) Medidas de dispersão;
- iii) Erros quadráticos e
- iii) Intervalo de confiança.

Para resumir os resultados obtidos nas simulações foram calculados os seguintes valores:

- i) as médias das amostras das estimativas dos contrastes entre os tratamentos;
- ii) a diferença absoluta entre os valores verdadeiros dos contrastes e a média das estimativas desses contrastes;
- iii) as médias das medidas de dispersão;
- iv) os desvios padrões das amostras das medidas de dispersão;
- v) as médias das amostras dos erros quadráticos calculados analiticamente;
- vi) os erros quadráticos obtidos por simulação de Monte Carlo e
- vii) o intervalo de cobertura de cada contraste, c_j , j=1,...,5, e do vetor de contrastes, $c=(c_1,...,c_5)$.

Foram ajustados nessas simulações os seguintes modelos:

- i) Modelo clássico ANOVA (i.i.d);
- ii) O modelo proposto por Papadakis e
- iii) O modelo Geoestatístico.

O modelo Markoviano foi ajustado somente a um conjunto de dados escolhido aleatoriamente entre os três mil gerados a cada nível de dependência espacial, porque o programa utilizado para ajustar esse modelo não é adequado para a automatização dos procedimentos.

4.3.1 Simulação Ia: com baixa dependência espacial

Tabela 12. Médias das estimativas dos contrastes resultantes da simulação com baixa dependência espacial

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,0458	5,9713	10,0171	1,9961	0,03016
PP	-4,0409	5,9774	10,0184	1,9960	0,0397
VAR-MQG	-4,0467	5,9722	10,0189	1,9981	0,0302
MV	-4,0438	5,9738	10,0177	1,9978	0,0335
MVR	-4,0467	5,9752	10,0219	1,9972	0,0348

Tabela 13. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com baixa dependência espacial

	$ c_1 - \overline{\hat{c}_1} $	$ c_2-\overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	0,0458	0,0287	0,0171	0,0039	0,0302
PP	0,0409	0,0226	0,0184	0,0041	0,0397
VAR-MQG	0,0467	0,0278	0,0189	0,0018	0,0302
MV	0,0438	0,0262	0,0177	0,0022	0,0335
MVR	0,0467	0,0248	0,0219	0,0027	0,0348

Tabela 14. Médias e desvios padrões da medida de dispersão (D), resultantes da simulação com baixa dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	9,4903	10,4434	9,4999	9,5611	9,6318
eq(D)	6,9174	7,6468	6,9017	6,9691	7,0294

Tabela 15. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente, da simulação com baixa dependência espacial

	$\overline{eq(c_1)}$	$\overline{eq(c_2)}$	$\overline{eq(c_3)}$	$\overline{eq(c_4)}$	$\overline{eq(c_5)}$
ANOVA	1,3071	1,3071	1,0672	1,5093	1,5093
PP	1,0775	1,0776	0,8803	1,2440	1,2445
VAR-MQG	1,1547	1,1548	0,9422	1,3325	1,3330
MV	1,1278	1,1278	0,92098	1,3019	1,3018
MVRE	1,2514	1,2515	1,0215	1,4439	1,4442

Tabela 16. Erros quadráticos das estimativas dos contrastes obtidos por simulação de Monte Carlo da simulação com baixa dependência espacial

	$eq(c_1)$	$eq(c_2)$	$eq(c_3)$	$eq(c_4)$	$eq(c_5)$
ANOVA	1,3269	1,3251	1,1142	1,5517	1,5239
PP	1,3781	1,3825	1,1637	1,6389	1,6095
VAR-MQG	1,3185	1,3222	1,1112	1,5627	1,5278
MV	1,3262	1,3265	1,1171	1,5633	1,5325
MVRE	1,3270	1,3283	1,1195	1,5722	1,5424

Tabela 17. Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados por (54), resultantes da simulação com baixa dependência espacial

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9474	0,9506	0,9430	0,9470	0,9514	0,801
PP	0,8798	0,8794	0,8650	0,8716	0,8720	0,5942
VAR-MQG	0,8988	0,9014	0,8858	0,8940	0,8976	0,6416
MV	0,8918	0,8936	0,8796	0,8838	0,8906	0,6212
MVRE	0,9238	0,9244	0,9152	0,9174	0,9214	0,7154

4.3.2 Simulação Ib: média dependência espacial

Tabela 18. Médias das estimativas dos contrastes resultantes da simulação com média dependência espacial

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,0488	5,9674	10,0161	1,9981	0,0172
PP	-4,0320	5,9812	10,0132	2,0017	0,0255
VAR-MQG	-4,0379	5,9793	10,0171	2,0041	0,0188
MV	-4,0382	5,9777	10,0159	2,0031	0,0233
MVR	-4,0385	5,9772	10,0157	2,0017	0,0201

Tabela 19. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com média dependência espacial

	$ c_1 - \overline{\hat{c}_1} $	$ c_2 - \overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	0,0488	0,0326	0,0161	0,0019	0,0172
PP	0,0320	0,0188	0,0132	0,0017	0,0255
VAR-MQG	0,0379	0,0207	0,0171	0,0041	0,0188
MV	0,0382	0,0223	0,0159	0,0031	0,0233
MVR	0,0385	0,0228	0,0157	0,0017	0,0201

Tabela 20. Médias e desvios padrões da medida de dispersão (D), resultantes da simulação com média dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	7,9245	6,5601	6,5360	6,5445	6,4694
eq(D)	6,1204	4,8786	4,9266	4,8894	4,8033

Tabela 21. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente, da simulação com média dependência espacial

	$\overline{eq(c_1)}$	$\overline{eq(c_2)}$	$\overline{eq(c_3)}$	$\overline{eq(c_4)}$	$\overline{eq(c_5)}$
ANOVA	1,1903	1,1903	0,9719	1,3744	1,3744
PP	1,2403	1,2409	1,0134	1,4323	1,4328
VAR-MQG	0,9964	0,9963	0,8124	1,1491	1,1496
MV	0,9314	0,9310	0,7596	1,0744	1,0744
MVR	1,0268	1,0263	0,8373	1,1842	1,1844

Tabela 22. Erros quadráticos das estimativas dos contrastes obtidos por simulação de Monte Carlo, simulação com média dependência espacial

	$eq(c_1)$	$eq(c_2)$	$eq(c_3)$	$eq(c_4)$	$eq(c_5)$
ANOVA	1,2208	1,2087	1,0131	1,4138	1,3948
PP	1,1002	1,0914	0,9157	1,3021	1,2742
VAR-MQG	1,1007	1,0888	0,9154	1,2936	1,2753
MV	1,0977	1,0892	0,9141	1,2974	1,2780
MVRE	1,0909	1,0807	0,9104	1,2909	1,2707

Tabela 23. Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados por (54), resultantes da simulação com média dependência espacial

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9484	0,9464	0,9484	0,9456	0,9490	0,7974
PP	0,8706	0,8744	0,8616	0,8606	0,8672	0,5744
VAR-MQG	0,9050	0,9134	0,9064	0,9042	0,9080	0,6784
MV	0,8894	0,8940	0,8870	0,8858	0,8882	0,6202
MVRE	0,9202	0,928	0,9166	0,9166	0,9224	0,7092

4.3.3 Simulação Ic: com alta dependência espacial

Tabela 24. Médias das estimativas dos contrastes resultante da simulação com alta dependência espacial

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,0451	5,9683	10,0135	1,9996	0,01307
PP	-4,0286	5,9826	10,0112	2,0033	0,0199
VAR-MQG	-4,0312	5,9814	10,0126	2,0039	0,0145
MV	-4,0316	5,9802	10,0118	2,0035	0,0159
MVR	-4,0323	5,9801	10,0124	2,0028	0,0136

Tabela 25. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com alta dependência espacial

	$ c_1-\overline{\hat{c}_1} $	$ c_2-\overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	0,04512	0,03164	0,01348	0,00024	0,01307
PP	0,0286	0,0174	0,0112	0,0033	0,0199
VAR-MQG	0,0312	0,0186	0,0126	0,0039	0,0145
MV	0,0316	0,0198	0,0118	0,0035	0,0159
MVR	0,0323	0,0199	0,0124	0,0028	0,0136

Tabela 26. Médias e desvios padrões da medida de dispersão (D), resultantes da simulação com alta dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	6,6007	5,0447	5,0160	5,0197	4,9396
eq(D)	5,2478	3,7668	3,7970	3,7764	3,6708

Tabela 27. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente, da simulação com alta dependência espacial

	$\overline{eq(c_1)}$	$\overline{eq(c_2)}$	$\overline{eq(c_3)}$	$\overline{eq(c_4)}$	$\overline{eq(c_5)}$
ANOVA	1,0855	1,0855	0,8863	1,2534	1,2534
PP	1,1309	1,1315	0,9240	1,3061	1,3064
VAR-MQG	0,8818	0,8816	0,7189	1,0170	1,0174
MV	0,8177	0,8175	0,6668	0,9432	0,9433
MVR	0,9007	0,9005	0,7344	1,0385	1,0388

Tabela 28. Erros quadráticos das estimativas dos contrastes obtidos por simulação de Monte Carlo da simulação com alta dependência espacial

	$eq(c_1)$	$eq(c_2)$	$eq(c_3)$	$eq(c_4)$	$eq(c_5)$
ANOVA	1,1154	1,1041	0,9223	1,2899	1,2732
PP	0,96708	0,9558	0,8004	1,1426	1,1176
VAR-MQG	0,9662	0,9529	0,7987	1,1350	1,1168
MV	0,9628	0,9518	0,7994	1,1378	1,1192
MVR	0,9553	0,9427	0,7927	1,1304	1,1097

Tabela 29. Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados por (54), resultantes da simulação com alta dependência espacial

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9484	0,9478	0,9504	0,9448	0,9504	0,8006
PP	0,8432	0,8506	0,8362	0,8380	0,8422	0,5116
VAR-MQG	0,9484	0,9478	0,9504	0,9448	0,9504	0,6854
MV	0,9460	0,8898	0,8966	0,8890	0,8870	0,6250
MVR	0,9218	0,9288	0,9212	0,9192	0,9214	0,7142

Nestas simulações foi possível constatar que quanto maior a dependência espacial entre as parcelas, maior a precisão das estimativas dos contrastes. Os métodos espaciais de estimação de parâmetros frequentistas apresentaram resultados muito próximos. Em comparação à análise clássica foram mais precisos quando ajustados aos dados com alta e média dependências espaciais, para os dados com baixa dependência espacial a análise clássica foi mais eficiente, o valor do parâmetro $\phi = 0,5$ não gerou dados com dependência espacial que justificasse tal análise. Ocorreram divergências entre as medidas de dispersão e intervalo de cobertura, utilizadas para comparar o desempenho dos métodos de estimação de parâmetros. Por exemplo, a análise clássica ANOVA (i.i.d), nos dados com média e alta dependência espacial, apresentou o melhor intervalo de cobertura, mas em compensação a pior medida de dispersão. Já o método de estimação de máxima verossilhança restrita forneceu a menor medida de dispersão e o segundo melhor intervalo de cobertura. Baseando-se nesses resultados o modelo geoestatístico junto ao método de máxima verossimilhança restrita foi o mais eficiente entre os modelos abordados segundo a Simulação I.

4.3.4 Análise frequentista versus Bayesiana - Simulação I

O modelo Markoviano que foi tratado segundo a análise Bayesiana foi ajustado a um conjunto de dados extraído da Simulação Ia:com baixa dependência espacial, a um da Simulação Ib: com média dependência espacial e a um da Simulação Ic: com alta dependência espacial. Somente para a simulação com alta dependência espacial os modelos espaciais foram mais eficientes do que a análise clássica. O modelo Markoviano perdeu em precisão para o modelo geoestatístico, contudo, seus resultados foram semelhantes. Os resultados obtidos na análise freqüentista versus Bayesiana, realizada a partir do conjunto de dados extraído da Simulação Ib: com média dependência espacial, não conferem com os obtidos nessa simulação, em que foram gerados e analisados 5000 conjuntos de dados, indicando os modelos espaciais mais precisos do que a análise clássica ANOVA.

4.3.4.1 Análise freqüentista versus Bayesiana - Simulação Ia

Tabela 30. Estimativas dos contrastes resultantes da simulação com baixa dependência espacial

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,1741	6,7100	10,8841	-0,1367	1,1863
PP	-4,2028	5,9030	10,1058	-1,1520	1,1659
VAR-MQG	-4,2754	6,2937	10,5692	-0,5268	1,2792
MV	-4,2243	6,2953	10,5197	-0,5043	1,2355
MVRE	-4,2164	6,2253	10,4417	-0,6577	1,2140
MCMC	-4,1550	6,8610	11,0160	-0,7760	1,3160

Tabela 31. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com baixa dependência espacial

	$ c_1-\hat{c}_1 $	$ c_2 - \hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	0,1741	0,7100	0,8841	2,1367	1,1863
PP	0,2028	0,0970	0,1058	3,1520	1,1659
VAR-MQG	0,2754	0,2937	0,5692	2,5268	1,2792
MV	0,2243	0,2953	0,5197	2,5043	1,2355
MVRE	0,2164	0,2253	0,4417	2,6577	1,2140
MCMC	0,1550	0,8610	1,0160	2,7760	1,3160

Tabela 32. Medidas de dispersão (D), resultantes da simulação com baixa dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR	MCMC
D	7,2890	8,5074	11,3561	8,2056	8,8298	11,2356

4.3.4.2 Análise frequentista versus Bayesiana - Simulação Ib

Tabela 33. Estimativas dos contrastes resultantes da simulação com média dependência espacial

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-3,9290	5,0229	8,9519	0,2452	-1,1873
PP	-3,8018	5,1038	8,9057	-0,5070	-2,6331
VAR-MQG	-3,9716	4,9640	8,9357	-0,2370	-1,8796
MV	-3,9223	5,0142	8,9364	0,0876	-1,4755
MVRE	-3,9246	4,9958	8,9204	-0,0363	-1,7182
MCMC	-4,0441	4,8819	8,9260	0,0980	-1,5660

Tabela 34. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com média dependência espacial

	$ c_1-\hat{c}_1 $	$ c_2 - \hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	0,0710	0,9771	1,0481	1,7548	1,1873
PP	0,1981	0,8962	1,0943	2,5070	2,6331
VAR-MQG	0,0284	1,0360	1,0643	2,2370	1,8796
MV	0,0777	0,9858	1,0636	1,9124	1,4755
MVRE	0,0754	1,0042	1,0796	2,0363	1,7183
MCMC	0,0441	1,1181	1,0740	1,9020	1,5660

Tabela 35. Medidas de dispersão (D), resultantes da simulação com média dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR	MCMC
D	6,5476	15,2583	10,7442	7,9432	9,2784	8,4755

4.3.4.3 Análise freqüentista versus Bayesiana - Simulação Ic

Tabela 36. Estimativas dos contrastes resultantes da simulação com alta dependência espacial

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-2,6380	4,5603	7,1983	2,2523	-0,3023
PP	-2,8322	5,2211	8,0532	2,0573	-0,4914
VAR-MQG	-2,9650	5,3554	8,3204	2,1636	-0,2705
MV	-2,9604	5,3434	8,3038	2,1722	-0,2723
MVRE	-2,9855	5,3730	8,3585	2,1697	-0,2531
MCMC	-3,0397	4,9968	8,0365	1,9670	-0,1280

Tabela 37. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com alta dependência espacial

	$ c_1-\hat{c}_1 $	$ c_2-\hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	1,3619	1,4397	2,8017	0,2522	0,3023
PP	1,1678	0,7789	1,9467	0,0573	0,4914
VAR-MQG	1,0350	0,6445	1,6796	0,1636	0,2705
MV	1,0396	0,6566	1,6962	0,1722	0,2723
MVRE	1,0145	0,6270	1,6415	0,1697	0,2531
MCMC	0,9603 v1,0032	1,9635	0,0330	0,1280	

Tabela 38. Medidas de dispersão (D), resultantes da simulação de dados com alta dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR	MCMC
D	11,9321	6,0051	4,4076	4,4929	4,2096	5,8014

4.4 Resultados da Simulação II

Para verificar as performances dos modelos ajustados aos dados do Experimento II, foram realizadas três simulações de 225 parcelas, supondo-as dispersas em uma área quadrada, com 15×15 parcelas conforme modelo (55). Tais simulações foram referidas como: i) Simulação IIa: com baixa dependência espacial, ii) Simulação IIb: com média dependência espacial e iii) Simulação IIc: com alta dependência espacial. O que diferenciou essas simulações foram os valores atribuídos ao parâmetro ϕ , que mede a dependência espacial entre as unidades experimentais, que consistiram em: 2; 4 e 6; respectivamente. Os valores adotados aos parâmetros de variâncias foram: $\sigma^2 = 5$ e $\tau^2 = 1$. Em cada simulação foram gerados 3000 conjuntos de dados. Assim, como na Simulação I, foram obtidas a partir dos 3000 conjuntos de dados gerados as medidas que se seguem:

- i) Estimativas dos contrastes;
- ii) Medidas de dispersão;
- iii) Erros quadráticos e
- iii) Intervalo de confiança.

Para resumir os resultados foram calculados os seguintes valores:

- i) as médias das amostras das estimativas dos contrastes entre os tratamentos;
- ii) a diferença absoluta entre os valores verdadeiros dos contrastes e a média das estimativas desses contrastes;
- iii) as médias das amostras das medidas de dispersão;
- iv) os desvios padrões das amostras das medidas de dispersão;
- v) as médias das amostras dos erros quadráticos calculados analiticamente;
- vi) os erros quadráticos obtidos por simulação de Monte Carlo e

vii) o intervalo de cobertura de cada contraste, c_j : j=1,...,5, e do vetor de contrastes, $c=(c_1,...,c_5)$.

Foram ajustados nessas simulações os seguintes modelos:

- i) Modelo clássico ANOVA (i.i.d);
- ii) O modelo proposto por Papadakis e
- iii) O modelo Geoestatístico.

O modelo Markoviano também foi ajustado a um conjunto de dados escolhido aleatoriamente entre os três mil gerados em cada nível de dependência espacial, porque o programa utilizado para ajustar esse modelo não é adequado para a automatização dos procedimentos.

Os resultados a cerca da Simulação II estão apresentados nas tabelas que se seguem.

4.4.1 Simulação IIa

Tabela 39. Médias das estimativas dos contrastes resultantes da simulação com baixa dependência espacial

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,0027	5,9982	10,0009	2,0153	-0,0045
PP	-3,9999	6,0009	10,0008	2,0042	0,0027
VAR-MQG	-4,0011	5,9973	9,9984	2,0078	0,0083
MV	-4,0006	5,9996	10,0003	2,0026	0,0011
MVR	-4,0004	5,9997	10,0002	2,0026	0,0012

Tabela 40. Diferenças absolutas entre os valores verdadeiros e as estimativas da simulação com baixa dependência espacial

	$ c_1 - \overline{\hat{c}_1} $	$ c_2 - \overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	0,0027	0,0018	0,0009	0,0153	0,0045
PP	0,0001	0,0009	0,0008	0,0042	0,0027
VAR-MQG	0,0011	0,0027	0,0016	0,0078	0,0083
MV	0,0006	0,0003	0,0003	0,0026	0,0011
MVRE	0,0004	0,0003	0,0002	0,0026	0,0012

Tabela 41. Médias e desvios padrões da medida de dispersão (D), resultantes da simulação com baixa dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	1,0065	0,5638	0,6183	0,5524	0,5526
eq(D)	0,7447	0,4103	0,4730	0,4027	0,4031

Tabela 42. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente, da simulação com baixa dependência espacial

	$\overline{eq(\hat{c}_1)}$	$\overline{eq(\hat{c}_2)}$	$\overline{eq(\hat{c}_3)}$	$\overline{eq(\hat{c}_4)}$	$\overline{eq(\hat{c}_5)}$
ANOVA	0,4309	0,4309	0,3518	0,4975	0,4975
PP	0,2557	0,2557	0,2088	0,2953	0,2953
VAR-MQG	0,3585	0,3584	0,2926	0,4141	0,4140
MV	0,3185	0,3185	0,2599	0,3678	0,3678
MVR	0,3218	0,3218	0,2626	0,3716	0,3716

Tabela 43. Erros quadráticos das estimativas dos contrastes obtidos por simulação de Monte Carlo da simulação com baixa dependência espacial

<u> </u>	$eq(\hat{c}_1)$	$eq(\hat{c}_2)$	$eq(\hat{c}_3)$	$eq(\hat{c}_4)$	$eq(\hat{c}_5)$
ANOVA	0,4312	0,4319	0,3448	0,5020	0,5130
PP	0,3242	0,3214	0,2606	0,3772	0,3813
VAR-MQG	0,3392	0,3380	0,2727	0,3923	0,4009
MV	0,3222	0,3185	0,2591	0,3726	0,3759
MVR	0,3225	0,3185	0,2591	0,3726	0,3760

Tabela 44. Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados por (54), resultantes da simulação com baixa dependência espacial

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9490	0,9510	0,9583	0,9520	0,9457	0,7967
PP	0,8763	0,8803	0,8813	0,8803	0,8767	0,5707
VAR-MQG	0,9593	0,9550	0,9650	0,9630	0,9570	0,8360
MV	0,9460	0,9493	0,9483	0,9467	0,9447	0,7840
MVR	0,9487	0,9520	0,9513	0,9480	0,9493	0,7950

4.4.2 Simulação IIb

Tabela 45. Médias das estimativas dos contrastes resultantes da simulação com média dependência espacial

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,0034	5,9990	10,0025	2,0171	-0,0050
PP	-4,0012	6,0004	10,0016	2,0046	0,0014
VAR-MQG	-4,0025	5,9979	10,0004	2,0034	-0,0017
MV	-4,0012	5,9996	10,0008	2,0025	0,0001
MVR	-4,0010	5,9997	10,0007	2,0025	0,0002

Tabela 46. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com média dependência espacial

	$ c_1 - \overline{\hat{c}_1} $	$ c_2 - \overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5 - \overline{\hat{c}_5} $
ANOVA	0,0034	0,0010	0,0025	0,0171	0,0050
PP	0,0012	0,0004	0,0016	0,0046	0,0014
VAR-MQG	0,0025	0,0021	0,0004	0,0034	0,0017
MV	0,0012	0,0004	0,0008	0,0025	0,0001
MVR	0,0010	0,0003	0,0007	0,0025	0,0002

Tabela 47. Médias e desvios padrões da medida de dispersão (D), resultantes da simulação com média dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	0,8911	0,4070	0,4021	0,3880	0,3880
eq(D)	0,6992	0,3005	0,3038	0,2866	0,2865

Tabela 48. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente, da simulação com média dependência espacial

	$\overline{eq(\hat{c}_1)}$	$\overline{eq(\hat{c}_2)}$	$\overline{eq(\hat{c}_3)}$	$\overline{eq(\hat{c}_4)}$	$\overline{eq(\hat{c}_5)}$
ANOVA	0,4028	0,4028	0,3288	0,4651	0,4651
PP	0,2168	0,2167	0,1769	0,2502	0,2502
VAR-MQG	0,2884	0,2884	0,2354	0,3331	0,3331
MV	0,2671	0,2671	0,2180	0,3085	0,3084
MVR	0,2698	0,2698	0,2202	0,3117	0,3116

Tabela 49. Erros quadráticos das estimativas dos contrastes obtidos por simulação de Monte Carlo na simulação com média dependência espacial

	$eq(\hat{c}_1)$	$eq(\hat{c}_2)$	$eq(\hat{c}_3)$	$eq(\hat{c}_4)$	$eq(\hat{c}_5)$
ANOVA	0,4041	0,4070	0,3206	0,4785	0,4799
PP	0,2753	0,2733	0,2194	0,3212	0,3244
VAR-MQG	0,2756	0,2720	0,2191	0,3170	0,3221
MV	0,2703	0,2665	0,2163	0,3128	0,3153
MVR	0,2704	0,2664	0,2163	0,3127	0,3153

Tabela 50. Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados por (54), resultantes da simulação com média dependência espacial

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9493	0,9493	0,9607	0,9493	0,9473	0,7973
PP	0,8810	0,8790	0,8830	0,8783	0,8757	0,5740
VAR-MQG	0,9567	0,9540	0,9630	0,9587	0,9573	0,8330
MV	0,9463	0,9487	0,9483	0,9457	0,9487	0,7903
MVR	0,9500	0,9517	0,9507	0,9477	0,9503	0,7993

4.4.3 Simulação IIc

Tabela 51. Médias das estimativas dos contrastes resultantes da simulação com alta dependência espacial

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-4,0035	5,9990	10,0025	2,0168	-0,0045
PP	-4,0019	6,0001	10,0020	2,0045	0,0009
VAR-MQG	-4,0025	5,9982	10,0007	2,0011	-0,0013
MV	-4,0014	5,9996	10,0010	2,0024	-0,0003
MVR	-4,0012	5,9997	10,0009	2,0024	-0,0002

Tabela 52. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, resultantes da simulação com alta dependência espacial

	$ c_1 - \overline{\hat{c}_1} $	$ c_2-\overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	0,0035	0,0010	0,0025	0,0168	0,0045
PP	0,0019	0,0001	0,0098	0,0045	0,0009
VAR-MQG	0,0025	0,0018	0,0007	0,0011	0,0013
MV	0,0014	0,0004	0,0010	0,0024	0,0003
MVR	0,0012	0,0003	0,0009	0,0024	0,0002

Tabela 53. Médias e desvios padrões da medida de dispersão (D), resultantes da simulação com alta dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	0,7860	0,3480	0,3351	0,3272	0,327
eq(D)	0,6359	0,2586	0,2516	0,2427	0,2426

Tabela 54. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente, da simulação com alta dependência espacial

	$\overline{eq(\hat{c}_1)}$	$\overline{eq(\hat{c}_2)}$	$\overline{eq(\hat{c}_3)}$	$\overline{eq(\hat{c}_4)}$	$\overline{eq(\hat{c}_5)}$
ANOVA	0,3774	0,3774	0,3081	0,4358	0,4358
PP	0,2008	0,2008	0,1640	0,2319	0,2319
VAR-MQG	0,2554	0,2553	0,2084	0,2949	0,2949
MV	0,2455	0,2455	0,2004	0,2836	0,2836
MVRE	0,2481	0,2481	0,2025	0,2865	0,2865

Tabela 55. Erros quadráticos das estimativas dos contrastes obtidos por simulação de Monte Carlo na simulação com alta dependência espacial

	$eq(\hat{c}_1)$	$eq(\hat{c}_2)$	$eq(\hat{c}_3)$	$eq(\hat{c}_4)$	$eq(\hat{c}_5)$
ANOVA	0,3792	0,3823	0,3002	0,4519	0,4491
PP	0,2546	0,2530	0,2022	0,2976	0,2996
VAR-MQG	0,2517	0,2480	0,1999	0,2907	0,2929
MV	0,2482	0,2447	0,1982	0,2876	0,2895
MVRE	0,2482	0,2445	0,1983	0,2876	0,2895

Tabela 56. Intervalos de cobertura de cada contraste e do vetor dos contrastes, dados por (54), resultantes da simulação com alta dependência espacial

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9510	0,9483	0,9610	0,9470	0,9467	0,7927
PP	0,8790	0,8810	0,8877	0,8743	0,8777	0,5800
VAR-MQG	0,9487	0,9497	0,9537	0,9513	0,9520	0,8100
MV	0,9460	0,9483	0,9480	0,9443	0,9493	0,7883
MVRE	0,9477	0,9513	0,9503	0,9480	0,9520	0,7983

Como ocorrido na Simulação I, existiram divergências entre a medida de dispersão e o intervalo de cobertura utilizados para comparar os métodos de estimação dos parâmetros. Por exemplo, independente do nível de dependência espacial, o método do variograma seguido de mínimos quadrados apresentou o melhor intervalo de cobertura e medidas de dispersão muito próximas aos valores obtidos pelos métodos de máxima verossilhança e de máxima verossimilhança restrita, que por sua vez apresentaram as menores medidas de dispersão e o segundo melhor intervalo de cobertura. Entretanto, o modelo geoestatístico utilizando os métodos de estimação de parâmetros de máxima verossimilhança e máxima verossimilhança restrita é preferível uma vez que o variograma seguido de mínimos quadrados generalizados é um método mais subjetivo. As estimativas dos contrastes obtidas nas simulações de 25 parcelas foram menos precisas em comparação às obtidas nas simulações de 225 parcelas. Ou seja, o aumento no tamanho da amostra também aumentou a precisão das estimativas dos contrastes entre os tratamentos. Quanto aos diferentes níveis de dependência, como verificado na Simulação I, quanto maior a dependência espacial entre as parcelas maior a precisão das estimativas para os contrastes entre os tratamentos.

4.4.4 Análise frequentista versus Bayesiana - Simulação II

Nesta simulação o modelo Markoviano proporcionou boas estimativas para os contrastes entre os tratamentos. Segundo a medida de dispersão, nos dados com baixa dependência espacial a análise Bayesiana ocupou o quarto lugar entre os seis métodos de estimação de parâmetros adotados, Tabela 59, nos dados com médio grau de dependência ficou na primeira posição, Tabela 62, e nos dados com alta dependência espacial ocupou o segundo lugar, Tabela 65.

4.4.4.1 Análise freqüentista versus Bayesiana - Simulação IIa

Tabela 57. Estimativas dos contrastes obtidas da simulação com baixo grau de dependência espacial

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_{5}
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-2,8712	7,0152	9,8865	2,1915	0,1642
PP	-3,3242	6,4980	9,8222	2,1045	0,0529
VAR-MQG	-3,1176	6,6054	9,7231	2,2684	-0,0731
MV	-3,3153	6,4496	9,7649	2,1213	-0,0723
MVR	-3,3136	6,4446	9,7582	2,1275	-0,0806
MCMC	-3,1920	6,4665	9,6585	2,1800	0,0150

Tabela 58. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, obtidas da simulação com baixo grau de dependência espacial

	$ c_1-\overline{\hat{c}_1} $	$ c_2 - \overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	1,1288	1,0152	0,1135	0,1915	0,1642
PP	0,6758	0,4980	0,1778	0,1045	0,0529
VAR-MQG	0,8824	0,6054	0,2769	0,2684	0,0731
MV	0,6847	0,4496	0,2350	0,1213	0,07232
MVR	0,6864	0,4446	0,2418	0,1275	0,0806
MCMC	0,8080	0,4665	0,3415	0,1800	0,0150

Tabela 59. Medidas de dispersão (D), obtidas da simulação com baixo grau de dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR	MCMC
\overline{D}	2,3814	0,7501	1,2992	0,7461	0,7500	1,0197

4.4.4.2 Análise frequentista versus Bayesiana - Simulação IIb

Tabela 60. Estimativas dos contrastes obtidas da simulação com médio grau de dependência espacial

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-2,8861	6,7710	9,6573	2,4676	0,0697
PP	-3,7177	6,2551	9,9728	1,8787	0,0902
VAR-MQG	-3,6497	6,185	9,8348	1,8216	0,1223
MV	-3,7957	6,1978	9,9935	1,7964	0,1099
MVR	-3,7881	6,1944	9,9825	1,7903	0,1121
MCMC	-3,8095	6,1210	9,9305	1,8850	0,1500

Tabela 61. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, obtidas da simulação com médio grau de dependência espacial

	$ c_1-\hat{c}_1 $	$ c_2-\hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	1,1139	0,7712	0,3427	0,4676	0,0697
PP	0,2823	0,2551	0,0272	0,1213	0,0902
VAR-MQG	0,3503	0,1851	0,1652	0,1784	0,1223
MV	0,2043	0,1978	0,0066	0,2037	0,1099
MVR	0,2119	0,1944	0,0175	0,2097	0,1121
MCMC	0,1905	0,1210	0,0695	0,1150	0,1500

Tabela 62. Medidas de dispersão (D), obtidas da simulação com médio grau de dependência espacial

-	ANOVA	PP	VAR-MQG	MV	MVR	MCMC
\overline{D}	2,1763	0,1684	0,2310	0,1344	0,1396	0,0915

4.4.4.3 Análise freqüentista versus Bayesiana - Simulação IIc

Tabela 63. Estimativas dos contrastes obtidas da simulação com alto grau de dependência espacial

	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-3,6230	6,3125	9,9355	1,5171	-0,6001
PP	-3,8439	6,2050	10,0490	1,7863	-0,7555
VAR-MQG	-3,8438	6,1694	10,0132	1,7806	-0,8405
MV	-3,8517	6,1753	10,0270	1,8238	-0,8208
MVRE	-3,8474	6,1745	10,0218	1,8091	-0,8296
MCMC	-3,8623	6,1557	10,0180	1,7320	-0,7860

Tabela 64. Diferenças absolutas entre os valores verdadeiros e as estimativas dos contrastes, obtidas da simulação com alto grau de dependência espacial

	$ c_1-\hat{c}_1 $	$ c_2 - \hat{c}_2 $	$ c_3-\hat{c}_3 $	$ c_4-\hat{c}_4 $	$ c_5-\hat{c}_5 $
ANOVA	0,3770	0,3125	0,0645	0,4829	0,6001
PP	0,1560	0,2050	0,0489	0,2137	0,7555
VAR-MQG	0,1562	0,1694	0,0132	0,2194	0,8405
MV	0,1483	0,1753	0,0270	0,1761	0,8208
MVR	0,1526	0,1745	0,0218	0,1909	0,8295
MCMC	0,1377	0,1557	0,0180	0,2680	0,7860

Tabela 65. Medidas de dispersão (D), obtidas da simulação com alto grau de dependência espacial

	ANOVA	PP	VAR-MQG	MV	MVR	MCMC
\overline{D}	0,8372	0,6852	0,8079	0,7582	0,7788	0,7331

4.5 Resultados da Simulação III

Essa simulação foi realizada a fim de investigar se os efeitos dos tratamentos podem ser confundidos com a variação espacial da área como indicada pelos resultados obtidos no Experimento I. Por isso, foram realizadas 5000 aleatorizações dos tratamentos de queimadas, apresentados na Tabela 2, e adicionados seus respectivos efeitos no ensaio em branco apresentado por Ver Hoef & Cressie (1993). Nessa simulação foram gerados 5000 valores para as seguintes medidas:

- i) estimativas dos contrastes entre os tratamentos;
- ii) médias dos erros quadráticos das estimativas dos contrastes;
- iii) média das medidas de dispersão e
- iv) intervalo de confiança.

A partir desses valores foram calculados:

- i) as médias das amostras das estimativas dos contrastes entre os tratamentos;
- ii) as diferenças absolutas entre os valores verdadeiros dos contrastes e a média das estimativas desses contrastes;
- iii) as médias das amostras das medidas de dispersão;
- iv) os desvios padrões das medidas de dispersão;
- v) as médias das amostras dos erros quadráticos calculados analiticamente;
- vi) os erros quadráticos obtidos por simulação de Monte Carlo e
- vii) o intervalo de cobertura de cada contraste, c_j : j = 1,...,5, e do vetor de contrastes, $c = (c_1,...,c_5)$.

Os resultados dessas medidas estão apresentados nas tabelas que se seguem:

Tabela 66. Médias das estimativas dos contrastes entre os tratamentos

	$\overline{\hat{c}_1}$	$\overline{\hat{c}_2}$	$\overline{\hat{c}_3}$	$\overline{\hat{c}_4}$	$\overline{\hat{c}_5}$
VERD	-4,0000	6,0000	10,0000	2,0000	0,0000
ANOVA	-3,9923	6,0154	10,0076	1,9622	-0,0011
PP	-3,9990	6,0062	10,0052	1,9808	-0,0126
VAR-MQG	-3,9961	6,0094	10,0056	1,9705	-0,0090
MV	-3,9975	6,0076	10,0051	1,9737	-0,0117
MVR	-3,9990	6,0062	10,0052	1,9760	-0,0125

Tabela 67. Diferenças absolutas entre os valores verdadeiros e a média das estimativas dos contrastes

	$ c_1 - \overline{\hat{c}_1} $	$ c_2-\overline{\hat{c}_2} $	$ c_3-\overline{\hat{c}_3} $	$ c_4-\overline{\hat{c}_4} $	$ c_5-\overline{\hat{c}_5} $
ANOVA	0,0077	0,0154	0,0076	0,0378	0,0011
PP	0,0010	0,0062	0,0052	0,0192	0,0126
VAR-MQG	0,0039	0,0094	0,0056	0,0295	0,0090
MV	0,0025	0,0076	0,0051	0,0263	0,0117
MVR	0,0010	0,0062	0,0052	0,0240	0,0125

Tabela 68. Médias e desvios padrões da medida de dispersão (D)

	ANOVA	PP	VAR-MQG	MV	MVR
$\overline{\overline{D}}$	9,4486	5,4830	6,2054	5,8613	5,5331
eq(D)	5,5169	3,7883	3,9816	3,9965	3,6941

Tabela 69. Médias dos erros quadráticos das estimativas dos contrastes calculados analiticamente

	$\overline{eq(\hat{c}_1)}$	$\overline{eq(\hat{c}_2)}$	$\overline{eq(\hat{c}_3)}$	$\overline{eq(\hat{c}_4)}$	$\overline{eq(\hat{c}_5)}$
ANOVA	1,3289	1,3289	1,0850	1,5345	1,5345
PP	0,7142	0,7139	0,5828	0,8243	0,8244
VAR-MQG	1,0627	1,0620	0,8682	1,2272	1,2294
MV	0,9390	0,9383	0,7672	1,0841	1,0867
MVR	1,0110	1,0103	0,8262	1,1673	1,1700

Tabela 70. Erros quadráticos das estimativas dos contrastes calculados por simulação de Monte Carlo

	/2 >	(0)	/A')	(0)	/2.\
	$eq(\hat{c}_1)$	$eq(\hat{c}_2)$	$eq(c_3)$	$eq(c_4)$	$eq(\hat{c}_5)$
ANOVA	1,3220	1,3060	1,0942	1,5586	1,5391
PP	1,0050	1,0056	0,8342	1,1809	1,1713
VAR-MQG	1,0677	1,0662	0,8872	1,2613	1,2453
MV	1,0420	1,0412	0,8619	1,2239	1,2046
MVR	1,0134	1,0122	0,8364	1,1892	1,1696

Tabela 71. Intervalos de cobertura de cada contraste e do vetor dos contrastes

	$ICB(c_1)$	$ICB(c_2)$	$ICB(c_3)$	$ICB(c_4)$	$ICB(c_5)$	ICB(c)
ANOVA	0,9544	0,9586	0,9514	0,9512	0,9520	0,8046
PP	0,8508	0,8436	0,8398	0,8382	0,8402	0,5050
VAR-MQG	0,9374	0,9414	0,9312	0,9324	0,9358	0,7564
MV	0,9150	0,9120	0,9050	0,9058	0,9114	0,6734
MVR	0,9400	0,9392	0,9352	0,9346	0,9398	0,7612

Os resultados desta simulação indicaram que os modelos espaciais forneceram estimativas mais precisas aos contrastes entre os tratamentos do que a
análise clássica, confirmando a importância da detecção da estrutura de correlação
e o uso de suas informações na análise de dados conforme havia sido observado
no Experimento I. Entretanto, no Experimento I o modelo geoestatístico junto ao
método de estimação de máxima verossimilhança foi o mais eficiente, enquanto que
na Simulação III, o modelo que apresentou melhor desempenho foi o Papadakis.

Em suma, todos os modelos espaciais propostos na metodologia foram satisfatoriamente ajustados, utilizando o pacote geoestatístico geoR (Ribeiro Jr. & Diggle, 2004), para os modelos tratados segundo a abordagem frequentista e o programa estatístico WinBUGS 1.4 (Spiegelhalter et al., 2002), para o modelo Markoviano. A partir dos modelos espaciais ajustados foram obtidas estimativas mais precisas dos contrastes entre os tratamentos do que as resultantes pela análise clássica ANOVA, indicando que a dependência espacial entre as parcelas quando detectada não deve ser ignorada. Com o desenvolvimento computacional a aplicabilidade dos modelos que incorporam a dependência espacial estão cada vez mais acessíveis. Por exemplo, foi observado neste estudo, que o modelo proposto por Papadakis oferece maior precisão às estimativas dos contrastes entre os tratamentos do que a análise clássica ANOVA (i.i.d.), e computacionalmente é implementado através de uma análise de covariância disponível em programas estatísticos.

É importante o esclarecimento de que o modelo Markoviano, segundo a análise Bayesiana, não foi o foco principal desse estudo, por isso pode ser imprudente compará-lo aos demais modelos com total convicção.

Entre os modelos espaciais frequentistas, de um modo geral, o modelo geoestatístico utilizando os métodos de estimação de parâmetros baseados na verossimilhança foram os mais eficientes.

5 CONCLUSÕES

Com base nos resultados apresentados foram obtidas as conclusões que se seguem:

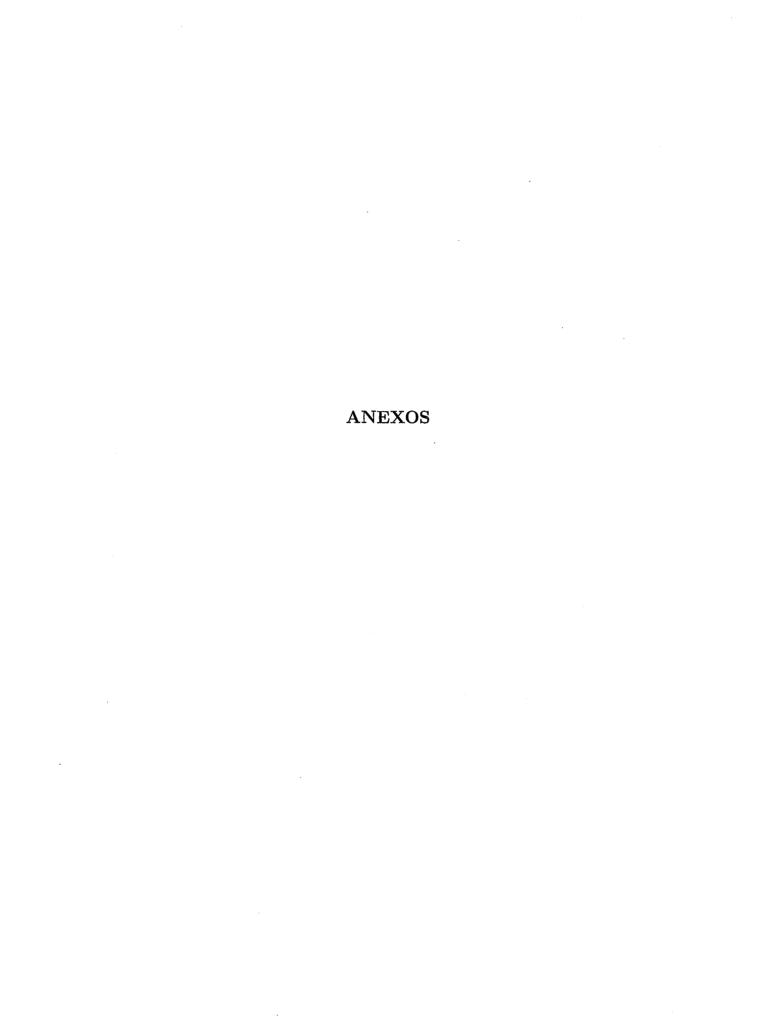
Os modelos espaciais forneceram estimativas mais precisas aos contrastes entre os tratamentos do que a análise clássica ANOVA (i.i.d.), mostrando a importância de se levar em consideração as informações da estrutura de correlação entre as parcelas quando essa foi detectada.

Os modelos espaciais apresentaram resultados muito próximos. O modelo geoestatístico junto aos métodos de estimação de parâmetros de máxima verossimilhança e máxima verossimilhança restrita, foram na maioria das vezes os mais eficientes entre os modelos e métodos de estimação de parâmetros espacias, tratados segundo a análise frequentista.

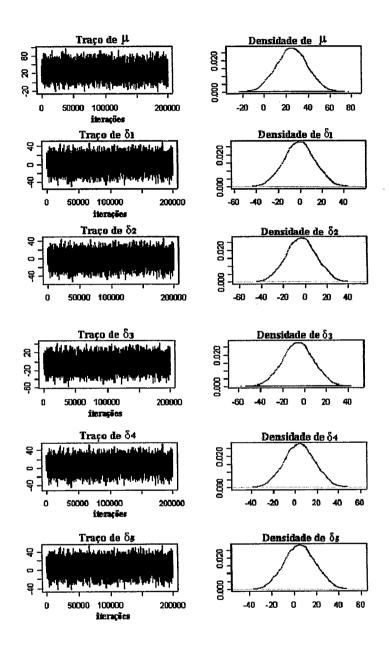
O modelo de Papadakis não apresentou o melhor desempenho entre os modelos espaciais, porém, foi mais preciso em relação à análise clássica, e trata-se de uma análise de covariância disponível em programas computacionais.

O modelo Markoviano apresentou-se eventualmente mais eficiente do que o modelo geoestatístico, porém, não foi suficientemente explorado.

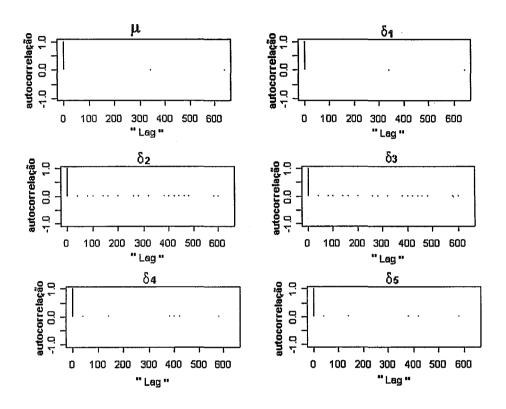
A eficiência dos modelos espaciais aumenta conforme maior o tamanho da amostra e mais alto o nível da dependência espacial entre as parcelas.



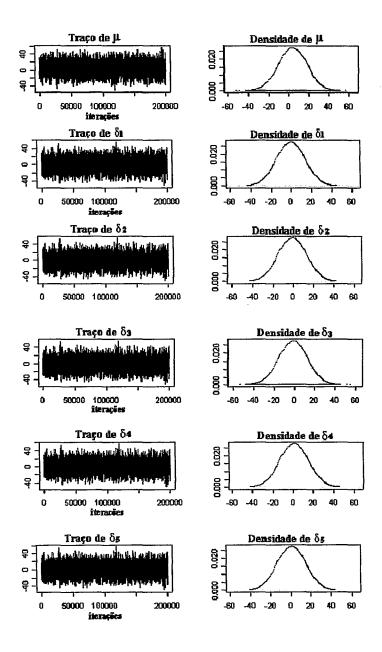
Anexo A - Traços e densidades da distribuição marginal a posteriori da média μ e dos efeitos dos tratamentos $\delta=(\delta_1,...,\delta_5)$



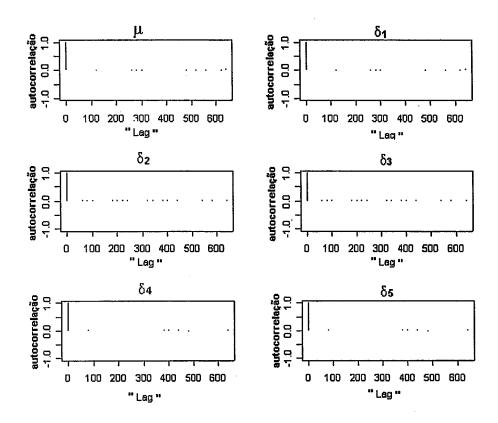
Anexo B - Autocorrelação da amostra gerada para os parâmetros $\mu \ {\rm e} \ \delta = (\delta_1,...,\delta_5) \ {\rm do} \ {\rm modelo} \ {\rm Markoviano}$



Anexo C - Traços e densidades da distribuição marginal a posteriori da média μ e dos efeitos dos tratamentos $\delta=(\delta_1,...,\delta_5)$



Anexo D - Autocorrelação da amostra gerada para os parâmetros $\mu \in \delta = (\delta_1,...,\delta_5) \text{ do modelo Markoviano}$



REFERÊNCIAS BIBLIOGRÁFICAS

- ASSUNÇÃO, R.M. Etatística espacial com aplicações em epidemiologia, economia, sociologia. In: ESCOLA DE MODELOS DE REGRESSÃO, 7., São Carlos, 2001. Minicurso. São Carlos: UFSCar, Departamento de estatística, 2001. 131p.
- ATKINSON, A.C. The use of residuals as a concomitant variable. **Biometrika**, v.56, p.33-41, 1969.
- BAIRD, D.; MEAD, R.; The empirical efficiency and validity of two neighbour models. Biometrics, v.47, p. 1473-1487, 1991.
- BARBIN, D. Planejamento e Análise Estatística de Experimentos Agronômicos. Arapongas: Midas, 2003. 194p.
- BARTLETT, M.S. The approximate recovery of information from replicated field experiments with large blocks. **Journal of Agricultural Science**, v.28, p. 418-427, 1938.
- BARTLETT, M.S. Nearest neighbour models in the analysis of field experiments with large blocks. Journal of the Royal Statistical Society, v.40, 147-174, 1978.
- BESAG, J. Spatial interaction and the statistical analysis of lattice (with discussion).

 Journal of the Royal Statistical Society Series B, v.36, p. 192-236, 1974.
- BESAG, J.; HIGDON, D. Bayesian analysis of agricultural field experiments.

 Journal of the Royal Statistical Society Series B, v.36, p. 691-746, 1999.

- BESAG, J.; KEMPTON, R.; Statistical analysis of field experiments using neighbouring plots. Biometrics, v.42, p. 231-251, 1986.
- BESAG, J.; KOOPERBERG, C. On conditional and intrinsic autoregressions. Biometrika, v.82, p. 733-746, 1995.
- BESAG, J.; YORK, J.C.; MOLLIÉ, A. Bayesian image restoration with two applications in spatial statistics (with discussion). Annals of the Institute of Statistical Mathematics, v.43, p. 1-59, 1991.
- BRAGA, L.P.V. Geoestatística e aplicações. In: SIMPÓSIO BRASILEIRO DE PRO-BABILIDADE E ESTATÍSTICA, 9., São Paulo, 1990. 36p.
- BROWNIE, C.; BOWMAN, D.T.; BURTON, J. W. Estimating spatial variation in analysis of data from yield trials: a comparison of methods. **Agronomy Journal**, v.85, n.6, p. 1244-1253, 1993.
- CARGNELUTTI FILHO, A.; STORCK, L.; LÚCIO, A.D. Ajustes de quadrado médio do erro em ensaios de competição de cultivares de milho pelo método de Papadakis. Pesquisa Agropecuária Brasileira, v.38, p.467-473, 2003.
- CLAYTON, D.; KALDOR, J.; Empirical bayes estimates age-standarlized relative risks for use in disease maping. **Biometrics**, v.43, p. 671-681, 1987.
- CLIF, A.D.; ORD, J.K. Spatial process: models & applications. London: Pion Press, 1981. 266p.
- CONLON, E. M.; WALLER, L. A. Flexible neighborhood structures in hierarchical models for disease mapping: research report 018. Minneapolis: University of Minnessota, School of Public Health, 1998. 25p.
- COSTA, J. R. Análise espacial e de vizinhança no melhoramento genético de plantas. Lavras, 2003. 82p. Tese (Doutorado)-Universidade Federal de Lavras.

- CRESSIE, N. Fitting variogram models by weighted least squares. **Journal of the**International Association of Mathematical Geology, v.17, p.563-586, 1985.
- CRESSIE, N. Statistics for spatial data. New York: John Wiley, 1993. 900p.
- CRESSIE, N.; CHAN, N.H. Spatial modeling of regional variables. **Journal of the**American Statistical Association, v.84, p. 393-401, 1989.
- CRESSIE, N.; LAHIRI, S.N. Asymptotics for REML estimation os spatial covariance parameters. Journal of Statistical Planning and Inference, v.50, p.327-341. 1996.
- CULLIS, B.R.; GLEESON, A.C. Spatial analysis of field experiments: an extension to two dimensions. **Biometrics**, v. 47, p. 1449-1460, 1991.
- DIGGLE, P.J.; RIBEIRO Jr., P.J. Model based geostatistics. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14. Caxambú, 2000. Minicurso. São Paulo: Associação Brasileira de Estatística, 2000. 102p.
- EISEBERG, B.E.; GAUCH, H.G.; ZOBEL, R.W.; KILIAN, W. Spatial of field experiments: fertilizer experiments with wheat (*Titium aetivum*) and tea (*Cammelia sinensis*). In KANG, M.S.; GAUCH, H.G (Ed.): Genotype by environment interation. Boca Raton: CRC Press, 1996. cap 14, p.373-404.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. I.E.E. transactions of Parttern Analysis and Machine Intelligence, v.6,p. 721-741, 1984.
- GILKS, W.R.; RICHARDSON, S.; SPIEGELHALTER, D.J. Markov Chain Monte Carlo in Pratice. London: Chapman and Hall, 1997. 486p.
- GRONDONA, M.O.; CRESSIE, N.; Using spatial considerations in the analysis of experiments. Technometrics, v.33, p. 381-392, 1991.

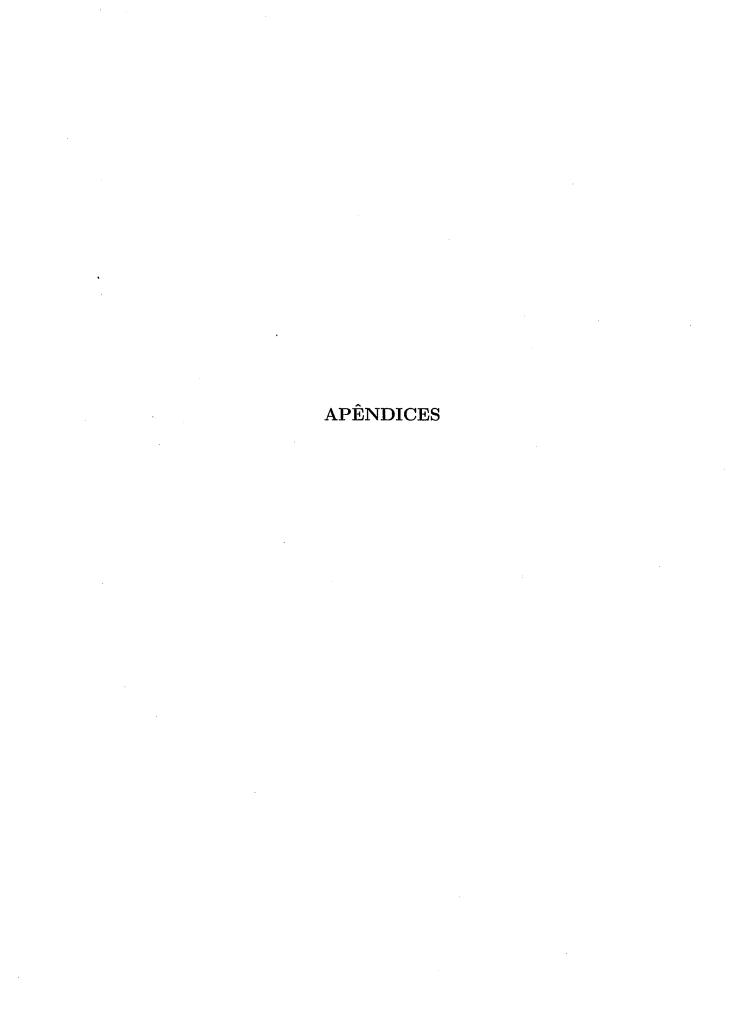
- GUERRA, P.A.G. Geoestatística operacional. Brasília: Ministério das Minas de Energia, 1988. 145p.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. Biometrika, v.57, p. 97-109, 1970.
- KELSALL, J.E.; WAKEFIELD, J.C. Discussion of "Bayesian models for spatially correlated disease and exposure data", by Best et al. In BERNARDO, J.M.; BERGER, J.O.; DAVID, A.P.; SMITH, A.F.M. (Ed.) Bayesian statistics 6. London: Oxford University Press, 1999. 151p.
- KITANIDIS, P.K.; SHEN, K.F. Geostatistical interpolation of chemical concentration. Advances in Water Resources, v.19, n.6, p.369-378, 1996.
- MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-639, 1986.
- MEAD, R.; CURNOW, R. N. Statistical methods in agriculture and experimental biology. New York: Chapman and Hall, 1983. 335 p.
- MERCER, W. B.; HALL, A. D. The experimental error of field trials. **Journal of Agricultural Science**, v.4, p.107-132, 1911.
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equations os state calculations by fast computing machines.

 Journal of Chemical Physics, v.21, p.1087-1091, 1953.
- OLIVEIRA, M.C.N. Métodos de estimação de parâmetros em modelos geoestatísticos com diferentes estruturas de covariâncias: uma aplicação ao teor de cálcio no solo. Piracicaba, 2003. 140p. Tese (Doutorado)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

- PAPADAKIS, J.S. Méthode statistique pour des expériences sur champ. Thessalonike: Institut d'Amélioration des Plantes à Salonique, 1937. 30p. (Bulletin, 23).
- PATHERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v.58, 545-554, 1971.
- PLUMMER M.; BEST, N.; COWLES K.; VINES, K. CODA: Output analysis and diagnostics for MCMC. version 0.5. Cambridge: Cambridge University, 2003. 34p.
- RAMALHO, M. A.; FERREIRA, D. F.; OLIVEIRA, A. C. de. A experimentação em genética e melhoramento de plantas. Lavras: UFLA, 2000. 326p.
- RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. **geoR**: A package for geostatistical data analysis. Version 1.4-6 http://www.est.ufpr.br/ paulojus/geoR/geoRdoc/geoR (05 Abr. 2004).
- RIBEIRO JUNIOR, P.J. Métodos estatísticos no estudo da variabilidade espacial de parâmetros do solo. Piracicaba, 1995. 99p. Dissertação(Mestrado)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- RUIZ CÁRDENAS, R. Modelagem da distribuição espaço-temporal da broca do café (*Hypothenemus hampei Ferrari*)em uma cultura da região central colombiana. Piracicaba, 2002. 120p. Dissertação(Mestrado)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SAMPER, F.J.S.; RAMIREZ, J.C. Geoestadística: Aplicaciones a la hidrogeología subterránea. Barcelona: Gráficas Torres, 1990. 484p.
- SOUZA, E. A. Aternativas experimentais na avaliação de progênies em programas de melhoramento genético vegetal. Piracicaba, 1997. 122p. Tese (Doutorado)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SPIEGELHALTER, D.; THOMAZ A.; BEST, N.; LUNN D. WinBUGS:user manual. Version 1.4. Cambridge: Cambridge University, 2002. 60p.

ş.

- STORCK, L.; GARCIA D. C.; LOPES S.J.; ESTEFANEL, V. Experimentação Vegetal. Santa Maria: Universidade Federal de Santa Maria, 2000. 128p.
- VENABLES, W.N.; SMITH, D.M. An introduction to R. Notes on R: a programing environment for data analysis and graphics. Version 1.7.0 http://cran.r-project.org/manuals.html (16 Abr. 2003).
- VER HOEF, J.M.; CRESSIE, N. Spatial statistics: Analysis of field experiments. In SHEINER, S.M.; GUREVITCH, J.(Ed.). **Design and analysis of ecological** experiments. London: Chapman and Hall, 1993. p.319-341.
- VIEIRA, S.R. Geoestatística. Curso de extensão universitária. Botucatu: UNESP, 1996. 71p.
- VIVALDI, L.J. Comparação entre métodos de análise espacial de experimentos de campo. Pesquisa Agropecuária Brasileira, v.25, p. 77-84, 1990.
- WALLER, L.A.; CARLIN, B.P.; XIA, H.; GELFAND, A.E. Hierarchical spational mapping of disease rates. Journal of the American Statistical Association, v.92, p. 607-617, 1997.
- WEBSTER, R. Quantitative spatial analysis of soil in the field. Advances in Soil Science, v.3, p. 1-70, 1985.
- WILKINSON, G. N.; ECKERT, S.R.; HANCOCK, T.W.; MAYO, O.; Nearest neighbour (NN) analysis of field experiments (with discussion). **Journal of the Royal** Statistical Society Series B, v.45, p.151-211, 1983.
- ZIMMERMAN, D.L.; HARVILLE, D.A.; A random field approach to the analysis of field-plot experiments and other spatial experiments. Biometrics, v.47, p.223-239, 1991.



APÊNDICE 1 - Subrotina geoR para a obtenção das análises: clássica, e geoestatística dos Experimentos I e II.

```
require(geoR)
data(hoef)
hoef.geo ← as:geodata(hoef, covar.col=4)
dat \leftarrow hoef\$dat
trat ← factor(hoef$trat)
# Matriz dos coeficiente dos contrastes
D \leftarrow matrix(c(...),5,5)
#ANOVA
reg \leftarrow lm(dat \sim (trat-1))
res \leftarrow residuals(reg)
X \leftarrow model.matrix(reg)
# Estimativa para os contrastes - análise clássica
estimativa \leftarrow function(X,dat,D){
est \leftarrow D%*%solve(crossprod(X))%*%t(X)%*%dat
return(est)
}
coan \leftarrow estimativa(X, dat, D)
# VARIOGRAMA SEGUIDO DE MÍNIMOS QUADRADOS
GENERALIZADOS (VAR-MQG)
# Construindo variograma empírico.
m \leftarrow cbind(hoef.geo\$coords,res)
resigeo \leftarrow as.geodata(m,coords.col=1:2,dat.col=3)
varemp \leftarrow variog(resigeo,coords=resigeo$coords,resigeo$dat, max.dist=5)
# Ajustando os variogramas teóricos com diferentes funções de correlação.
varteo \leftarrow variofit(varemp,ini.cov.pars=matrix(c(...),dim),
nug=1,cov.model="exp")
varteol \leftarrow variofit(varemp,ini.cov.pars=matrix(c(...),dim),
nug=1,cov.model="mat",kappa=1)
```

```
varteo2 \leftarrow variofit(varemp,ini.cov.pars=matrix(c(...),dim),
nug=1,cov.model="mat",kappa=2)
#Gráficos com os semivariogramas ajustados
plot(varemp)
lines(varteo,col=2)
lines(varteo1,col=3)
lines(varteo2,col=4)
legend(15,0.10, c("Exponencial", "Matérn(1)", "Matérn(2)"), fill=2:4)
# Construindo a inversa da matriz de covariância
inv \leftarrow varcov.spatial(m[,1:2],cov.pars=varteo$cov.pars,inv=TRUE,
nug=varteo$nugget)
inv1 \leftarrow varcov.spatial(m[,1:2],cov.pars=varteo1$cov.pars,inv=TRUE,
nug=varteo$nugget)
inv2 \(\lefta\) varcov.spatial(m[,1:2],cov.pars=varteo2\(\sigma\)cov.pars,inv=TRUE,
nug=varteo$nugget)
# Estimativas de mínimos quadrados generalizados para os contrastes.
estimativa \leftarrow function(X,inv,dat,D){
est \leftarrow D%*%solve(t(X)%*%inv%*%X)%*%t(X)%*%inv%*%dat
return(est)
}
covar \leftarrow estimativa (X,inv$inverse,dat,D)
covar1 \leftarrow estimativa (X,inv1\$inverse,dat,D)
covar2← estimativa (X,inv2$inverse,dat,D)
# Erros quadráticos das estimativas dos contrastes.
variancia \leftarrow function (X,inv){
var \longleftarrow D \%*\%solve(t(X)\%*\%inv\%*\%X)\%*\%t(D)
erro.q \leftarrow sqrt (diag(var))
return(erro.q)}
erroq ← variancia (X,inv$inverse)
errog1 \leftarrow variancia (X,inv1$inverse)
erroq2 ← variancia (X,inv2$inverse)
```

```
# MÁXIMA VEROSSIMILHANÇA (MV)
mv \leftarrow likfit(hoef.geo, ini=c(7,0.8), nug = 1, trend=\sim trat-1,
cov.model="mat",kappa=0.5)
mv1 \leftarrow likfit(hoef.geo, ini=c(7,0.8), nug = 1,trend=\sim trat-1,
cov.model="mat",kappa=1)
mv2 \leftarrow likfit(hoef.geo, ini=c(7,0.8), nug = 1, trend=\sim trat-1,
cov.model="mat",kappa=2)
comv ← D %*%mv$beta
comv1 \leftarrow D \%*\%mv1\$beta
comv2← D %*%mv2$beta
# Erros quadráticos das estimativas dos contrastes.
eq.mv \leftarrow sqrt (diag(D%*%mv2$beta.var%*%t(D)))
# MÁXIMA VEROSSIMILHANÇA RESTRITA (MVR).
mvr \leftarrow likfit(hoef.geo, ini=c(6,0.5), nug = 1,trend=\sim trat-1,
method.lik = "REML",cov.model="mat",kappa=0.5)
mvr1 \leftarrow likfit(hoef.geo, ini=c(6,0.5), nug = 1,trend=\sim trat-1,
method.lik = "REML",cov.model="mat",kappa=1)
mvr2 \leftarrow likfit(hoef.geo, ini=c(6,0.5), nug = 1,trend=\sim trat-1,
method.lik = "REML",cov.model="mat",kappa=2)
# Erros quadráticos das estimativas dos contrastes.
eq.mvr \leftarrow sqrt (diag(D%*%mvr1$beta.var%*%t(D)))
APÊNDICE 2 - Subrotina geoR para a obtenção da análise de
          Papadakis Experimentos I e II.
require(geoR)
```

data(hoef)

 $hoef.geo \leftarrow as.geodata(hoef, covar.col=4)$

```
trat ← factor(hoef$trat)
y \leftarrow - (hoef\$dat)
reg \leftarrow lm(y \sim trat-1)
res \leftarrow residuals(reg)
X \leftarrow \text{model.matrix(reg)}
n \leftarrow nx*ny
nx \leftarrow 5 \# número das abscissas.
ny <br/> \longleftarrow 5 # número das ordenadas.
#Matriz dos resíduos
matrizres \leftarrow function(res, nl, nc)
mres \leftarrow matrix(res,c(nl,nc))
l1 \leftarrow matrix(c(rep(0,nc+2)),c(1,nc+2))
c1 \leftarrow matrix(c(rep(0,nl)),c(nl,1))
meio1 \leftarrow cbind(c1, mres, c1)
matres \leftarrow rbind(l1,meio1,l1)
return(matres)
}
# Matriz das somas dos resíduos das parcelas vizinhas.
soma \leftarrow function(matres, nl, nc)
draw1 \leftarrow NULL
for(j in 2:(nc+1)){
for(i in 2:(nl+1)){
resi \longleftarrow (matres[i+1,j] + matres[i-1,j] + matres[i,j+1] + matres[i,j+1] + matres[i,j+1])
somares \leftarrow matrix(draw1,c(nl,nc))
return(somares)
}
# Matriz das médias dos resíduos das parcelas vizinhas
mediares \leftarrow function(somares,nl,nc){}
A \leftarrow somares[1,1]/3
```

```
B \leftarrow somares[1,2:(nc-1)]/4
C \leftarrow somares[1,nc]/3
D \leftarrow somares[2:(nl-1),1]/4
E \leftarrow somares[2:(nl-1),2:(nc-1)]/5
F \leftarrow somares[2:(nl-1),nc]/4
G \leftarrow somares[nl,1]/3
H \leftarrow somares[nl,2:(nc-1)]/4
I \leftarrow somares[nl,nc]/3
meio \leftarrow matrix(c(D,E,F),c((nl-2),nc))
media \leftarrow rbind(c(A,B,C),meio,c(G,H,I))
return(media)
}
erro.PP \leftarrow function(X,A,Z,R,var){
var.a \leftarrow A\%*\%((solve(crossprod(X))+solve(crossprod(X)))\%*\%
crossprod(X,Z)\%*\%solve(crossprod(R))\%*\%crossprod(Z,X)
%*\%solve(crossprod(X)))* var)%*%t(A)
e.qua \leftarrow sqrt(diag(var.a))
return(e.qua)
}
#Estimativas dos contrastes.
matres \leftarrow matrizres(res, nx, ny)
somares \leftarrow soma(matres, nx, ny)
medres \leftarrow mediares(somares,nx,ny)
Z \leftarrow matrix(medres,c(n,1)) \# matriz cm o calculo da covariavel
P \leftarrow (diag(n))-(X\%*\%solve(crossprod(X))\%*\%t(X))
b \leftarrow solve(crossprod(Z,P)\%*\%Z)\%*\%crossprod(Z,P)\%*\%y
coop \leftarrow A\%*\%(solve(crossprod(X))\%*\%crossprod(X,y))
-(solve(crossprod(X))\%*\%crossprod(X,Z)\%*\%b)
# Erros quadráticos das estimativas dos contrastes
R \leftarrow P\%*\%Z
Ra \leftarrow crossprod(y,X)\%*\%solve(crossprod(X))\%*\%crossprod(X,y)
```

```
SSRB \leftarrow crossprod(y,R)\%*\%solve(crossprod(R))\%*\%crossprod(R,y)
vPP \leftarrow (crossprod(y)-(Ra+SSRB))/(n-reg\$rank-qr(Z)\$rank)
var \leftarrow vPP[1]
eq.pp \leftarrow erro.PP(X,A,Z,R,var)
APÊNDICE 3 - Subrotina geoR para as Simulações I e II.
require(geoR)
# Generalização das Simulações.
#Estipular valores para: nx; ny; phi e ite.
mu \leftarrow 25
nx \leftarrow nx
ny ← ny
n \leftarrow nx * ny
nrep \leftarrow (n/5)
sigma \longleftarrow 5
phi ← phi
nugg \longleftarrow 1
ite \leftarrow ite
D \longleftarrow matrix(c(.....),dim)
D \leftarrow t(D)
verd \leftarrow D %*%rbind(0,-3,-5,6,6)
par.ori \leftarrow par(no.readonly = TRUE)
set.seed(123)
# FUNÇÕES GERAIS
# Gerando os dados geoestatísticos
tratamentos \leftarrow rep(1:5,each=nrep)
efeitos ← function(tratamentos) {
trat ← sample(tratamentos,n, replace = FALSE)
vetor ← numeric(length(tratamentos))
vetor[trat == 1] \leftarrow 0
```

```
vetor[trat == 2] \leftarrow -3
vetor[trat == 3] \leftarrow -5
vetor[trat == 4] \leftarrow 6
vetor[trat == 5] \leftarrow 6
S \leftarrow grf(n, grid="reg", nx=nx, ny=ny, xlims = c(1, nx),
ylims = c(1, ny),cov.pars=c(sigma, phi),messages=FALSE)
Z \leftarrow rnorm(n,0,nugg)
data \leftarrow mu + vetor + S\$data + Z
res ← as.data.frame(cbind(S$coords,data,trat))
res$trat ← factor(res$trat)
return(res)
}
# Contrastes ANOVA
contrastes \leftarrow function(F,y,D) {
cont \leftarrow D \%*\%(solve(crossprod(F)))\%*\%crossprod(F,y)
return(cont)
}
# Erros quadráticos - ANOVA
eq.an \leftarrow function(F,D) {
v \leftarrow ((sum((reg\$residuals)^2))/(n-5))
var \leftarrow v*(D \%*\% (solve(crossprod(F))) \%*\% t(D))
e.qua \leftarrow sqrt(diag(var))
return(e.qua)
}
# Contrastes VAR-MQG
contrastesv \leftarrow function(F,inv,y,D)  {
covar \leftarrow D \%*\% (solve(crossprod(F,inv))
%*%F)) %*% crossprod(F,inv) %*%y
return(covar)
 }
 # Erros quadráticos MQG
```

```
eq.var \leftarrow function(F,inv,D) {
var \leftarrow D \%*\%solve(crossprod(F,inv) \%*\%F) \%*\%t(D)
e.qual \leftarrow sqrt(diag(var))
return(e.qual)
# Covariáveis do modelo de PP
#Matriz dos resíduos
matrizres \leftarrow function(res, nl, nc)  {
mres \leftarrow matrix(res,c(nl,nc))
11 \leftarrow \text{matrix}(c(\text{rep}(0,\text{nc}+2)),c(1,\text{nc}+2))
c1 \leftarrow matrix(c(rep(0,nl)),c(nl,1))
meio1 \leftarrow cbind(c1, mres, c1)
matres \leftarrow rbind(l1,meio1,l1)
return(matres)
# Matriz das somas dos resíduos das parcelas vizinhas.
soma \leftarrow function(matres, nl, nc) 
draw1 \leftarrow NULL
for(j in 2:(nc+1)) {
for(i in 2:(nl+1)) {
resi \leftarrow
matres[i+1,j]+matres[i-1,j]+matres[i,j]+matres[i,j-1]+matres[i,j+1]
draw1 \leftarrow c(draw1,resi)
}}
somares \leftarrow matrix(draw1,c(nl,nc))
return(somares)
 # Matriz das médias dos resíduos das parcelas vizinhas
mediares \leftarrow function(somares,nl,nc) {
 A \leftarrow somares[1,1]/3
 B \leftarrow somares[1,2:(nc-1)]/4
```

```
C \leftarrow somares[1,nc]/3
D \leftarrow somares[2:(nl-1),1]/4
E \leftarrow somares[2:(nl-1),2:(nc-1)]/5
F \leftarrow somares[2:(nl-1),nc]/4
G \leftarrow somares[nl,1]/3
H \leftarrow somares[nl,2:(nc-1)]/4
I \leftarrow \text{somares[nl,nc]/3}
meio \leftarrow matrix(c(D,E,F),c((nl-2),nc))
media \leftarrow rbind(c(A,B,C),meio,c(G,H,I))
return(media)
}
erro.PP \leftarrow function(X,A,Z,R,vPP) {
var.a \leftarrow A \%*\%((solve(crossprod(X))+solve(crossprod(X)))\%*\%
crossprod(X,Z)\%*\%solve(crossprod(R))\%*\%crossprod(Z,X)
%*\% solve(crossprod(X)))*vPP) %*\%t(A)
e.qua \leftarrow sqrt(diag(var.a))
return(e.qua)
}
# Objetos que guardam os resultados gerados em cada iteração.
ANOVA
co \leftarrow matrix(0, 5, ite)
dimnames(co) ← list(paste("c", 1:5, sep=""),NULL)
eq.anova \leftarrow matrix(0, 5, ite)
dimnames(eq.anova) \( \leftarrow \text{list(paste("c", 1:5, sep=""),NULL)} \)
ic.anova \leftarrow NULL
ic1.anova \leftarrow NULL
ic2.anova \leftarrow NULL
ic3.anova \leftarrow NULL
ic4.anova \longleftarrow NULL
ic5.anova \leftarrow NULL
# VAR-MQG
```

```
covar \leftarrow matrix(0, 5, ite)
dimnames(covar) ← list(paste("c", 1:5, sep=""), NULL)
eq.variog \leftarrow matrix(0, 5, ite)
dimnames(eq.variog) ← list(paste("c", 1:5, sep=""),NULL)
ic1.var \leftarrow - NULL
ic2.var \longleftarrow NULL
ic3.var \leftarrow NULL
ic4.var \leftarrow NULL
ic5.var \longleftarrow NULL
# MV
comv \leftarrow matrix(0, 5, ite)
dimnames(comv) ← list(paste("c", 1:5, sep=""),NULL)
eq.mv \leftarrow matrix(0, 5, ite)
dimnames(eq.mv) ← list(paste("c", 1:5, sep=""),NULL)
ic1.mv \leftarrow - NULL
ic2.mv \leftarrow NULL
ic3.mv \leftarrow NULL
ic4.mv \longleftarrow NULL
ic5.mv \leftarrow NULL
# MVR
comvre \leftarrow matrix(0, 5, ite)
dimnames(comvre) ← list(paste("c", 1:5, sep=""),NULL)
eq.mvr \leftarrow matrix(0, 5, ite)
dimnames(eq.mvr) ← list(paste("c", 1:5, sep=""),NULL)
ic1.mvr \leftarrow NULL
ic2.mvr \leftarrow NULL
ic3.mvr ← NULL
ic4.mvr \leftarrow NULL
ic5.mvr \leftarrow NULL
# PP
coop \leftarrow matrix(0, 5, ite)
```

```
dimnames(coop) ← list(paste("c", 1:5, sep=""),NULL)
eq.PP \leftarrow matrix(0, 5, ite)
dimnames(eq.PP) \(\leftarrow\) list(paste("c", 1:5, sep=""),NULL)
ic1.PP \longleftarrow NULL
ic2.PP \longleftarrow NULL
ic3.PP \longleftarrow NULL
ic5.PP \leftarrow NULL
# MEDIDAS DE DISPERSÃO
disco \leftarrow matrix(0, ite, 1)
dimnames(disco) ← list(paste("D", 1:ite, sep=""),NULL)
disvar \leftarrow matrix(0, ite, 1)
dimnames(disvar) ← list(paste("D", 1:ite, sep=""),NULL)
dismv \leftarrow matrix(0, ite, 1)
dimnames(dismv) \( \leftarrow \text{list(paste("D", 1:ite, sep=""),NULL)} \)
dismvre \leftarrow matrix(0, ite, 1)
dimnames(dismvre) ← list(paste("D", 1:ite, sep=""),NULL)
disPP \leftarrow matrix(0,ite,1)
dimnames(disPP) ← list(paste("D", 1:ite, sep=""),NULL)
# PARÂMETROS DE CORRELAÇÃO
parsvar \leftarrow matrix(0, 3, ite)
dimnames(parsvar) ← list(paste("p", 1:3, sep=""),NULL)
parsmv \leftarrow matrix(0, 3, ite)
dimnames(parsmv) \leftarrow list(paste("p", 1:3, sep=""),NULL)
parsmvr \leftarrow matrix(0, 3, ite)
 dimnames(parsmvr) \leftarrow list(paste("p", 1:3, sep=""), NULL)
 dadosy \leftarrow NULL
 cta \longleftarrow 0
 repeat {
 cta \longleftarrow cta+1
```

```
dados \leftarrow efeitos(tratamentos)
dados.geo \leftarrow as.geodata(dados, covar.col=4)
y \longleftarrow (dados\$data)
tratt \leftarrow cbind(tratt,dados\$trat)
trat ← factor(dados$trat)
dadosy \leftarrow cbind(dadosy,dados\$data)
reg \leftarrow lm(data \sim (trat-1), data = dados)
res \leftarrow residuals(reg)
m \leftarrow cbind(dados.geo\$coords,res)
resigeo \leftarrow as.geodata(m,coords.col=1:2,data.col=3)
F \leftarrow model.matrix(reg)
#ANOVA
#Estimativas dos contrastes
coanova \leftarrow contrastes(F,y,D)
co[,cta] \leftarrow coanova
# Erro quadrático das estimativas dos contrastes
erro.q1 \leftarrow eq.an(F,D)
eq.anova[,cta] \leftarrow erro.q1
# Medida de dispersão
disco1 \leftarrow sum((coanova-verd)^2)
disco[cta,] \leftarrow disco1
# Intervalos de confiança
ic1.ano \leftarrow coanova[1,] + qt(c(0.025, 0.975),(n-5)) * erro.q1[1]
ic1.anova ← rbind(ic1.anova,ic1.ano)
ic2.ano \leftarrow coanova[2,] + qt(c(0.025, 0.975),(n-5)) * erro.q1[2]
ic2.anova ← rbind(ic2.anova,ic2.ano)
ic3.ano \leftarrow coanova[3,] + qt(c(0.025, 0.975),(n-5)) * erro.q1[3]
ic3.anova ← rbind(ic3.anova,ic3.ano)
ic4.ano \leftarrow coanova[4,] + qt(c(0.025, 0.975),(n-5)) * erro.q1[4]
ic4.anova ← rbind(ic4.anova,ic4.ano)
ic5.ano \leftarrow coanova[5,] + qt(c(0.025, 0.975),(n-5)) * erro.q1[5]
```

```
ic5.anova ← rbind(ic5.anova,ic5.ano)
#MÉTODO DO VARIOGRAMA SEGUIDO DE MÍNIMOS
QUADRADOS GENERALIZADOS - MQG
# Construção dos variogramas
dist.eu \leftarrow max(dist(dados.geo\$coords))
varemp \leftarrow variog(resigeo, coords = resigeo coords, resigeo data,
max.dist=(dist.eu)*0.75,messages=FALSE)
varteo \leftarrow variofit(varemp,ini.cov.pars=c(6,0.5),nug=1,cov.model="exp",
limits=pars.limits(phi=c(0, dist.eu)),messages=FALSE)
# Construção da inversa da matriz de covariância entre as parcelas
inv \leftarrow varcov.spatial(m[,1:2],cov.pars=varteo$cov.pars,inv=TRUE,
nug=varteo$nugget)
# Estimativas dos contrastes
covariog \leftarrow contrastesv(F,inv$inverse,y,D)
covar[,cta] \leftarrow covariog
# Erros quadráticos das estimativas dos contrastes
erro.q2 \leftarrow eq.var(F,inv$inverse,D)
eq.variog[,cta] \leftarrow erro.q2
#Medida de dispersão
disvar1 \leftarrow sum((covariog-verd)^2)
disvar[cta,] \leftarrow disvar1
# Intervalos de confiança
ic1.variog \leftarrow covariog[1,] + qnorm(c(0.025, 0.975)) * erro.q2[1]
ic1.var ← rbind(ic1.var,ic1.variog)
ic2.variog \leftarrow covariog[2] + qnorm(c(0.025, 0.975)) * erro.q2[2]
ic2.var \leftarrow rbind(ic2.var,ic2.variog)
ic3.variog \leftarrow covariog[3,] + qnorm(c(0.025, 0.975)) * erro.q2[3]
ic3.var \leftarrow rbind(ic3.var,ic3.variog)
 ic4.variog \leftarrow covariog[4,] + qnorm(c(0.025, 0.975)) * erro.q2[4]
 ic4.var \leftarrow rbind(ic4.var,ic4.variog)
 ic5.variog \leftarrow covariog[5,] + qnorm(c(0.025, 0.975)) * erro.q2[5]
```

```
ic5.var \leftarrow rbind(ic5.var,ic5.variog)
#Parâmetros de covariância estimados pelo variograma
sig.phi.nug \leftarrow c(varteo\$cov.pars, varteo\$nugget)
parsvar[,cta] ← sig.phi.nug
# MÁXIMA VEROSSIMILHANÇA (MV)
# Estimativas dos contrastes.
mv \leftarrow likfit(dados.geo, ini=c(10,0.5), nug = 1, trend=\sim(trat-1),cov.model="exp",
limits=pars.limits(phi=c(0,dist.eu)),messages=FALSE)
comvero ← D %*%mv$beta
comv[,cta] \leftarrow comvero
# Erros quadráticos das estimativas dos contrastes
erro.q3 \leftarrow sqrt(diag(D \%*\%mv\$beta.var\%*\%t(D)))
eq.mv[,cta] \leftarrow erro.q3
# Medida de dispersão
dismv1 \leftarrow sum((comvero-verd)^2)
dismv[cta,] \leftarrow dismv1
# Intervalos de cobertura
ic1mv \leftarrow comvero[1,] + qnorm(c(0.025, 0.975)) * erro.q3[1]
ic1.mv \leftarrow rbind(ic1.mv,ic1mv)
ic2mv \leftarrow comvero[2,] + qnorm(c(0.025, 0.975)) * erro.q3[2]
ic2.mv \leftarrow rbind(ic2.mv,ic2mv)
ic3mv \leftarrow comvero[3,] + qnorm(c(0.025, 0.975)) * erro.q3[3]
ic3.mv \leftarrow rbind(ic3.mv,ic3mv)
ic4mv \leftarrow comvero[4,] + qnorm(c(0.025, 0.975)) * erro.q3[4]
ic4.mv \leftarrow rbind(ic4.mv,ic4mv)
ic5mv \leftarrow comvero[5,] + qnorm(c(0.025, 0.975)) * erro.q3[5]
ic5.mv \leftarrow rbind(ic5.mv,ic5mv)
#Parâmetros de correlação estimados pelo método da MV.
sig.phi.nug \leftarrow c(mv$cov.pars,mv$nugget)
parsmv[,cta] ← sig.phi.nug
# MÁXIMA VEROSSIMILHANÇA RESTRITA
```

```
# Estimativas dos contrastes
mvr \leftarrow likfit(dados.geo, ini=c(5,0.5), nug =1,trend=\sim(trat-1), cov.model=
"exp",method.lik="REML",limits=pars.limits(phi=c(0,dist.eu)),messages=FALSE)
comverore ← D %*%mvr$beta
comvre[,cta] \leftarrow comverore
# Erros quadráticos dos contrastes
erro.q4 \leftarrow sqrt(diag(D \%*\%(mvr\$beta.var) \%*\%t(D)))
eq.mvr[,cta] \leftarrow erro.q4
# Medida de dispersão
dismvre1 ← sum((comverore-verd) ^2
dismvre[cta,] \longleftarrow dismvre1
#Intervalo de cobertura
ic1mvr \leftarrow comverore[1,] + qnorm(c(0.025, 0.975)) * erro.q4[1]
ic1.mvr \leftarrow rbind(ic1.mvr,ic1mvr)
ic2mvr \leftarrow comverore[2,] + qnorm(c(0.025, 0.975)) * erro.q4[2]
ic2.mvr \leftarrow rbind(ic2.mvr,ic2mvr)
ic3mvr \leftarrow comverore[3,] + qnorm(c(0.025, 0.975)) * erro.q4[3]
ic3.mvr \leftarrow rbind(ic3.mvr,ic3mvr)
ic4mvr \leftarrow comverore[4,] + qnorm(c(0.025, 0.975)) * erro.q4[4]
ic4.mvr \leftarrow rbind(ic4.mvr,ic4mvr)
ic5mvr \leftarrow comverore[5,] + qnorm(c(0.025, 0.975)) * erro.q4[5]
ic5.mvr ← rbind(ic5.mvr,ic5mvr)
#Parâmetros de correlação estimados pelo método da MVR.
sig.phi.nug \leftarrow c(mvr$cov.pars,mvr$nugget)
parsmvr[,cta] \leftarrow sig.phi.nug
# Estimativas dos contrastes pela covariância proposta por Papadakis (PP).
matres \leftarrow matrizres(res, nx, ny)
somares \longleftarrow soma(matres,nx,ny)
medres \leftarrow mediares(somares,nx,ny)
Z \leftarrow \text{matrix}(\text{medres}, c(n, 1))
P \leftarrow (diag(n))-F \%*\%solve(crossprod(F)) \%*\%t(F)
```

```
Pinv \leftarrow solve(crossprod(Z,P) \%*\%Z)
b \leftarrow Pinv \%*\%crossprod(Z,P) \%*\%y
coPP \leftarrow D \%*\%(solve(crossprod(F))) \%*\%crossprod(F,y)
- (solve(crossprod(F))) %*%crossprod(F,Z) %*%b
coop[,cta] \leftarrow coPP
# Erros quadráticos das estimativas dos contrastes
R \leftarrow P \%*\%Z
Ra \leftarrow crossprod(y,F) \%*\%solve(crossprod(F)) \%*\%crossprod(F,y)
SSRB \leftarrow crossprod(y,R) \%* \%solve(crossprod(R)) \%*\%crossprod(R,y)
varP \leftarrow (crossprod(y)-(Ra+SSRB))/(n-reg\$rank-qr(Z)\$rank)
varP \leftarrow varP[1]
erro.q5 \leftarrow erro.PP(F,D,Z,R,varP)
eq.PP[,cta] \leftarrow erro.q5
# Intervalos de confiança
ic1.P \leftarrow coPP[1,] + qt(c(0.025, 0.975),(n-5)) * erro.q5[1]
ic1.PP \leftarrow rbind(ic1.PP,ic1.P)
ic2.P \leftarrow coPP[2,] + qt(c(0.025, 0.975),(n-5)) * erro.q5[2]
ic2.PP \leftarrow rbind(ic2.PP,ic2.P)
ic3.P \leftarrow coPP[3,] + qt(c(0.025, 0.975),(n-5)) * erro.q5[3]
ic3.PP \leftarrow rbind(ic3.PP,ic3.P)
ic4.P \leftarrow coPP[4,] + qt(c(0.025, 0.975),(n-5)) * erro.q5[4]
ic4.PP \leftarrow rbind(ic4.PP,ic4.P)
ic5.P \leftarrow coPP[5,] + qt(c(0.025, 0.975),(n-5)) * erro.q5[5]
ic5.PP \leftarrow rbind(ic5.PP,ic5.P)
disPP1 \leftarrow sum((coPP-verd)^2)
disPP[cta,] ← disPP1
if(cta==ite)break
# Medidas que resumem os resultados
medco \leftarrow apply(co,1,mean)
medcovar \leftarrow apply(covar,1,mean)
```

```
medcomv \leftarrow apply(comv,1,mean)
medcomvre \leftarrow apply(comvre,1,mean)
medcoop \leftarrow apply(coop,1,mean)
#Diferença entre contrastes verdadeiros e a média
dos contrastes estimados
df.anova \leftarrow abs(verd-medco)
df.var \leftarrow abs(verd-medcovar)
df.mv \leftarrow abs(verd-medcomv)
df.mvre ← abs(verd-medcomvre)
df.PP \leftarrow abs(verd-medcoop)
# Média dos erros quadráticos analíticos
med.eq.anova \leftarrow apply(eq.anova,1,mean)
med.eq.variog \leftarrow apply(eq.variog,1,mean)
med.eq.mv \leftarrow apply(eq.mv,1,mean)
med.eq.mvr ← apply(eq.mvr,1,mean)
med.eq.PP \leftarrow apply(eq.PP,1,mean)
# Desvio padrao MC
sdco \leftarrow apply(co,1,sd)
sdcovar \leftarrow apply(covar, 1, sd)
sdcomv \leftarrow apply(comv,1,sd)
sdcomvre \leftarrow apply(comvre,1,sd)
sdcoop \leftarrow apply(coop,1,sd)
# Medida de dispersão
med.disco \leftarrow mean(disco)
med.disvar \longleftarrow mean(disvar)
med.dismv \leftarrow mean(dismv)
med.dismvre \leftarrow mean(dismvre)
med.disPP \leftarrow mean(disPP)
 #desvio padrao medida de dispersão
sd.disco \leftarrow sd(disco)
 sd.disvar \leftarrow sd(disvar)
```

```
sd.dismv \leftarrow sd(dismv)
sd.dismvre \leftarrow sd(dismvre)
sd.disPP \leftarrow sd(disPP)
# INTERVALO DE COBERTURA PARA CADA CONTRASTE.
int.cob ← function(ic,cv) {
vetor \leftarrow numeric(ite)
vetor[ic[,1] < cv \& cv < ic[,2]] \leftarrow 1
prob ← sum(vetor)/ite
return(vetor,prob)
}
# Anova
prob.anovac1 \leftarrow int.cob(ic1.anova,-4)
prob.anovac2 \leftarrow int.cob(ic2.anova,6)
prob.anovac3 \leftarrow int.cob(ic3.anova,10)
prob.anovac4 \leftarrow int.cob(ic4.anova,2)
prob.anovac5 \leftarrow int.cob(ic5.anova,0)
# VAR-MQG
prob.variogc1 \leftarrow int.cob(ic1.var,-4)
prob.variogc2 \leftarrow int.cob(ic2.var,6)
prob.variogc3 \leftarrow int.cob(ic3.var,10)
prob.variogc4 \leftarrow int.cob(ic4.var,2)
prob.variogc5 \leftarrow int.cob(ic5.var,0)
# MV
prob.mvc1 \leftarrow int.cob(ic1.mv,-4)
prob.mvc2 \leftarrow int.cob(ic2.mv,6)
prob.mvc3 \leftarrow int.cob(ic3.mv,10)
prob.mvc4 \leftarrow int.cob(ic4.mv,2)
prob.mvc5 \leftarrow int.cob(ic5.mv,0)
#MVR
prob.mvrc1 \leftarrow int.cob(ic1.mvr,-4)
prob.mvrc2 \leftarrow int.cob(ic2.mvr,6)
```

```
prob.mvrc3 ← int.cob(ic3.mvr,10)
prob.mvrc4 \leftarrow int.cob(ic4.mvr,2)
prob.mvrc5 \leftarrow int.cob(ic5.mvr,0)
#PP
prob.PPc1 \leftarrow int.cob(ic1.PP,-4)
prob.PPc2 \leftarrow int.cob(ic2.PP,6)
prob.PPc3 \leftarrow int.cob(ic3.PP,10)
prob.PPc4 \leftarrow int.cob(ic4.PP,2)
prob.PPc5 \leftarrow int.cob(ic5.PP,0)
# INTERVALO DE COBERTURA PARA O VETOR
DE CONTRASTES.
int.cob.vetor ← function(vetorc1, vetorc2, vetorc3, vetorc4, vetorc5) {
vet \leftarrow cbind(vetorc1, vetorc2, vetorc3, vetorc4, vetorc5)
todos \leftarrow numeric(ite)
todos[vet[,1]==1 \& vet[,2]==1 \& vet[,3]==1
& \text{vet}[,4] = =1 & \text{vet}[,5] = =1] \leftarrow 1
prob.v \leftarrow (sum(todos))/ite
return(prob.v)
}
# ANOVA
prob.v.anova — int.cob.vetor(prob.anovac1$vetor,prob.anovac2$vetor,
prob.anovac3$vetor,prob.anovac4$vetor,prob.anovac5$vetor)
# VAR-MQG
prob.v.variog \(\lefta\) int.cob.vetor(prob.variogc1\(\frac{1}{2}\)vetor,prob.variogc2\(\frac{1}{2}\)vetor,
prob.variogc3$vetor,prob.variogc4$vetor,prob.variogc5$vetor)
# MV
prob.v.mv ← int.cob.vetor(prob.mvc1$vetor,prob.mvc2$vetor,
prob.mvc3$vetor, prob.mvc4$vetor,prob.mvc5$vetor)
# MVR
prob.v.mvr ← int.cob.vetor(prob.mvrc1$vetor,prob.mvrc2$vetor,
prob.mvrc3$vetor,prob.mvrc4$vetor,prob.mvrc5$vetor)
```

```
# PP
prob.v.PP ← int.cob.vetor(prob.PPc1$vetor,prob.PPc2$vetor,
prob. PPc3\$ vetor, prob. PPc4\$ vetor, prob. PPc5\$ vetor)
APÊNDICE 4 - Simulação III: Diferentes aleatorizações.
require(geoR)
set.seed(123)
#matriz dos coeficientes dos contrastes
D \longleftarrow matrix(c(...),dim)
D \leftarrow t(D)
nx \leftarrow -5
ny---5
n←-nx*ny
ite \leftarrow 5000
verd \longleftarrow\! D
#Simulando diferente aleatorização
tratamentos \leftarrow rep(1:5, each=5)
efeitos←—function(tratamentos){
trat←sample(tratamentos,25, replace = FALSE)
vetor \leftarrow numeric(length(tratamentos))
vetor[trat == 1] \leftarrow 0
vetor[trat == 2] \leftarrow -3
vetor[trat == 3] \leftarrow -5
vetor[trat == 4] \leftarrow 6
vetor[trat == 5] \leftarrow 6
x1 \leftarrow rep(1:5,5)
x2 \leftarrow rep(1:5, each=5)
 dadosreais\leftarrowc(25,..., 24)
 data \leftarrow -dadosreais + vetor
 res \leftarrow as.data.frame(cbind(x1,x2,data,trat))
```

```
res$trat ← factor(res$trat)
return(res)
}
# Contrastes ANOVA
#Daqui em diante segue-se como no APÊNDICE 3.
APÊNDICE 5 - Subrotina WinBugs para o ajuste do modelo Markoviano
model
s[1:N] \sim car.normal(adj[], weights[], num[], taus)
for (i in 1:N) {
O[i]~dnorm(pi[i], tau)
pi[i] \longleftarrow alpha0 + alpha1 * x1[i] + alpha2*x2[i] + alpha3 * x3[i] +
alpha4 * x4[i]+ alpha5 * x5[i] + s[i] }
for(k in 1:sumNumNeigh) {
weights[k] \leftarrow 1
alpha0~dflat()
alpha1 \sim dnorm(0,0.001)
alpha2\sim dnorm(0,0.001)
alpha3\sim dnorm(0,0.001)
alpha4\sim dnorm(0,0.001)
alpha5\sim dnorm(0, 0.001)
taus~dgamma(0.001, 0.001)
tau \sim dgamma(0.001, 0.001)
sigmas \leftarrow 1/sqrt(taus)
sigmay \leftarrow 1/sqrt(tau)
}
list(N=25, sumNumNeigh = 144,
```

```
\begin{aligned} &\mathrm{O}{=}\mathrm{c}(25.61307,...,30.85184),\\ &\mathrm{adj}=\mathrm{c}(\ 7,\ 6,\ 2,\\ &.......,\\ &24,\ 20,\ 19\ ),\\ &\mathrm{num}=\mathrm{c}(3,...,3),\\ &\mathrm{x1}{=}\mathrm{c}(0,...,0),\\ &\mathrm{x2}{=}\mathrm{c}(1,...,0),\\ &\mathrm{x3}{=}\mathrm{c}(0,...,0),\\ &\mathrm{x4}{=}\mathrm{c}(0,...,0),\\ &\mathrm{x5}{=}\mathrm{c}(0,...,1))\\ &\mathrm{list}(\ \mathrm{tau}=1,\ \mathrm{taus}=1,\ \mathrm{alpha0}=0,\ \mathrm{alpha1}=0,\ \mathrm{alpha2}=0,\\ &\mathrm{alpha3}=0,\ \mathrm{alpha4}=0,\ \mathrm{alpha5}=0,\ \mathrm{s}{=}\mathrm{c}(0,...,0))\\ &\} \end{aligned}
```