

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos geoestatísticos gaussianos bivariados

Bruno Henrique Fernandes Fonseca

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2008

Bruno Henrique Fernandes Fonseca
Bacharel em Estatística

Modelos geoestatísticos gaussianos bivariados

Orientador:
Prof^a. Dr^a. **Paulo Justiniano Ribeiro Jr.**

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração: Es-
tatística e Experimentação Agronômica

Piracicaba
2008

Dedicatória

À minha avó e meu pai,
Alzira Modesto Fernandes (in memoriam) e
Ezequias Silva da Fonseca (in memoriam),

AGRADECIMENTOS

À minha mãe e à minha madrinha, Maria José Fernandes e Maria Aparecida Fernandes, o incentivo e o amor incondicional.

À minha tia, Lourdes, o apoio e o imenso carinho.

Aos meus tios, Raimundo e Cícero, pelos ensinamentos.

Ao Prof. Dr. Paulo Justiniano Ribeiro Jr. pela compreensão, a paciência e amizade.

À CAPES pelo apoio financeiro, que facilitou bastante a realização dessa pesquisa.

Ao engenheiro agrônomo Bruno Gherardi, que cedeu o conjunto de dados utilizado nessa pesquisa.

Ao conselho do programa de Pós-Graduação em Estatística e Experimentação Agronômica.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, Prof. Dr. Carlos Tadeu dos Santos Dias, Prof^a Dr^a Clarice Garcia Borges Demétrio, Prof. Dr. Décio Barbin, Prof. Dr. Edwin Moisés Ortega, Prof. Dr. Gerson Barreto, Prof^a Dr^a Roseli Aparecida Leandro, Prof. Dr. Silvio Sandoval Zocchi, Prof^a. Dr^a. Sônia Maria Stefano Piedade e Prof. Dr. Vitor Augusto Ozaki.

À Prof^a Dr^a Silvia Emiko Shimakura do Departamento de Estatística da UFPR pelo acolhimento e atenção.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, as secretárias Solange de Assis Paes Sabadin e Luciane Brajão e aos técnicos em informática Jorge Alexandre Wiendl e Eduardo Bonilha, pelos auxílios permanentes.

Aos colegas de turma de mestrado, Cássio, Cláudia, Elizabeth, Elton, Guilherme, Mariana, Marisol, Michele, Paula Klefens, Paula Sanches, Raphael e Sérgio, pelo companheirismo durante o curso.

À Maíra pelo carinho.

Aos amigos de Piracicaba, Eduardo, Fernanda, Juliana, Luiz Alberto, Luiz Eduardo e Luiz Trench, pela presença em muitos momentos.

SUMÁRIO

RESUMO	6
ABSTRACT	7
LISTA DE FIGURAS	8
LISTA DE TABELAS	9
1 INTRODUÇÃO	12
2 REVISÃO DE LITERATURA	14
2.1 Campos aleatórios	14
2.2 Modelos geoestatísticos gaussianos univariados	15
2.2.1 Estimação dos parâmetros	16
2.2.2 Krigagem	17
2.3 Modelos geoestatísticos gaussianos bivariados	18
2.3.1 Modelo gaussiano bivariado com componente de correlação parcialmente comum	18
2.3.2 Modelo bivariado de co-regionalização	19
3 MATERIAL E MÉTODOS	21
3.1 Estudo de Simulação	21
3.2 Dados sobre a qualidade do solo	22
4 RESULTADOS E DISCUSSÃO	24
4.1 Estudo de simulação	24
4.1.1 Modelos bivariados com dados co-locados e balanceados	24
4.1.2 Modelos bivariados com dados co-locados e desbalanceados	27
4.1.3 Modelos bivariados com dados 50% co-locados e balanceados	30
4.1.4 Estudo cruzado dos modelos bivariados com dados co-locados e balanceados	32
4.2 Estudo observacional sobre qualidade do solo	35
4.2.1 Análise exploratória	35
4.2.2 Modelagem univariada do pH	38
4.2.3 Modelagem univariada da saturação por bases	39
4.2.4 Krigagens	40
4.2.5 Análise de resíduos	41
4.2.6 Modelagem bivariada para o pH e a saturação por bases	42
4.2.7 Estudo empírico de comparação entre as abordagens univariada e bivariadas	45
5 CONSIDERAÇÕES FINAIS	47
REFERÊNCIAS	48
APÊNDICE	50

RESUMO

Modelos geoestatísticos gaussianos bivariados

Os modelos geoestatísticos bivariados são um conjunto de ferramentas, que de uma forma geral, tenta propor uma boa função matemática com respaldo probabilístico para dois processos estocásticos com localizações espaciais conhecidas. Por simplicidade inferencial, pode ser adotada a suposição da existência de um campo aleatório gaussiano latente para cada variável aleatória, a distribuição de probabilidade conjunta das variáveis latentes também é gaussiana. A matriz de covariância dessa distribuição deve ser positiva definida e possuir a estrutura de variabilidade espacial entre e dentre os atributos, Gelfand et al. (2004) e Diggle e Ribeiro Jr. (2006) propuseram estratégias para estruturar essa matriz. Não existem muitos relatos sobre o uso e avaliações comparativas entre essas abordagens. Neste trabalho foi conduzido um estudo de simulação de modelos geoestatísticos bivariados em conjunto com estimação por máxima verossimilhança e krigagem ordinária, sob diferentes configurações amostrais de localizações espaciais. Também foram utilizados dados provenientes da análise de solo de uma propriedade agrícola com 51,8ha de área, onde foram amostradas 67 localizações georeferenciadas e mensurados o pH e a saturação por bases do solo, que foram submetidas à análise descritiva espacial, modelagens geoestatísticas univariadas e bivariadas, predições espaciais e análise de pressupostos. Para verificar vantagens quanto a adoção de modelos univariados ou bivariados, a amostra da saturação por bases, que possui coleta mais dispendiosa, foi dividida em uma subamostra de modelagem e uma subamostra de controle, a primeira foi utilizada para fazer as modelagens novamente e a segunda foi utilizada para comparar as precisões das predições espaciais nas localizações omitidas no processo de modelagem. Todas análises foram conduzidas no ambiente R de programação, o *geoR* foi o pacote mais utilizado.

Palavras-chaves: Campos aleatórios gaussianos; Estimadores de máxima verossimilhança; Estudo de simulação; Modelos geoestatísticos bivariados

ABSTRACT

Bivariate gaussian geostatistic models

The bivariate geostatistical models are a set of tools, that generally, try to propose a mathematical function with probabilistic support for two stochastic processes with spatial locations known. For inferential simplicity, can be adopted the assumption of the existence of a latent Gaussian random field for each random variable, the jointly distribution of latent variables is Gaussian. The covariance matrix of this distribution should be positive-definite and to have the spatial variability structure between and among the attributes. Gelfand et al. (2004) and Diggle e Ribeiro Jr. (2006) suggested strategies for structuring this matrix. There are few reports on the use and benchmarking between these approaches. In this search was conducted a simulation study of bivariate models together with maximum likelihood estimators and spatial prediction under different settings of sampling locations space. Were also used data from the analysis of soil from an agricultural property with 51.8 hectares, where two soil quality attributes were observed in 67 spatial locations. The data were submitted to spatial descriptive analysis, univariate and bivariate modeling, spatial prediction and suppose analysis. To verify advantages as the adoption of univariate or bivariate models, the sample of the more expensive variable was divided into a subsample of modeling and a subsample of control, the first was used to fit geostatistical models again, and the second was used to compare the spatial prediction precisions in the locations not used in the modeling process. All analysis were conducted in the programming environment R, the *geoR* package was the more used.

Keywords: Bivariate geostatistical Models; Gaussian random fields; Maximum likelihood estimators; Simulation study

LISTA DE FIGURAS

Figura 1 - Gráficos de círculos - o gráfico a esquerda é para a variável pH e o da direita é da saturação por bases	36
Figura 2 - Gráficos descritivos do padrão espacial do pH	37
Figura 3 - Gráficos descritivos do padrão espacial da saturação por bases	37
Figura 4 - Gráficos de possíveis transformações de variáveis de Box-Cox, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases	38
Figura 5 - Gráficos de predições espaciais, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases	41
Figura 6 - Densidade e gráfico de quartis dos resíduos para o modelo final do pH	41
Figura 7 - Densidade e gráfico de quartis dos resíduos para o modelo final da saturação por bases	42
Figura 8 - Gráficos de predições espaciais com o BGCCM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases	44
Figura 9 - Gráficos de predições espaciais com o BCRM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases	44

LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	24
Tabela 2 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	24
Tabela 3 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	25
Tabela 4 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	25
Tabela 5 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	26
Tabela 6 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	26
Tabela 7 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	26
Tabela 8 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	26
Tabela 9 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	27
Tabela 10 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	27
Tabela 11 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados	27
Tabela 12 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados	28
Tabela 13 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados	28
Tabela 14 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados	28
Tabela 15 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados	29

Tabela 16 -Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados	29
Tabela 17 -Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados	29
Tabela 18 -Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados	29
Tabela 19 -Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados	29
Tabela 20 -Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados	30
Tabela 21 -Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados	30
Tabela 22 -Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados	30
Tabela 23 -Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados	31
Tabela 24 -Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados	31
Tabela 25 -Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados	31
Tabela 26 -Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados	31
Tabela 27 -Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados	32
Tabela 28 -Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados	32
Tabela 29 -Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados	32
Tabela 30 -Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados	32

Tabela 31 -Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	33
Tabela 32 -Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	33
Tabela 33 -Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	34
Tabela 34 -Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	34
Tabela 35 -Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	34
Tabela 36 -Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	34
Tabela 37 -Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados	35
Tabela 38 -Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados	35
Tabela 39 -Médias, medianas e desvios padrões do pH e da saturação por bases	35
Tabela 40 -Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com média constantes	39
Tabela 41 -Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com tendência na média induzida pela área de manejo	39
Tabela 42 -Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com tendência na média induzida pelas coordenadas oeste-leste	39
Tabela 43 -Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com média constantes	40
Tabela 44 -Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com tendência na média induzida pela área de manejo	40
Tabela 45 -Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com tendência na média induzida pelas coordenadas oeste-leste	40
Tabela 46 -Estimativas de máxima verossimilhança dos parâmetros associados ao BGCCM para a saturação por bases e o pH	43
Tabela 47 -Estimativas de máxima verossimilhança dos parâmetros associados ao BCRM para a saturação por bases e o pH	43
Tabela 48 -Valores observados da saturação por bases e não utilizados na modelagem e suas respectivas predições utilizando os modelo univariado, o BGCCM e o BCRM	45
Tabela 49 -Médias e desvios padrões dos erros de krigagem da saturação por bases nas 20 localizações omitidas no processo de estimação	46

1 INTRODUÇÃO

A modelagem estatística é um conjunto de ferramentas muito importante em diversos campos do conhecimento, que utilizam essas técnicas para tentar descrever o comportamento de um ou mais atributos que não possuem um modelo determinístico. De uma forma geral, os modelos estatísticos tentam explicar, o máximo possível, a variabilidade dos processos estocásticos através de uma ou mais variáveis explanatórias que possuam alguma associação ou correlação com a resposta de interesse.

Os primeiros modelos estatísticos propostos foram os lineares univariados, que assumem erros aleatórios independentes e identicamente distribuídos com uma distribuição de probabilidade gaussiana e todos os termos da parte explanatória do modelo fixos. Essas simplificações não são válidas na maioria dos processos naturais, logo, surgiu a necessidade de desenvolver técnicas mais sofisticadas para tentar modelar processos que possuem estruturas mais complexas de variabilidade.

Um campo de pesquisas que teve grande evolução nos últimos tempos foi a estatística espacial, que é formada por três grandes áreas de estudo: geoestatística, dados de área e processos pontuais, que são utilizadas conforme o tipo de dados em questão, neste trabalho será estudada apenas a primeira. A modelagem geoestatística é um conjunto de técnicas que tenta encontrar uma boa função matemática para um ou mais atributos que possuem localizações espaciais conhecidas, sendo assim, essas ferramentas são úteis para capturar a correlação entre as observações das variáveis aleatórias de interesse, onde existe uma forte suspeita de que pontos espaciais mais próximos possuem valores dos atributos mais parecidos. A abordagem geoestatística de modelagem se diferencia dos modelos lineares nos pressupostos, onde agora todas as observações não são independentes e existe efeito aleatório latente na parte explanatória do modelo.

Diversas pesquisas de distintas áreas podem possuir mais de uma variável resposta de interesse, ou seja, os pesquisadores possuem dois ou mais atributos que devem ser modelados. Se esses atributos sob estudo forem independentes deve-se propor um modelo estatístico para cada um deles. Se há evidências de que esses processos não sejam independentes e existindo uma explicação prática, modelos multivariados podem ser propostos, ou seja, os modelos estatísticos devem capturar ao máximo a correlação entre as variáveis respostas, para tal, algumas técnicas têm sido utilizadas, assim como, distribuições de probabilidades conjuntas.

Existindo interesse em mais de uma variável aleatória com padrão espacial de variabilidade, pode-se adotar modelos geoestatísticos multivariados, que devem capturar a correlação entre e dentre as variáveis aleatórias. Na literatura existem algumas formas de estudar esse tipo de problema. Devido a complexidade dos modelos geoestatísticos multivariados e o número elevado de parâmetros podem existir problemas para estimação paramétrica e, além disso, em alguns casos pode ocorrer problemas com a identificabilidade do modelo. Sendo assim, este trabalho, inicialmente apresenta um estudo de simulação de modelos geoestatísticos bivariados, dessa forma pode-se fazer uma comparação entre as metodologias, e detectar quais são as vantagens, probabilísticas e computacionais, de cada método em diversas configurações.

A agricultura de precisão é um campo de pesquisa que pode utilizar os modelos geoestatísticos e de forma bem sucinta, é o conjunto de métodos aplicados ao manejo da variabilidade. SCHUELLER (1992) definiu como um método de administração cuidadosa e detalhada do solo e da cultura para adequar as diferentes condições encontradas em cada pedaço de lavoura, tendo em vista a desuniformidade intrínseca dos solos. Vários autores concluíram que a variabilidade espacial existe, mesmo em áreas consideradas homogêneas, inúmeros trabalhos de campo têm mostrado a importância do estudo das variações das condições do solo como aspecto fundamental para se implementar uma agricultura mais eficiente e rentável. Estes trabalhos mostram que a

variabilidade do solo não é puramente aleatória, apresentando correlação ou dependência espacial. Nesse contexto, diversas abordagens podem ser utilizadas para estruturar a variabilidade espacial das variáveis químicas do solo, em agricultura de precisão, tradicionalmente é utilizada apenas medidas descritivas e a intuição pessoal dos pesquisadores para tal, o que pode gerar o problema de não ser possível mensurar a precisão dos resultados. Uma solução recente para esse tipo de problema, que proporciona resultados com respaldo probabilístico, é a modelagem geoestatística em conjunto com estimação por máxima verossimilhança.

Além do estudo de simulação, essa pesquisa utiliza ferramentas geoestatísticas para estudar a variabilidade espacial de duas variáveis químicas do solo de uma propriedade agrícola, sendo que, a utilização de modelos geoestatísticos bivariados é possível devido a natureza dos atributos, que são fortemente correlacionados e, além disso, uma das resposta é mais dispendiosa para ser observada. Com a estrutura conjunta de correlação espacial estabelecida, em monitoramentos futuros da mesma propriedade agrícola será possível diminuir os gastos com a coleta de informações.

2 REVISÃO DE LITERATURA

2.1 Campos aleatórios

Um campo aleatório é um processo estocástico que existe em algum espaço real d -dimensional, geralmente bi ou tri-dimensional sua definição é dada por:

$$\{Z(s_i) : s_i \in G \subset R^d\},$$

sendo $Z(s_i)$ a notação a variável aleatória Z na localização s_i do espaço sob estudo G .

Segundo Schmidt e Sansó (2006) e Le e Zidek (2006), a descrição de um campo aleatório é obtida através das distribuições acumuladas finito-dimensionais F , para qualquer conjunto de localizações (s_1, s_2, \dots, s_n) pertencentes à região G e qualquer inteiro n :

$$F_{S_1, S_2, \dots, S_n}(z_1, z_2, \dots, z_n) \equiv P(Z(s_1) \leq z_1, Z(s_2) \leq z_2, \dots, Z(s_n) \leq z_n)$$

Devido a sua simplicidade inferencial, a distribuição de probabilidade gaussiana é uma das mais utilizadas na literatura. Sendo assim, um campo aleatório é dito ser gaussiano se Z segue uma distribuição Normal em todas localizações do espaço sob estudo G . Para qualquer conjunto finito de localizações $s = (s_1, s_2, \dots, s_n)$ pertencente a G , $Z(s)$ segue uma distribuição normal n -variada e é completamente especificado pelo vetor de média $n \times 1$, notado por μ , e pela matriz de covariâncias $n \times n$, notada por Σ , que no contexto de geoestatística possui o comportamento empírico de que quanto maior a distância euclidiana entre duas localizações s_l e s_k quaisquer, menor a correlação entre $Z(s_l)$ e $Z(s_k)$. Não é trivial encontrar uma forma para gerar esse comportamento e, ao mesmo tempo, assegurar que a matriz de covariâncias fique positiva definida. Diggle e Ribeiro (2006) mostram maiores detalhes sobre campos aleatórios gaussianos.

Em pesquisas de geoestatística, geralmente, não é possível ter mais de uma realização do processo, sendo assim, outras suposições devem ser impostas sobre o campo aleatório gaussiano para a realização de inferências. A restrição mais utilizada é que o processo estocástico é estacionário, ou seja, a distribuição de probabilidade associada ao campo aleatório não depende da grandeza de escala das coordenadas, logo, a distribuição conjunta de $(Z(s_1), Z(s_2), \dots, Z(s_n))$ é igual a distribuição conjunta de $(Z(s_1 + h), Z(s_2 + h), \dots, Z(s_n + h))$, para qualquer incremento h . Outra definição menos restritiva é que a média do campo aleatório é igual em toda a região sob estudo e a correlação entre $Z(s_l)$ e $Z(s_k)$, para quaisquer s_l e s_k , só depende da distância entre as localizações, ou seja, a grandeza de escala de Z não influencia na estrutura de correlação espacial. Esse tipo de estacionariedade é conhecido na literatura como estacionariedade fraca ou de segunda ordem. Uma observação importante é que a primeira restrição implica na segunda, no entanto, o contrário não é válido, a não ser que o processo espacial seja gaussiano, que produz equivalência entre as duas restrições. No entanto, nem sempre é fácil verificar as restrições de estacionariedade forte ou fraca, logo, outra possibilidade menos restritiva é assumir que os incrementos $[Z(s) - Z(s + h)]$ possuem estacionariedade. Esta característica é denominada de estacionariedade intrínseca (SCHANBENBERGER; GOTWAY, 2005). Sendo assim, um campo aleatório é dito ser intrinsecamente estacionário se para todo s_i pertencente a G , $E[Z(s_i)] = \mu$ e $Var[Z(s_i) - Z(s_i + h)] = 2\gamma(h)$, sendo $\gamma(h) = Var(Z(s_i)) - Cov(Z(s_i); Z(s_i + h))$ e denominado de semivariograma.

Em alguns casos a suposição de estacionariedade não é válida, sendo assim, diversas abordagens são propostas para contornar esse problema. Quando a média do processo estocástico não é constante na região sob estudo, geralmente, utiliza-se a inclusão de covariáveis na modelagem, onde a média é tratada como efeito fixo, e sua interpretabilidade é igual a de modelos lineares,

Diggle e Ribeiro Jr. (2006) detalham melhor essa técnica. Já para problemas com variâncias e covariâncias não constantes, uma técnica mais simples é a utilização de transformação nos valores observados do campo aleatório (CHRISTENSEN; DIGGLE; RIBEIRO Jr., 2001).

Mesmo que um campo aleatório possua algum tipo de estacionariedade, o padrão espacial de correlação pode depender das distâncias e das direções envolvidas entre as localizações, a essa característica é dado o nome de anisotropia, em problemas práticos de geoestatística não é fácil identificar tal característica nos dados observacionais, no entanto, a natureza de algumas variáveis exige a utilização de tal abordagem, como por exemplo, estudos sobre poluição, onde o padrão dos ventos gera correlação espacial dependente das direções entre localizações amostradas. Silva (2006) mostra algumas técnicas utilizadas em problemas de anisotropia geométrica.

Um campo aleatório gaussiano é dito ser homogêneo se ele for estacionário e seu padrão de correlações não depende das direções. Utilizando essa suposição o processo estocástico fica bastante restritivo, porém consegue modelar diversos problemas naturais.

Com a suposição de homogeneidade de um campo aleatório gaussiano, é necessário estabelecer uma função matemática que dependa apenas das distâncias entre as localizações amostradas do espaço sob estudo e que estruture a matriz de covariâncias, de forma que esta seja positiva definida e com o comportamento empírico utilizado em geoestatística. Devido a complexidade para encontrar tais funções, diversas propostas são sugeridas.

Função de Matérn

Essa família de funções de correlação foi proposta por Berfil Matérn (1986) e possui a seguinte forma:

$$\rho(h) = 2^\kappa - \Gamma(\kappa)^{-1} (h/\phi)^\kappa K_\kappa(h/\phi),$$

sendo h a distância euclidiana entre duas localizações quaisquer do campo aleatório, os parâmetros dessa função são $\phi > 0$ e $\kappa > 0$, sendo o primeiro vinculado ao alcance das correlações e quanto menor o parâmetro menor o alcance das correlações, ou seja, somente observações muito próximas possuem correlação significativa. Já o segundo parâmetro é vinculado a suavidade do processo, quanto maior κ , maior a suavidade. E por último, $K_\kappa(\cdot)$ é a função Bessel de ordem κ .

Função Exponencial Potência

$$\rho(h) = \exp(-h/\phi)^\kappa,$$

essa família possui as mesmas características da função de Matérn, sendo que neste caso κ é limitado no intervalo $[0, 2]$.

Essas duas funções são muito utilizadas devido a capacidade de produzir comportamentos distintos quanto a suavidade do processo, ou seja, é possível modelar processos mais ou menos diferenciáveis. Além disso, sob estacionariedade fraca, essas funções possuem propriedade conhecidas e desejáveis, Schabenberger e Gotway (2005) apresentam e discutem detalhes sobre tais propriedades. Silva (2006) apresenta diversas outras funções de correlação conhecidas e válidas.

2.2 Modelos geoestatísticos gaussianos univariados

Considerando que em alguma área G exista um campo aleatório gaussiano Z latente, ou seja, o processo existe mas não é observável, é necessário fazer uma amostragem de n localizações espaciais dentro da área G e observar valores de atributo de interesse nas localizações amostradas. Sendo assim, existe um vetor $Y(s)$ $n \times 1$ de valores observados em $s = (s_1, s_2, \dots, s_n)$, que segundo Diggle e Ribeiro Jr (2006) pode ser modelado como:

$$Y(s) = \mu + Z(s) + \epsilon, \tag{1}$$

onde $\mu = X\beta$, onde X é uma matriz $n \times q$ contendo $q - 1$ possíveis covariáveis, β um vetor $q \times 1$ de parâmetros associados a X , $Z(s)$ um campo aleatório gaussiano que possui vetor de médias $n \times 1$ nulo e matriz de covariâncias Σ de dimensão $n \times n$, onde cada elemento $\Sigma_{i,j}$ é igual a $Cov(Z(s_i); Z(s_j))$, para todo s_i e s_j pertencentes a s , e ϵ um vetor $n \times 1$ de ruídos brancos, que por suposição são independentes e identicamente distribuídos com distribuição de probabilidade normal com média zero e desvio padrão τ .

Utilizando (1) é possível encontrar a distribuição de probabilidade de $Y(s)$, que é gaussiana n -variada, com vetor de médias $X\beta$ e matriz de covariâncias $\Sigma_Y = \Sigma + \tau^2 I$, onde I é uma matriz identidade $n \times n$. Logo, existe um vetor de parâmetros $\theta = (\beta, \sigma^2, \phi^*, \tau^2)$ a ser estimado, onde ϕ^* é um vetor de parâmetros associados a função de correlação utilizada. Cabe observar que $Cov(Z(s_i); Z(s_j)) = \rho(Z(s_i); Z(s_j))\sigma^2$, para todo s_i e s_j pertencentes a s e $\sigma^2 > 0$.

2.2.1 Estimação dos parâmetros

Estabelecidas as estruturas paramétricas, o próximo passo é fazer a estimação dos parâmetros. Se o campo aleatório é intrinsecamente estacionário pode-se trabalhar com uma estimação para o semivariograma, abaixo segue a expressão de uma estimativa empírica para o semivariograma através dos estimadores do método de momentos:

$$\hat{\gamma}(h) = \frac{\sum_{|N(h)|} (Z(s_i) - Z(s_j))^2}{2|N(h)|} \quad (2)$$

em que $|N(h)|$ é o número de pontos abrangidos pela distância h .

Devido a relação entre o semivariograma empírico e as funções de correlação válidas, muitos trabalhos aplicados de geoestatística utilizam um modelo em função dos parâmetros de variabilidade e correlação que se ajuste aos valores calculados em (2), isso pode ser feito por meio de métodos "AD-HOC". Essa abordagem para estimar os parâmetros de variabilidade e da função de correlação podem não ser muito precisos, pois os valores dos semivariogramas empíricos podem se afastar muito do semivariograma real e desconhecido, devido ao tamanho e acaso amostral. Esse tipo de estimação deve ser utilizado, na maioria dos casos, apenas como análise descritiva inicial, Diggle e Ribeiro Jr. (2006) discutem com mais detalhes esse assunto.

Por outro lado, assumindo que o campo aleatório possui estacionariedade forte, pode-se optar por estimadores de máxima verossimilhança ou máxima verossimilhança restrita, que consiste em utilizar os valores observados da variável resposta para encontrar um vetor $\hat{\theta}$ que seja o ponto de máximo da função de verossimilhança associada a θ . Por simplicidade matemática, normalmente utiliza-se o logaritmo da função de verossimilhança para fazer a estimação, que associada a (1) tem a seguinte forma:

$$l(\theta; Y(s)) = -0.5(n \ln(2\pi) + \ln(|\Sigma_Y|)) + (Y(s) - X\beta)^t \Sigma_Y^{-1} (Y(s) - X\beta).$$

Maiores detalhes sobre técnicas e propriedades da estimação por máxima verossimilhança são expostos por Azzalini (1996) e Bickel e Doksum (1976). No contexto de geoestatística, Diggle e Ribeiro Jr. (2006) propuseram a utilização da reparametrização $\nu = \tau/\sigma$, a qual facilita a estimação de θ . O vetor de parâmetros ser estimado passar a ser $\theta^* = (\beta, \sigma^2, \phi^*, \nu^2)$ e Σ_Y pode ser escrita como $\sigma^2 V$, onde V é uma matriz $n \times n$ que depende apenas de ν e ϕ^* . O logaritmo da função de verossimilhança fica da seguinte forma:

$$l(\theta^*; Y(s)) \propto -0.5(n \ln(\sigma^2) - \ln(|V|) - \sigma^{-2} (Y - X\beta)^t V^{-1} (Y - X\beta)) \quad (3)$$

sendo que existem formas analíticas apenas para os estimadores de β e σ^2 :

$$\widehat{\beta} = (X^t V^{-1} X)^{-1} (X^t V^{-1} Y) \quad \widehat{\sigma}^2 = n^{-1} (Y - X\widehat{\beta})^t V^{-1} (Y - X\widehat{\beta})$$

Observe que $\widehat{\beta}$ e $\widehat{\sigma}^2$ são funções dos demais parâmetros e além das formas fechadas para os estimadores, é trivial encontrar a matriz de informação de Fischer observada para os mesmos, logo, é fácil encontrar a matriz de covariâncias associada a $(\widehat{\beta}, \widehat{\sigma}^2)$ (DIGGLE; RIBEIRO Jr., 2006).

Para ϕ^* e ν^2 não existem formas analíticas para os estimadores. Utilizando $\widehat{\beta}$ e $\widehat{\sigma}^2$ em (3), tem-se o logaritmo da função de verossimilhança concentrada, que depende apenas de $\theta_c = (\nu, \phi^*)$.

Para encontrar $\widehat{\theta}_c$ é possível utilizar métodos numéricos de maximização de funções, como por exemplo, o método de Nelder e Mead (1965), o qual calcula $\widehat{\phi}^*$ e $\widehat{\nu}^2$ e a matriz Hessiana estimada, denotada por H . Com os parâmetros da função de máxima verossimilhança concentrada estimados, por invariância é possível encontrar as estimativas de β e σ^2 e τ^2 .

Utilizando as propriedades assintóticas dos estimadores de máxima verossimilhança e o método Delta pode-se encontrar a distribuição de probabilidade de $\widehat{\theta}$, que é $N(\theta; \Sigma_\theta)$, sendo $\Sigma_\theta = \Delta^t \Sigma_{\theta^*} \Delta$, onde a i -ésima coluna de Δ é o vetor $\frac{\partial l(\theta_i)}{\partial \theta^*}$ e:

$$\Sigma_{\theta^*} = \begin{bmatrix} \Sigma_{\beta, \sigma^2} & O \\ O^t & \Sigma_{\theta_c} \end{bmatrix}$$

sendo Σ_{β, σ^2} a matriz de covariâncias de $(\widehat{\beta}, \widehat{\sigma}^2)$, que possui forma analítica, $\Sigma_{\theta_c} = -H^{-1}$ é a matriz de covariâncias de $\widehat{\theta}_c$, e O uma matriz de zeros, pois não é possível recuperar as covariâncias entre (β, σ^2) e θ_c .

Geralmente o objetivo final dos estudos com dados georeferenciados é calcular as predições espaciais em localizações não amostradas. Esse processo de predição é denominado de krigagem e é baseado nas propriedades inferenciais da distribuição normal multivariada.

2.2.2 Krigagem

A krigagem nada mais é do que o processo de predição do campo aleatório em localizações não amostradas. O nome krigagem é uma homenagem ao pesquisador sul-africano D.G. Krige que foi um dos pioneiros em estudos de predição espacial.

Antes de expor as técnicas de krigagem, é necessário utilizar algumas propriedades inferenciais da distribuição gaussiana multivariada. Suponha que o interesse é fazer predição para Z nas localizações s^* , ou seja, devemos fazer predição para $Z(s^*)$ tal que o erro quadrático médio seja mínimo, Diggle e Ribeiro Jr. (2006) mostram que a distribuição de probabilidade de $(Z(s^*)|Y(s))$ gera as predições com melhor precisão, sendo assim:

$$E(Z(s^*)|Y(s)) = \mu_Z + \Sigma_{Z,Y} \Sigma_Y^{-1} (Y(s) - X\beta) \quad (4)$$

sendo μ_Z a média de $Z(s^*)$ e $\Sigma_{Z,Y}$ é a matriz de covariâncias cruzadas entre $Z(s^*)$ e $Y(s)$. Além disso a variância preditiva é conhecida:

$$Var(Z(s^*)|Y(s)) = \Sigma_Z - \Sigma_{Z,Y} \Sigma_Y^{-1} \Sigma_{Z,Y}^t \quad (5)$$

sendo Σ_Z a matriz de covariâncias de $Z(s^*)$. Cabe ressaltar que as expressões (4) e (5) são obtidas através das propriedades da distribuição gaussiana multivariada, Diggle e Ribeiro Jr. (2006) mostram mais detalhes.

Do ponto de vista geoestatístico, as krigagens mais utilizadas são a simples e a ordinária, que utilizam (3) e (4), mas que se diferenciam quanto a suposição de conhecimento sobre os parâmetros (DIGGLE; RIBEIRO Jr., 2006).

2.3 Modelos geoestatísticos gaussianos bivariados

Quando existem dois campos aleatórios gaussianos de interesse, a intuição inicial é que seja ajustado um modelo geoestatístico univariado para cada vetor de valores observados. Porém, em alguns casos existe correlação estatística entre as variáveis aleatórias, o que leva a possibilidade de adotar modelos geoestatísticos bivariados. Cabe ressaltar que somente com essa justificativa estatística não é vantajoso utilizar tal abordagem, é necessário que existam justificativas e vantagens práticas para esse aumento de complexidade dos modelos geoestatísticos, que nesses casos devem estruturar as correlações entre e dentre os campos aleatórios.

Supondo a existência de dois campos aleatórios gaussianos de interesse, pode-se utilizar seguinte o modelo:

$$Y_i = \mu_i + Z_i, \quad i = 1, 2 \quad (6)$$

sendo Y_i um vetor $n_i \times 1$ de valores observados do campo aleatório gaussiano latente Z_i , que possui vetor de médias nulo $n_i \times 1$ e matriz de covariâncias Σ_i $n_i \times n_i$, μ_i é um vetor, que possui os parâmetros de médias associados a Y_i . Observe que em (6) não existem termos de ruídos brancos.

A distribuição de $Y = (Y_1, Y_2)$ é de interesse final e possui distribuição gaussiana n -variada, sendo $n = n_1 + n_2$, com vetor de médias $\mu = (\mu_1, \mu_2)$ e matriz de covariâncias Σ_Y , positiva definida que possui o comportamento empírico de correlações utilizado em geoestatística e que pode ser particionada da forma:

$$\Sigma_Y = \begin{bmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^t & \Sigma_2 \end{bmatrix}$$

onde Σ_i é uma matriz $n_i \times n_i$ das covariâncias da variável Y_i , $i = 1, 2$, e $\Sigma_{1,2}$ uma matriz $n_1 \times n_2$ com as covariâncias cruzadas entre Y_1 e Y_2 .

Nesse contexto, a maior dificuldade é propor uma matriz Σ_Y válida. Modelos separáveis é a abordagem mais simples para resolver esse problema e de forma geral utiliza decomposições os termos latentes de (6).

2.3.1 Modelo gaussiano bivariado com componente de correlação parcialmente comum

Diggle e Ribeiro Jr. (2006) propuseram uma abordagem para problemas geoestatísticos bivariados, a qual é abreviada nesse trabalho por BGCCM e que utiliza as seguintes decomposições dos campos aleatórios latentes de (6):

$$Z_i = \sigma_{0i}S_{0i} + \sigma_i S_i, \quad i = 1, 2, \quad (7)$$

sendo $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$ um vetor de parâmetros de dispersão associados a (6) e S_{01} , S_1 , S_{02} e S_2 campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variâncias unitárias e correlações determinadas por funções de correlação válidas quaisquer. As quais vão gerar empiricamente as correlações cruzadas entre Y_1 e Y_2 , uma vez que, as funções de correlação adotadas para S_{01} e S_{02} devem ser idênticas, sendo assim, três funções de correlação válidas devem ser adotadas e utiliza-se somas das mesmas para gerar uma matriz Σ_Y válida.

Logo, (6) fica da seguinte maneira:

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_{01} + \sigma_1S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_{02} + \sigma_2S_2 \end{cases}$$

Definindo $Y_i(s_l)$ e $Y_i(s_k)$ como observações de Y_i feitas em duas localizações quaisquer s_l e s_k , que estão separados por uma distância euclidiana h , para todo $l, k = 1, 2, \dots, n_i$ e $i = 1, 2$, tem-se que o elemento $\Sigma_{i,(l,k)}$ é dado por $Cov(h) = \sigma_{0i}^2\rho_0(h) + \sigma_i^2\rho_i(h)$, sendo ρ_0 e ρ_i as funções de correlação adotadas para S_{0i} e S_i , respectivamente. Utilizando propriedades básicas de covariâncias, pode-se encontrar $\Sigma_{1,2}$, que é igual a $\sigma_{01}\sigma_{02}R_0$, sendo R_0 uma matriz $n_1 \times n_2$ das correlações cruzadas entre as respostas e depende da função de correlação adotada para S_{01} e S_{02} . Dessa forma Σ_Y fica completamente estruturada. Por simplicidade, suponha que existam apenas duas localizações amostradas, logo:

$$\Sigma_Y = \begin{bmatrix} \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}^2\rho_0(h) + \sigma_1^2\rho_1(h) & \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) \\ \sigma_{01}^2\rho_0(h) + \sigma_1^2\rho_1(h) & \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} \\ \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{02}^2 + \sigma_2^2 & \sigma_{02}^2\rho_0(h) + \sigma_2^2\rho_2(h) \\ \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} & \sigma_{02}^2\rho_0(h) + \sigma_2^2\rho_2(h) & \sigma_{02}^2 + \sigma_2^2 \end{bmatrix}$$

Sendo assim, a distribuição de probabilidade do vetor $Y = (Y_1, Y_2)$ está estabelecida e depende do vetor de parâmetros $\theta = (\beta^*, \sigma^*, \phi_0^*, \phi_1^*, \phi_2^*)$, onde β^* é um vetor de parâmetros associado a μ e ϕ_j^* é um vetor de parâmetros associado a escolha da função ρ_j , para todo $j = 0, 1, 2$. Para fazer a estimação dos parâmetros e a predição espacial pode-se utilizar as mesmas técnicas envolvidas nos modelos geoestatísticos univariados. No entanto, é necessário utilizar mais reparametrizações de θ para encontrar a função de verossimilhança concentrada.

2.3.2 Modelo bivariado de co-regionalização

Gelfand et al. (2004) propôs uma abordagem concorrente ao BGCCM para encontrar uma estrutura paramétrica válida para Σ_Y , que é abreviada por BCRM. Nesta abordagem os termos latentes de (6) são decompostos da seguinte forma:

$$\begin{cases} Z_1 = \sigma_{11}S_{11} \\ Z_2 = \sigma_{12}S_{12} + \sigma_{22}S_{22} \end{cases}$$

sendo S_{11} , S_{12} e S_{22} campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulo, variância unitária e com correlações determinadas pela escolha de funções de correlação conhecidamente válidas. A escolha para S_{11} e S_{12} deve ser idêntica, ou seja, agora a variável Y_1 possui apenas um termo de correlação, que é comum as duas variáveis. Gelfand et al. (2004) mostram mais detalhes sobre essa abordagem do ponto de vista de modelos geoestatísticos.

Utilizando o BCRM a expressão (6) é representada da seguinte forma:

$$\begin{cases} Y_1 = \mu_1 + \sigma_{11}S_{11} \\ Y_2 = \mu_2 + \sigma_{12}S_{12} + \sigma_{22}S_{22} \end{cases}$$

Definindo $Y_i(s_l)$ e $Y_i(s_k)$ como observações feitas da resposta Y_i em duas localizações quaisquer s_l e s_k , separadas por uma distância euclidiana h , para todo $l, k = 1, 2, \dots, n_i$ e $i = 1, 2$, tem-se que o elemento $\Sigma_{1,(l,k)}$ é dado por $Cov(h) = \sigma_{11}^2\rho_1(h)$ e o elemento $\Sigma_{2,(l,k)}$ é dado por

$Cov(h) = \sigma_{12}^2 \rho_1(h) + \sigma_{22}^2 \rho_2(h)$, onde ρ_1 e ρ_2 são as funções de correlação adotadas. Utilizando propriedades básicas de covariâncias, pode-se encontrar $\Sigma_{1,2}$, que é igual a $\sigma_{11} \sigma_{12} R_1$, onde R_1 é uma matriz $n_1 \times n_2$ das correlações cruzadas entre as respostas e depende apenas da função de correlação adotada para Y_1 . Dessa forma Σ_Y fica completamente estruturada. Por simplicidade, suponha que existam apenas duas localizações amostradas, logo:

$$\Sigma_Z = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}^2 \rho_1(h) & \sigma_{11} \sigma_{12} & \sigma_{11} \sigma_{12} \rho_1(h) \\ \sigma_{11}^2 \rho_1(h) & \sigma_{11}^2 & \sigma_{11} \sigma_{12} \rho_1(h) & \sigma_{11} \sigma_{12} \\ \sigma_{11} \sigma_{12} & \sigma_{11} \sigma_{12} \rho_1(h) & \sigma_{12}^2 + \sigma_{22}^2 & \sigma_{12}^2 \rho_1(h) + \sigma_{22}^2 \rho_2(h) \\ \sigma_{11} \sigma_{12} \rho_1(h) & \sigma_{11} \sigma_{12} & \sigma_{12}^2 \rho_1(h) + \sigma_{22}^2 \rho_2(h) & \sigma_{12}^2 + \sigma_{22}^2 \end{bmatrix}$$

Sendo assim, a distribuição de probabilidade do vetor $Y = (Y_1, Y_2)$ está estabelecida e depende do vetor de parâmetros $\theta = (\beta^*, \sigma^*, \phi_1^*, \phi_2^*)$, onde β^* é um vetor de parâmetros associado a μ e ϕ_j^* é um vetor de parâmetros associado a escolha da função ρ_j , para todo $j = 1, 2$. Para fazer a estimação dos parâmetros e a predição espacial pode-se utilizar as mesmas técnicas envolvidas nos modelos geoestatísticos univariados. No entanto, é necessário utilizar mais reparametrizações de θ para encontrar a função de verossimilhança concentrada.

3 MATERIAL E MÉTODOS

Alguns trabalhos recentemente publicados vêm utilizando modelos geoestatísticos bivariados para estruturar a dependência espacial entre e dentre campos aleatórios de interesse. No entanto, não existem pesquisas comparando o BGCCM e o BCRM sob o enfoque frequentista de estimação. Esse trabalho concentra-se em utilizar tais modelos e verificar suas vantagens e desvantagens em análise de dados.

Com o intuito de verificar a qualidade das estimativas por máxima verossimilhança e das krigagens utilizando os modelos geoestatísticos bivariados, inicialmente foi conduzido um estudo de simulação com diversas configurações para o BGCCM e o BCRM.

Após o estudo de simulação foram analisados dados observacionais provenientes de uma pesquisa sobre qualidade do solo de uma propriedade agrícola.

Cabe ressaltar que todas as análises e resultados foram obtidos através do ambiente R de programação (R Development Core Team, 2006), sendo que o pacote mais utilizado foi o *geoR* (Ribeiro Jr. e Diggle, 2001), no entanto, para os modelos bivariados a maioria das técnicas não está implementada nos pacotes do R, sendo assim, consta no apêndice as programações utilizadas.

3.1 Estudo de Simulação

Essa fase do trabalho é muito importante para detectar se existem possíveis problemas com os modelos geoestatísticos bivariados e verificar as vantagens e desvantagens de cada abordagem considerada. Nesse contexto, foram selecionadas 3 configurações amostrais de localizações espaciais para cada abordagem e foram simulados 1000 conjuntos de dados dentro de um quadrado de lado igual a 1 e com grid irregular para cada configuração. Para avaliar as precisões das estimativas e das predições espaciais, cada vetor $Y = (Y_1, Y_2)$ de dados simulados foi dividido em um grupo de modelagem e um grupo de controle. O grupo de modelagem foi utilizado para ajustar os modelos bivariados e o grupo de controle, que possui 20 observações de cada resposta e suas respectivas localizações espaciais, foi utilizado para avaliar os erros de predição nas localizações desse grupo.

Para simular observações com o BGCCM, foram considerados os parâmetros de média e variabilidade sempre fixos em $\mu_1 = 150$, $\mu_2 = 60$, $\sigma_{01} = 8$, $\sigma_1 = 4$, $\sigma_{02} = 5$ e $\sigma_2 = 2$, logo, $Var(Y_1) = \sigma_{01}^2 + \sigma_1^2 = 80$, $Var(Y_2) = \sigma_{02}^2 + \sigma_2^2 = 29$. A proposta de Matérn foi adotada para as três funções de correlação, $\phi_0 = 0,25$, $\phi_1 = 0,2$, $\phi_2 = 0,2$, $\kappa_0 = 0,5$, $\kappa_1 = 0,5$ e $\kappa_2 = 0,5$. Utilizando essa configuração paramétrica tem-se que os dois campos aleatórios são homogêneos e os parâmetros escolhidos para estruturar as correlações geram alcances práticos, distância h em que as $\rho(h) = 0.05$, aproximadamente iguais a 0,75, 0,60 e 0,60 para ρ_0 , ρ_1 e ρ_2 , respectivamente. Já para a amostragem de localizações, foram consideradas três configurações distintas: na primeira foram utilizados $n_1 = n_2 = 100$ e as localizações das duas respostas completamente coincidentes, dados co-locados; na segunda foram utilizados dados completamente co-locados, porém desbalanceados, sendo $n_1 = 100$ e $n_2 = 50$; e na última configuração foi considerado $n_1 = n_2 = 100$ porém apenas metade dos dados co-locados.

Com o BCRM, foram consideradas as configurações das localizações e dos parâmetros de média idênticas as do BGCCM. Os demais parâmetros foram fixos em $\sigma_1 = 9$, $\sigma_{12} = 5$, $\sigma_2 = 2$, $\phi_1 = 0,25$, $\phi_2 = 0,2$, $\kappa_1 = 0,5$ e $\kappa_2 = 0,5$. Tais configurações foram utilizadas para gerar características similares as da outra abordagem, sendo assim, comparações entre as abordagens são mais fáceis de ser realizadas.

Os parâmetros foram estimados por máxima verossimilhança em cada uma das ob-

servações de Y utilizando o mesmo modelo do qual a observação é proveniente. Para facilitar a estimação de θ foram utilizadas reparametrizações. No BGCCM foram definidos $\sigma = \sigma_{01}$, $\eta = \sigma_{02}/\sigma_{01}$, $\nu_1 = \sigma_1/\sigma_{01}$ e $\nu_2 = \sigma_2/\sigma_{01}$, já no BCRM foram utilizados $\sigma = \sigma_{11}$, $\nu_1 = \sigma_{12}/\sigma_{11}$ e $\nu_2 = \sigma_{22}/\sigma_{11}$. Nas duas abordagens pode-se escrever $\Sigma_Y = \sigma^2 V$, sendo assim, existe forma analítica para $\hat{\mu}$ e $\hat{\sigma}$ e V é função dos demais parâmetros, os quais foram estimados maximizando a função de máxima verossimilhança concentrada pelo método de Nelder e Mead (1965).

As técnicas para encontrar $\hat{\theta}$ e sua distribuição de probabilidade assintótica são similares as utilizadas nos modelos geoestatísticos univariados, onde foi utilizada a propriedade de invariância para calcular todas as estimativas e o método Delta para encontrar a matriz de variância de $\hat{\theta}$. Cabe ressaltar que, em nenhum caso foram estimados os parâmetros de suavidade das funções de correlação, os mesmos foram considerados fixos.

Para avaliar a qualidade das estimativas, foram utilizados erros absolutos médios e erros quadráticos médios.

Cada vetor de parâmetros estimados foi utilizado para fazer as krigagens dos campos aleatórios nas localizações do grupo de controle. Sendo assim, para cada modelo ajustado, existem 20 valores observados de cada resposta e suas previsões espaciais. Para avaliar a precisão dessas previsões, foram calculados intervalos de confiança marginais para as previsões, erros quadráticos médios e erros absolutos médios.

3.2 Dados sobre a qualidade do solo

Os dados analisados com as técnicas apresentadas foram obtidos por meio de uma pesquisa realizada em julho 2006 na fazenda Tupã, localizada no município de Echaporan/SP, que possui 51,8ha de área, solo argissolo de textura média e dois históricos de manejo distintos, soja numa região e pastagem na outra. Para analisar as variáveis químicas do solo, foram amostradas 67 localizações, com grid regular a cada hectare, georeferenciadas no sistema Universal Transverse Mercatur (UTM). Com a utilização de um aparelho GPS foram medidos os atributos de interesse.

Dois parâmetros químicos do solo levantados pelos pesquisadores serão modelados nesse trabalho: saturação por bases e pH do solo, a primeira variável é uma medida de capacidade do solo reter bons nutrientes N, P, K, Ca, Mg e a segunda variável mede a acidez do solo, sendo que, existe uma forte correlação entre essas duas respostas, o que justifica a tentativa de modelagem bivariada.

Uma análise exploratória inicial dos dados foi conduzida utilizando gráficos e medidas descritivas, os quais detectaram padrão espacial das variáveis químicas. Do ponto de vista geoestatístico, o pH e a saturação por bases são dois campos aleatórios gaussianos e latentes. A intuição inicial é que seja ajustado individualmente o modelo (1) para cada vetor de valores observados, que foram notados por Y_1 e Y_2 para a saturação por base e o pH, respectivamente. Então Y_i segue uma distribuição gaussiana 67-variada, com matriz de covariâncias Σ_i e vetor de médias $\mu_i = X_i \beta_i^*$, onde X_i é uma matriz $67 \times p$ com o intercepto e $p - 1$ possíveis covariáveis e β_i^* um vetor $p \times 1$ com os parâmetros associados a μ_i .

Utilizando os resultados da análise exploratória, foram consideradas três formas diferentes para X_i , média constante, média com tendência induzida pela área de manejo e média com tendência induzida pela coordenada oeste-leste das localizações amostradas. Foi utilizada a família de funções de correlação de Matérn para estruturar a matriz 67×67 de correlações, considerando o parâmetro de suavidade κ fixo nos valores 0,5, 1, 1.5, 2, 2.5, logo, ϕ_i^* depende apenas de ϕ_i . Combinando as diferentes escolhas de κ e X_i foram ajustados diversos modelos. As estimativas paramétricas foram calculadas por máxima verossimilhança, onde foi utilizado a reparametrização $\nu_i = \tau_i/\sigma_i$ e o método numérico de Nelder e Mead (1965) para encontrar as esti-

mativas $\hat{\phi}_i$ e $\hat{\nu}_i$ que maximizam a função de máxima verossimilhança concentrada, logo, utilizando a propriedade de invariância dos estimadores em questão foram feitas as demais estimativas de interesse.

Os valores dos máximos estimados da função de máxima verossimilhança concentrada e o Critério de Informação de Akaike (AIC) foram utilizados para selecionar o modelo final de cada atributo. Com os parâmetros estimados dos modelos finais foram conduzidas as krigagens ordinárias para as variáveis químicas do solo. Por último, uma análise de resíduos foi realizada para verificação de pressupostos dos ruídos brancos, os quais se comportaram conforme o esperado.

Após estudo individual de cada variável química, foram ajustados modelos bivariados para o vetor $Y = (Y_1, Y_2)$, utilizando as proposições de Diggle e Ribeiro Jr. (2006) e de Matheron (1982). Nas duas abordagens foi utilizado $\mu = X\beta^*$, onde X é uma matriz $134 \times p$ contendo um intercepto para cada resposta e $p - 2$ possíveis covariáveis e β^* é um vetor $p \times 1$ com os parâmetros de média associados a Y . Para estruturar Σ_Y somente funções de correlação da família de Matérn foram utilizadas, sendo que o parâmetro κ foi sempre considerado fixo em valores similares aos modelos univariados.

Combinando diferentes de escolhas de κ e X , diversos modelos foram ajustados. Em todos os casos o vetor de parâmetros θ foi estimado por máxima verossimilhança, sendo que foram utilizadas as mesmas técnicas dos modelos univariados. No BGCCM foi lançado mão das reparametrizações $\sigma = \sigma_{01}$, $\eta = \sigma_{02}/\sigma_{01}$, $\nu_1 = \sigma_1/\sigma_{01}$ e $\nu_2 = \sigma_2/\sigma_{01}$ e no BCRM das reparametrizações $\sigma = \sigma_1$, $\nu_1 = \sigma_{12}/\sigma_{11}$ e $\nu_2 = \sigma_{22}/\sigma_{11}$. Então, nas duas abordagens existe forma analítica para os estimadores de β^* e σ^2 , que são função dos demais parâmetros, os quais foram estimados pelo método de Nelder e Mead (1965). Utilizando a propriedade de invariância dos estimadores foi possível encontrar estimativas para todos os parâmetros dos modelos. Quanto a seleção de modelos e a krigagem, as mesmas ferramentas dos modelos univariados foram utilizadas.

A abordagem de modelos geoestatísticos bivariados é uma possibilidade para esse conjunto de dados, uma vez que, os dois atributos químicos em questão são altamente correlacionados e a coleta da informação sobre a saturação por bases é mais dispendiosa. Como a fazenda continuará sendo monitorada periodicamente, é interessante estruturar a correlação espacial entre e dentre os atributos, pois nas próximas análises de solo é possível amostrar menos pontos georeferenciados para observar a saturação por bases e utilizar a informação do pH para fazer boas inferências. Para exemplificar tal técnica, a amostra da saturação por bases foi dividida em uma subamostra de modelagem e uma subamostra de controle, que contêm 47 e 20 observações, respectivamente. O grupo de modelagem foi utilizado para modelar novamente a saturação por bases sob as abordagens uni e bivariadas, cabe ressaltar que nos modelos bivariados foi utilizada a informação completa do pH. Com os parâmetros estimados foram calculadas as previsões da saturação por bases nas localizações do grupo de controle. Para comparar as metodologias, por último foram calculados os erros de krigagem sob cada abordagem.

4 RESULTADOS E DISCUSSÃO

4.1 Estudo de simulação

Este tópico tem o objetivo de avaliar a qualidade das estimativas paramétricas por máxima verossimilhança e da capacidade de predição espacial sob as duas abordagens de modelos geoestatísticos bivariados considerados nesse estudo.

4.1.1 Modelos bivariados com dados co-locados e balanceados

As Tabelas 1 e 2 mostram as estatísticas descritivas das estimativas paramétricas por máxima verossimilhança do BGCCM e do BCRM, respectivamente. As estimativas dos parâmetros de média se comportaram bem, uma vez que os valores de média e mediana se aproximam muito dos valores utilizados na simulação e existe uma pequena amplitude nas estimativas. Com relação aos parâmetros de variabilidade e de correlação, nas duas abordagens as estimativas possuem uma maior amplitude e em sua maioria ocorre uma pequena subestimação em média.

Tabela 1 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	137,03	147,39	150,01	149,99	152,48	161,82
β_{02}	52,18	58,56	60,02	59,98	61,43	66,32
σ_{01}	4,13	6,24	7,22	7,41	8,35	16,47
σ_1	> 0,001	2,14	3,58	3,18	4,29	7,81
σ_{02}	2,12	3,82	4,44	4,58	5,17	9,11
σ_2	> 0,001	0,38	1,65	1,54	2,40	5,20
ϕ_0	0,05	0,14	0,19	0,22	0,27	1,00
ϕ_1	> 0,001	0,11	0,15	0,19	0,22	1,00
ϕ_2	> 0,001	0,11	0,17	0,21	0,28	1,00

Tabela 2 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	138,39	147,67	150,07	149,99	152,45	162,05
β_{02}	53,86	58,63	60,03	59,98	61,37	67,17
σ_{11}	5,19	7,40	8,16	8,41	9,12	16,00
σ_{12}	2,69	4,10	4,54	4,68	5,14	9,55
σ_{22}	1,22	1,66	1,84	1,88	2,05	3,21
ϕ_1	0,07	0,16	0,20	0,22	0,26	0,79
ϕ_2	0,06	0,13	0,16	0,18	0,22	0,58

As Tabelas 3 e 4 mostram as estatísticas descritivas das variâncias das estimativas paramétricas sob as duas abordagens de modelos considerados. De forma geral, as variâncias das estimativas são relativamente pequenas. O BCRM resultou variâncias das estimativas com menor amplitude.

Tabela 3 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,78	4,97	8,58	12,03	15,10	137,48
β_{02}	0,40	1,81	3,06	4,40	5,55	43,38
σ_{01}	0,04	0,10	0,13	0,14	0,17	0,68
σ_1	> 0,001	0,04	1,93	28,97	9,63	2467,42
σ_{02}	0,01	0,07	0,33	4,78	1,67	386,51
σ_2	> 0,001	0,12	0,42	6,93	1,92	797,07
ϕ_0	0,00	0,00	0,002	0,01	0,009	2,48
ϕ_1	0,00	0,00	0,003	0,04	0,01	2,56
ϕ_2	0,00	0,00	0,002	0,07	0,02	7,70

Tabela 4 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,96	5,42	8,55	12,36	14,22	106,43
β_{02}	0,59	2,07	3,10	4,31	5,03	36,35
σ_{11}	0,07	0,14	0,17	0,18	0,21	0,64
σ_{12}	0,04	0,08	0,09	0,10	0,11	0,37
σ_{22}	0,04	0,13	0,19	0,23	0,29	1,41
ϕ_1	> 0,001	0,003	0,005	0,02	0,01	0,48
ϕ_2	> 0,001	0,002	0,003	0,007	0,07	0,15

Com as variâncias das estimativas calculadas, a primeira intuição é que sejam calculados intervalos de confiança marginais ou conjuntos para os parâmetros. Esses intervalos não são precisos e existem outros problemas com esse tipo de análise. Como em modelos geoestatísticos geralmente existe apenas uma observação de um campo aleatório, há problema com a teoria assintótica para os estimadores de máxima verossimilhança. Outro problema é que a distribuição assintótica de $\hat{\theta}$ considera independência entre os parâmetros estimados com fórmulas analíticas e os parâmetros estimados por métodos numéricos. Além disso, a matriz de covariâncias dos parâmetros associados a função de verossimilhança concentrada é estimada através da matriz de informação observada de Fischer, que nesse caso é estimada através da matriz hessiana resultante do método numérico, que em muitos casos possui comportamento estranho, possivelmente devido a instabilidade computacional.

Para contornar os problemas com os intervalos de confiança, pode-se adotar a abordagem com funções de verossimilhança perfilada. Essa técnica demanda muito tempo computacional, fato que nesse estudo de simulação inviabilizou sua utilização.

As Tabelas 5 e 6 trazem os erros absolutos e quadráticos médios dos parâmetros nas 1000 estimações paramétricas de cada abordagem. As estimativas erraram relativamente pouco com relação aos parâmetros de simulação, sendo que, o BCRM estima os parâmetros com um pouco mais de precisão.

Tabela 5 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

Estadística	β_{01}	β_{02}	σ_{01}	σ_1	σ_{02}	σ_2	ϕ_0	ϕ_1	ϕ_2
<i>EAM</i>	2,98	1,71	1,44	1,41	0,93	1,01	0,094	0,102	0,116
<i>EQM</i>	14,04	4,79	3,17	3,65	1,27	1,52	0,015	0,021	0,024

Tabela 6 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

Estadística	β_{01}	β_{02}	σ_{11}	σ_{12}	σ_{22}	ϕ_1	ϕ_2
<i>EAM</i>	2,89	1,67	1,33	0,76	0,26	0,08	0,06
<i>EQM</i>	13,17	4,46	2,66	0,88	0,10	0,01	0,005

Após as estimações paramétricas foram conduzidas as krigagens dos dois campos aleatórios nas localizações dos grupos de controle. Para cada amostra foram calculados os erros absolutos e quadráticos médios de predição. As Tabelas 7 e 8 apresentam as estatísticas descritivas dessas medidas de erro. Os dois modelos se mostraram muito eficientes para predição espacial.

Tabela 7 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
<i>EAM_{Y₁}</i>	1,79	3,29	3,78	3,83	4,30	6,21
<i>EAM_{Y₂}</i>	1,14	2,00	2,30	2,31	2,58	4,06
<i>EQM_{Y₁}</i>	5,67	17,29	22,24	23,56	28,55	54,45
<i>EQM_{Y₂}</i>	2,05	6,32	8,13	8,50	10,20	21,95

Tabela 8 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
<i>EAM_{Y₁}</i>	1,35	3,24	3,71	3,76	4,21	6,87
<i>EAM_{Y₂}</i>	1,00	2,00	2,27	2,29	2,58	3,75
<i>EQM_{Y₁}</i>	3,06	16,52	21,40	22,68	27,33	57,87
<i>EQM_{Y₂}</i>	1,64	6,22	7,97	8,38	10,04	21,09

As Tabelas 9 e 10 corroboram a idéia de boa precisão das predições, uma vez que, as médias das variâncias de predição são pequenas. Com essas variâncias foram construídos intervalos de confiança para as krigagens, nas duas abodagens mais de 95% de todos os intervalos de confiança contemplam os valores simulados das duas respostas dos grupos de controle.

Tabela 9 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,70	4,52	4,74	4,75	4,99	5,83
\overline{EP}_{Y_2}	2,21	2,71	2,84	2,85	2,98	3,68

Tabela 10 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,80	4,46	4,68	4,70	4,93	5,93
\overline{EP}_{Y_2}	2,26	2,71	2,86	2,86	3,00	3,54

4.1.2 Modelos bivariados com dados co-locados e desbalanceados

As Tabelas 11 e 12 mostram as estatísticas descritivas das estimativas para essa configuração amostral. Os resultados são bem parecidos com os da configuração do tópico anterior.

As Tabelas 13 e 14 apresentam os resultados das variâncias das estimativas por máxima verossimilhança, os resultados também são similares aos da configuração com dados co-locados e balanceados.

As Tabelas 15 e 16 apresentam as estatísticas descritivas das medidas de erro de estimação das duas abordagens de modelos bivariados. Os dois modelos estão errando pouco na estimação, mas o BCRM se mostrou um pouco mais eficiente do que o BGCCM.

Com relação as predições espaciais, os resultados são bem parecidos aos da última configuração analisada. Os dois modelos são eficientes para fazer krigagem. As Tabelas de 17 à 20 apresentam os resultados.

Tabela 11 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	137,94	147,31	149,96	149,92	152,32	161,64
β_{02}	52,05	58,55	60,01	59,95	61,38	66,44
σ_{01}	3,68	6,22	7,22	7,43	8,42	16,46
σ_1	> 0,001	2,09	3,54	3,12	4,18	8,77
σ_{02}	2,22	3,81	4,42	4,55	5,17	8,50
σ_2	> 0,001	0,01	1,59	1,47	2,35	5,62
ϕ_0	0,05	0,14	0,20	0,22	0,27	0,98
ϕ_1	> 0,001	0,10	0,16	0,20	0,24	1,00
ϕ_2	> 0,001	0,09	0,16	0,20	0,28	1,00

Tabela 12 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	138,39	147,62	150,00	149,99	152,46	162,05
β_{02}	54,00	58,52	60,01	59,97	61,43	67,45
σ_{11}	5,00	7,38	8,16	8,39	9,12	16,00
σ_{12}	2,76	4,05	4,55	4,68	5,17	10,30
σ_{22}	1,11	1,62	1,83	1,85	2,04	3,46
ϕ_1	0,07	0,16	0,20	0,22	0,26	0,79
ϕ_2	0,03	0,12	0,16	0,17	0,21	0,80

Tabela 13 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,78	4,95	8,82	11,98	14,95	137,19
β_{02}	0,35	1,78	3,12	4,37	5,40	35,87
σ_{01}	0,05	0,13	0,17	0,19	0,24	0,90
σ_1	> 0,001	0,05	1,46	27,11	8,85	2164,17
σ_{02}	0,02	0,08	0,31	4,51	1,59	595,88
σ_2	> 0,001	> 0,001	0,30	5,74	1,50	460,65
ϕ_0	0,00	0,00	0,001	0,01	0,007	0,80
ϕ_1	0,00	0,00	0,002	0,18	0,01	74,62
ϕ_2	0,00	0,00	0,001	0,18	0,02	53,89

Tabela 14 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,93	5,39	8,49	12,31	14,20	106,45
β_{02}	0,58	2,08	3,18	4,32	5,06	42,01
σ_{11}	0,08	0,18	0,22	0,24	0,28	0,85
σ_{12}	0,06	0,12	0,15	0,16	0,18	0,65
σ_{22}	0,05	0,14	0,20	0,25	0,31	2,69
ϕ_1	> 0,001	0,003	0,006	0,02	0,01	0,47
ϕ_2	> 0,001	0,002	0,004	0,01	0,01	0,61

Tabela 15 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados

Estadística	β_{01}	β_{02}	σ_{01}	σ_1	σ_{02}	σ_2	ϕ_0	ϕ_1	ϕ_2
<i>EAM</i>	2,91	1,71	1,44	1,41	0,94	1,02	0,095	0,109	0,123
<i>EQM</i>	13,50	4,85	3,10	3,81	1,30	1,58	0,015	0,024	0,026

Tabela 16 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados

Estadística	β_{01}	β_{02}	σ_{11}	σ_{12}	σ_{22}	ϕ_1	ϕ_2
<i>EAM</i>	2,89	1,67	1,34	0,78	0,29	0,08	0,07
<i>EQM</i>	13,05	4,42	2,71	0,94	0,12	0,01	0,007

Tabela 17 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
<i>EAM</i> _{Y₁}	1,83	3,29	3,76	3,82	4,28	6,18
<i>EAM</i> _{Y₂}	1,22	2,09	2,40	2,41	2,70	4,04
<i>EQM</i> _{Y₁}	5,94	17,10	22,11	23,40	28,13	54,63
<i>EQM</i> _{Y₂}	2,30	6,96	8,90	9,29	11,17	23,56

Tabela 18 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
<i>EAM</i> _{Y₁}	1,35	3,23	3,71	3,76	4,21	6,87
<i>EAM</i> _{Y₂}	1,11	2,03	2,31	2,34	2,60	3,91
<i>EQM</i> _{Y₁}	3,06	16,52	21,28	22,66	27,33	57,87
<i>EQM</i> _{Y₂}	1,81	6,44	8,28	8,74	10,55	21,09

Tabela 19 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados colocados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP} _{Y₁}	3,71	4,54	4,74	4,77	5,01	6,01
\overline{EP} _{Y₂}	2,22	2,77	2,94	2,96	3,16	3,81

Tabela 20 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e desbalanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,76	4,46	4,68	4,70	4,93	5,93
\overline{EP}_{Y_2}	2,15	2,76	2,91	2,92	3,08	3,76

4.1.3 Modelos bivariados com dados 50% co-locados e balanceados

As Tabelas de 21 à 30 mostram que os resultados com essa configuração são parecidos com os resultados das demais configurações.

De forma geral, a configuração das localizações amostrais aparentemente não influencia muito na estimação dos parâmetros e na predição espacial.

Tabela 21 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	137,86	147,60	150,17	150,12	152,54	163,84
β_{02}	52,90	58,57	60,15	60,09	61,52	66,54
σ_{01}	4,02	6,22	7,21	7,42	8,30	16,64
σ_1	> 0,001	2,21	3,56	3,13	4,24	8,18
σ_{02}	2,30	3,92	4,49	4,59	5,20	9,21
σ_2	> 0,001	0,01	1,62	1,50	2,45	4,82
ϕ_0	0,06	0,15	0,20	0,22	0,27	1,00
ϕ_1	> 0,001	0,10	0,15	0,18	0,23	1,00
ϕ_2	> 0,001	0,10	0,18	0,22	0,30	1,00

Tabela 22 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	135,74	147,46	149,67	149,81	152,10	163,08
β_{02}	52,13	58,58	59,79	59,82	61,16	67,26
σ_{11}	5,34	7,45	8,20	8,42	9,21	15,77
σ_{12}	2,75	4,12	4,56	4,68	5,15	8,63
σ_{22}	1,19	1,63	1,81	1,85	2,01	3,27
ϕ_1	0,08	0,17	0,20	0,22	0,26	0,75
ϕ_2	0,004	0,12	0,16	0,17	0,21	0,69

Tabela 23 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,67	4,85	8,32	11,06	13,27	140,18
β_{02}	0,30	1,84	2,95	4,04	5,01	42,79
σ_{01}	0,04	0,10	0,13	0,14	0,17	0,70
σ_1	> 0,001	0,04	1,54	25,95	7,18	3974,537
σ_{02}	0,01	0,06	0,32	4,26	1,30	500,87
σ_2	> 0,001	> 0,001	0,31	11,50	1,52	3236,42
ϕ_0	0,00	0,00	0,001	0,01	0,007	1,59
ϕ_1	0,00	0,00	0,002	0,13	0,01	28,20
ϕ_2	0,00	0,00	0,001	0,19	0,02	47,58

Tabela 24 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	1,04	5,40	8,08	11,06	13,76	102,51
β_{02}	0,63	1,96	2,86	3,77	4,69	30,74
σ_{11}	0,07	0,14	0,17	0,18	0,21	0,62
σ_{12}	0,05	0,10	0,12	0,12	0,14	0,41
σ_{22}	0,05	0,13	0,18	0,24	0,27	2,22
ϕ_1	> 0,001	0,003	0,005	0,01	0,01	0,38
ϕ_2	> 0,001	0,002	0,004	0,009	0,008	0,31

Tabela 25 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados

Estatística	β_{01}	β_{02}	σ_{01}	σ_1	σ_{02}	σ_2	ϕ_0	ϕ_1	ϕ_2
<i>EAM</i>	2,90	1,71	1,41	1,37	0,88	1,06	0,087	0,103	0,125
<i>EQM</i>	13,41	4,47	3,03	3,63	1,18	1,63	0,013	0,019	0,029

Tabela 26 - Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados

Estatística	β_{01}	β_{02}	σ_{11}	σ_{12}	σ_{22}	ϕ_1	ϕ_2
<i>EAM</i>	2,70	1,57	1,24	0,71	0,29	0,07	0,07
<i>EQM</i>	11,41	3,95	2,28	0,74	0,12	0,008	0,007

Tabela 27 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
EAM_{Y_1}	1,78	3,12	3,55	3,59	4,06	5,94
EAM_{Y_2}	0,91	1,78	2,02	2,04	2,30	3,29
EQM_{Y_1}	4,90	15,05	19,53	20,60	25,46	51,98
EQM_{Y_2}	1,25	5,00	6,47	6,83	8,26	17,96

Tabela 28 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
EAM_{Y_1}	1,72	2,97	3,36	3,40	3,78	5,82
EAM_{Y_2}	0,99	1,77	2,01	2,03	2,28	3,88
EQM_{Y_1}	4,77	13,86	17,86	18,68	22,16	50,21
EQM_{Y_2}	1,73	4,83	6,35	6,74	8,01	23,04

Tabela 29 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,55	4,25	4,45	4,45	4,65	5,50
\overline{EP}_{Y_2}	2,03	2,45	2,57	2,57	2,70	3,13

Tabela 30 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados 50% co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,38	4,07	4,26	4,26	4,44	5,14
\overline{EP}_{Y_2}	1,97	2,44	2,55	2,55	2,66	3,06

4.1.4 Estudo cruzado dos modelos bivariados com dados co-locados e balanceados

Nas seções anteriores foram expostos os resultados das estimações paramétricas e das previsões espaciais utilizando o mesmo modelo do qual os dados foram simulados. Esta seção se dedica a investigar se cada uma das abordagens de modelos bivariados é robusto para estimar parâmetros e fazer krigagem com dados provenientes da abordagem concorrente.

As Tabelas 31 e 32 apresentam as estatísticas descritivas das estimativas paramétricas. Os parâmetros de variância estimados pelo BGCCM mostraram que a maior parte das variabilidades das respostas está sendo explicada pelos parâmetros σ_{01} e σ_{02} , além disso, em média a

estimativa do parâmetro σ_2 foi muito pequena, fato que inflacionou as estimativas de ϕ_2 .

As Tabelas 33 e 34 mostram que as variâncias das estimativas são relativamente pequenas. Cabe ressaltar que não foi possível calcular as medidas de erro das estimativas, pois os parâmetros estimados são da abordagem concorrente da utilizada na simulação dos dados.

Como em estudos com dados georeferenciados o maior interesse é na predição espacial, foram conduzidas krigagens nas localizações dos grupos controles do processo de simulação. Os dois modelos bivariados se mostraram robustos para fazer predição com a utilização de modelos diferentes nos processos de simulação e estimação, no entanto, o BGCCM apresentou resultados um pouco melhores. As Tabelas de 35 à 38 mostram os resultados.

Tabela 31 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	137,13	147,52	150,01	150,08	152,71	162,00
β_{02}	53,05	58,55	60,07	60,04	61,53	67,17
σ_{01}	4,45	6,56	7,40	7,67	8,41	16,28
σ_1	> 0,001	2,34	2,79	2,50	3,15	5,47
σ_{02}	2,70	4,22	4,75	4,88	5,44	11,06
σ_2	> 0,001	> 0,001	0,005	0,54	1,19	3,30
ϕ_0	0,05	0,14	0,19	0,21	0,25	0,77
ϕ_1	> 0,001	0,12	0,16	0,20	0,23	1,00
ϕ_2	> 0,001	0,21	0,36	0,38	0,50	1,00

Tabela 32 - Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	137,99	147,68	149,99	149,96	152,19	161,53
β_{02}	52,06	58,70	60,00	59,96	61,22	66,38
σ_{11}	4,89	7,46	8,24	8,40	9,17	16,43
σ_{12}	2,12	3,56	3,96	4,05	4,45	7,99
σ_{22}	1,78	2,46	2,72	2,78	3,02	5,16
ϕ_1	0,07	0,16	0,20	0,21	0,25	0,93
ϕ_2	0,06	0,13	0,17	0,18	0,21	0,69

Tabela 33 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,76	4,79	7,83	11,34	13,28	120,41
β_{02}	0,38	1,76	2,88	4,28	4,97	55,20
σ_{01}	0,05	0,11	0,14	0,15	0,18	0,66
σ_1	> 0,001	0,02	0,02	3,44	0,36	510,80
σ_{02}	0,02	0,05	0,07	0,41	0,11	62,36
σ_2	> 0,001	> 0,001	> 0,001	2,85	0,07	859,72
ϕ_0	0,00	0,00	0,00	0,005	> 0,001	0,94
ϕ_1	0,00	0,00	0,00	0,04	0,001	6,75
ϕ_2	0,00	0,00	0,00	0,15	> 0,001	42,84

Tabela 34 - Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
β_{01}	0,75	5,43	8,72	11,47	14,04	136,50
β_{02}	0,51	2,08	3,02	3,73	4,53	32,92
σ_{11}	0,06	0,14	0,17	0,18	0,21	0,67
σ_{12}	0,05	0,10	0,12	0,13	0,15	0,50
σ_{22}	0,11	0,28	0,40	0,51	0,61	5,24
ϕ_1	> 0,001	0,003	0,005	0,01	0,01	0,81
ϕ_2	> 0,001	0,002	0,003	0,008	0,007	0,30

Tabela 35 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
EAM_{Y_1}	1,35	3,25	3,71	3,75	4,22	6,87
EAM_{Y_2}	1,01	2,00	2,27	2,28	2,57	3,78
EQM_{Y_1}	3,05	16,42	21,46	22,61	27,36	57,87
EQM_{Y_2}	1,64	6,23	7,98	8,36	10,03	21,13

Tabela 36 - Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
EAM_{Y_1}	1,79	3,28	3,75	3,80	4,29	6,21
EAM_{Y_2}	1,13	2,00	2,28	2,30	2,57	4,06
EQM_{Y_1}	5,80	17,32	22,04	23,37	28,21	54,42
EQM_{Y_2}	1,92	6,33	8,07	8,45	10,10	22,09

Tabela 37 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BGCCM com dados colocados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,69	4,53	4,76	4,77	5,02	5,96
\overline{EP}_{Y_2}	2,21	2,73	2,86	2,87	3,00	3,70

Tabela 38 - Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BCRM com dados colocados e balanceados

θ	mínimo	1º quartil	mediana	média	3º quartil	máximo
\overline{EP}_{Y_1}	3,76	4,46	4,69	4,70	4,94	5,93
\overline{EP}_{Y_2}	2,24	2,70	2,86	2,85	2,99	3,56

4.2 Estudo observacional sobre qualidade do solo

Com o intuito de apresentar as técnicas da geoestatística aplicadas à agricultura de precisão, esse tópico da pesquisa traz uma análise geoestatística detalhada do conjunto de dados provenientes da fazenda Tupã.

4.2.1 Análise exploratória

A Figura 1 mostra os gráficos de círculos das variáveis sob estudo, onde os círculos representam os valores mensurados dos atributos nas 67 localização amostrada da propriedade agrícola. Quanto maior o círculo mais elevado é o valor observado do atributo. Essa primeira análise gráfica evidência que existe um possível padrão espacial nas variáveis, onde há uma suavidade dos processos ao longo da fazenda. A linha no meio dos gráficos separa as áreas com históricos de manejo distintos: cultivo de soja e pastagem, para as regiões a direita e a esquerda, respectivamente. A primeira área possui valores mais elevados, o que leva a possibilidade de ser considerada essa informação como covariável no processo de modelagem. Além disso, os gráficos são bem parecidos para as duas variáveis, o que corrobora a idéia de que os dados são fortemente correlacionados.

A Tabela 39 mostra as estatísticas descritivas do pH e da saturação por bases, nos dois casos a mediana é próxima da média e existe uma variabilidade relativamente pequena dos dados. De uma maneira geral, os atributos de solo, estão bem comportados, onde há uma simetria moderada dos dados com relação a média amostral e não existem valores observados muito discrepantes.

Tabela 39 - Médias, medianas e desvios padrões do pH e da saturação por bases

Variável	Mediana	Média	D.P.
Saturação	56,00	53,27	10,05
pH	5,00	4,94	0,42

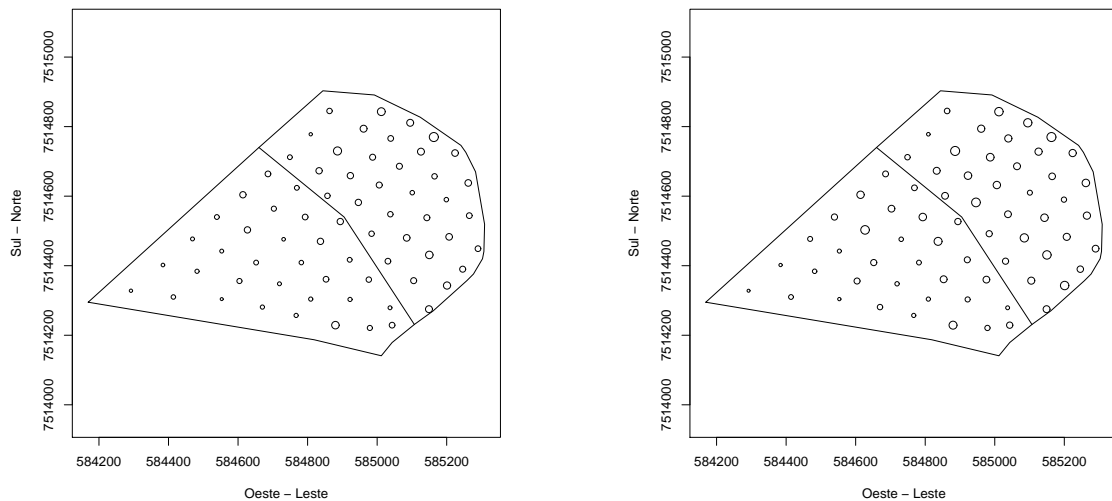


Figura 1 - Gráficos de círculos - o gráfico a esquerda é para a variável pH e o da direita é da saturação por bases

Os gráficos dos cantos superiores esquerdos das Figuras 2 e 3 categorizam os dados nos quartis amostrais das observações, onde quanto mais fria a cor menor o quartil amostral, essas imagens corroboram a idéia que de existe padrão espacial nos dados, uma vez que, existem conglomerados das categorias.

Os gráficos dos cantos inferiores esquerdos das Figuras 2 e 3 mostram que quanto maior o valor da coordenada x , oeste-leste, maior os valores observados dos atributos. Logo, essas coordenadas foram consideradas como possível covariável no processo de modelagem.

Para análise exploratória do pressuposto de gaussianidade das respostas foram analisados os gráficos dos cantos inferiores direitos da Figuras 2 e 3, os quais mostram as densidades amostrais dos dados desconsiderando o possível padrão espacial.

Como existe uma pequena fulga da distribuição gaussiana, análises de possíveis transformações de variáveis foram conduzidas, a Figura 4 mostra qual transformação de Box-Cox maximiza a probabilidade dos dados serem provenientes da distribuição normal, no entanto, como os intervalos de confiança para λ contemplam o valor 1, não foi utilizada nenhuma transformação das variáveis e a suposição de campos aleatórios gaussianos homogêneos foi considerada como válida.

Essa análise descritiva dos dados evidenciou que as variáveis químicas do solo aparentemente possuem padrão espacial de variabilidade e indicou que as coordenadas oeste-leste e a área de manejo são possíveis covariáveis associadas as respostas.

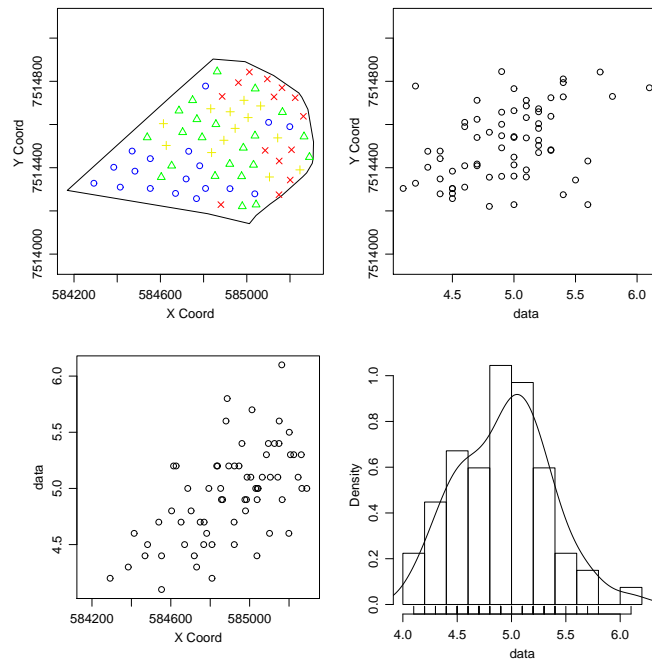


Figura 2 - Gráficos descritivos do padrão espacial do pH

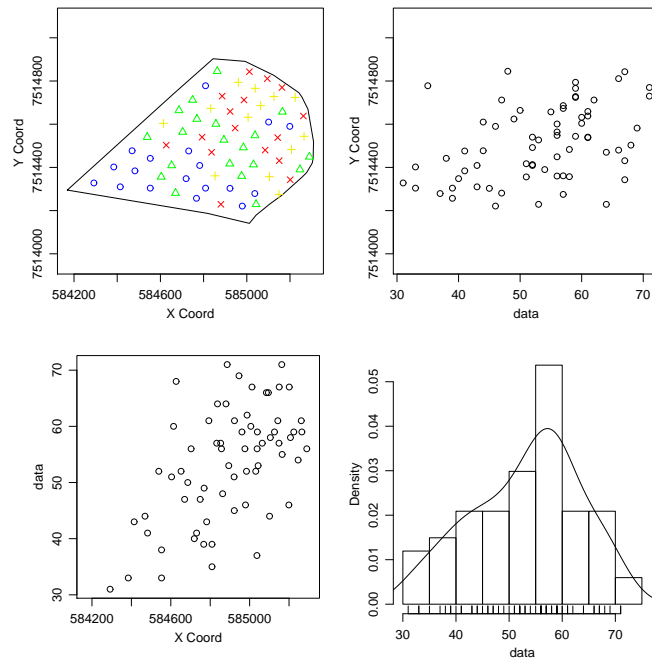


Figura 3 - Gráficos descritivos do padrão espacial da saturação por bases

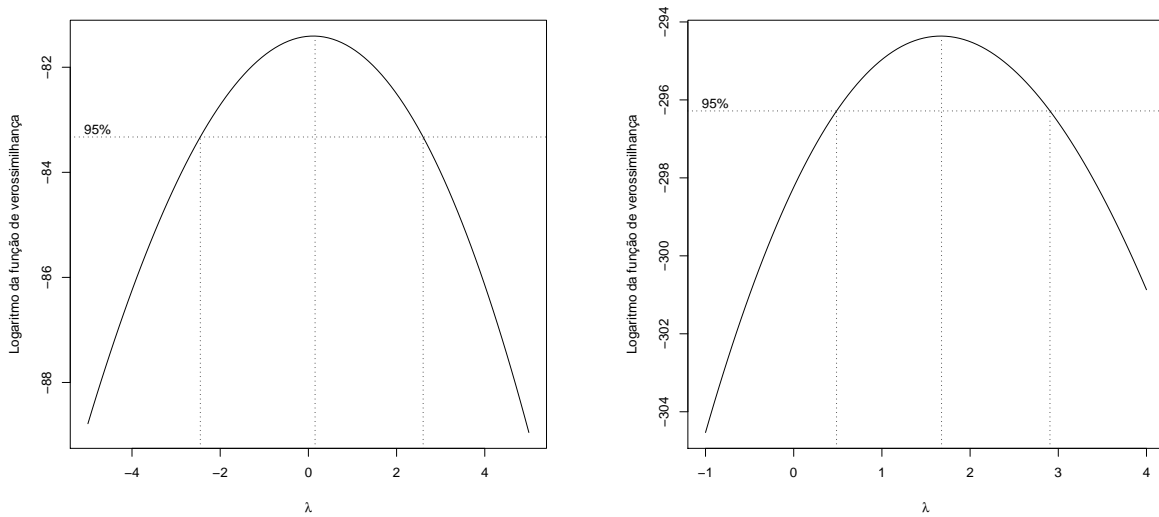


Figura 4 - Gráficos de possíveis transformações de variáveis de Box-Cox, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

4.2.2 Modelagem univariada do pH

As Tabelas 40, 41 e 42 mostram as estimativas por máxima verossimilhança para os parâmetros dos modelos univariados ajustados para o pH considerando as três possíveis formas para a matriz de delineamento X : média constantes, média com tendência induzida pela área de manejo e média induzida pela coordenada oeste-leste. Utilizando as diferentes abordagens para a média, as diferentes escolhas para κ não produzem muita diferença nas estimativas dos parâmetros de média e de variabilidade, no entanto o parâmetro ϕ é mais afetado, o que é normal, uma vez que, os parâmetros provenientes da função de correlação Mátern não são ortogonais. No entanto os alcances práticos das correlações são parecidos para essas estimativas. Com a inclusão de covariáveis no modelo muda um pouco a relação entre σ^2 e τ^2 , onde os dois valores estimados ficaram bem próximos e a estimativa ϕ indica que o alcance das correlações é menor, ou seja, as covariáveis estão explicando, aparentemente, uma parte significativa da variabilidade dos dados. O próximo passo é selecionar o melhor modelo geoestatístico univariado para o pH. Cabe ressaltar que os valores estimados dos máximos dos logaritmos das funções de verossimilhança concentrada não são diretamente comparáveis, pois existem números diferentes de parâmetros associados aos modelos conforme a escolha de X .

Analisando os valores estimados dos máximos do logaritmo da função de verossimilhança e os seus respectivos AIC, foi selecionado o modelo final, que é o com tendência induzida pela coordenada oeste-leste e $\kappa = 2,5$, uma vez que, esse ajuste gerou o maior máximo da função alvo e o menor valor do critério de penalidade por número de parâmetros associado ao modelo.

Além de selecionar o modelo final dentre os que consideram o padrão espacial, foi conduzida uma regressão linear simples considerando a coordenada oeste-leste como covariável, ou seja, o padrão dos dados é completamente aleatório. Sobre esse enfoque o AIC ficou igual a 49,8, o que afirma a idéia de que o pH possui variabilidade espacial significativa.

Tabela 40 - Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com média constantes

β	τ^2	σ^2	ϕ	κ	$l(\theta)$	AIC
4,907	0,077	0,163	560	0,5	-25,49	58,98
4,905	0,101	0,149	510,6	1	-25,69	59,38
4,904	0,106	0,251	650,5	1,5	-25,57	59,15
4,903	0,106	0,288	569,7	2	-25,45	58,90
4,903	0,106	0,297	491,1	2,5	-25,36	58,73

Tabela 41 - Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com tendência na média induzida pela área de manejo

β_0	β_1	τ^2	σ^2	ϕ	κ	$l(\theta)$	AIC
4,725	0,397	0	0,125	49,19	0,5	-23,44	56,89
4,723	0,402	0	0,125	33,42	1	-23,35	56,70
4,723	0,403	0,014	0,111	28,43	1,5	-23,33	56,66
4,779	0,227	0,110	0,061	349,3	2	-24,86	59,73
4,786	0,204	0,109	0,086	353,1	2,5	-24,83	59,65

Tabela 42 - Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com tendência na média induzida pelas coordenadas oeste-leste

β_0	β_1	τ^2	σ^2	ϕ	κ	$l(\theta)$	AIC
-607,48	0,001	0	0,111	49,32	0,5	-19,66	49,31
-608,51	0,001	0	0,112	34,28	1	-19,44	48,89
-608,33	0,001	0	0,112	27,58	1,5	-19,39	48,77
-608,36	0,001	0,008	0,104	24,53	2	-19,36	48,73
-608,46	0,001	0,015	0,097	22,46	2,5	-19,35	48,70

4.2.3 Modelagem univariada da saturação por bases

Para essa variável foram consideradas as mesmas análises conduzidas na seção anterior. As Tabelas 43, 44 e 45 mostram os resultados. Novamente o melhor modelo foi o que considera tendência na média induzida pela coordenada oeste-leste e $\kappa = 2,5$. O modelo de regressão linear simples com a coordenada oeste-leste como covariável gerou um AIC de 480,2, sendo assim, para essa variável também existe padrão espacial de variabilidade significativo.

Tabela 43 - Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com média constantes

β	τ^2	σ^2	ϕ	κ	$l(\theta)$	AIC
49,53	47,66	105,33	700,00	0,5	-239,53	487,05
48,53	59,57	120,62	625,58	1	-239,76	487,51
47,98	62,40	124,31	516,78	1,5	-239,79	487,57
47,27	63,29	149,69	489,58	2	-239,75	487,50
47,18	63,51	139,18	400,00	2,5	-239,70	487,40

Tabela 44 - Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com tendência na média induzida pela área de manejo

β_0	β_1	τ^2	σ^2	ϕ	κ	$l(\theta)$	AIC
47,91	8,80	24,35	53,68	73,68	0,5	-238,05	486,10
47,92	8,83	38,78	39,28	59,18	1	-238,01	486,01
47,92	8,86	43,35	34,69	50,33	1,5	-237,98	485,96
47,92	8,88	45,53	32,47	44,34	2	-237,96	485,92
47,93	8,89	46,79	31,18	39,98	2,5	-237,95	485,90

Tabela 45 - Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com tendência na média induzida pelas coordenadas oeste-leste

β_0	β_1	τ^2	σ^2	ϕ	κ	$l(\theta)$	AIC
-14297,01	0,025	8,98	59,98	54,51	0,5	-234,75	479,50
-14300,66	0,025	27,56	41,46	45,42	1	-234,72	479,44
-14300,92	0,025	33,53	35,53	39,51	1,5	-234,70	479,39
-14300,17	0,025	36,39	32,68	35,33	2	-234,68	479,36
-14298,98	0,025	38,06	31,03	32,19	2,5	-234,67	479,34

4.2.4 Krigagens

Com os modelos finais estabelecidos, foram calculadas as previsões espaciais para as duas variáveis químicas do solo em uma malha de 10000 localizações espaciais, sendo que, as estimativas paramétricas foram substituídas nas fórmulas de krigagem e variância preditiva. A Figura 5 mostra os mapas preditivos dos campos aleatórios. As imagens ficaram condizentes com os dados observados, ou seja, localizações com menores valores para a coordenada oeste-leste possuem, em sua maioria, valores preditos menores. Além disso, as variâncias preditivas da saturação por bases estão todas no intervalo [12,35;30,21], já para o pH os valores máximo e mínimo das variâncias de krigagens ficaram iguais a 0,10 e 0,01, respectivamente. Esses valores das variâncias preditivas são relativamente pequenos com relação a grandeza de escala das variáveis, sendo assim, a precisão das krigagens ficaram em um patamar aceitável.

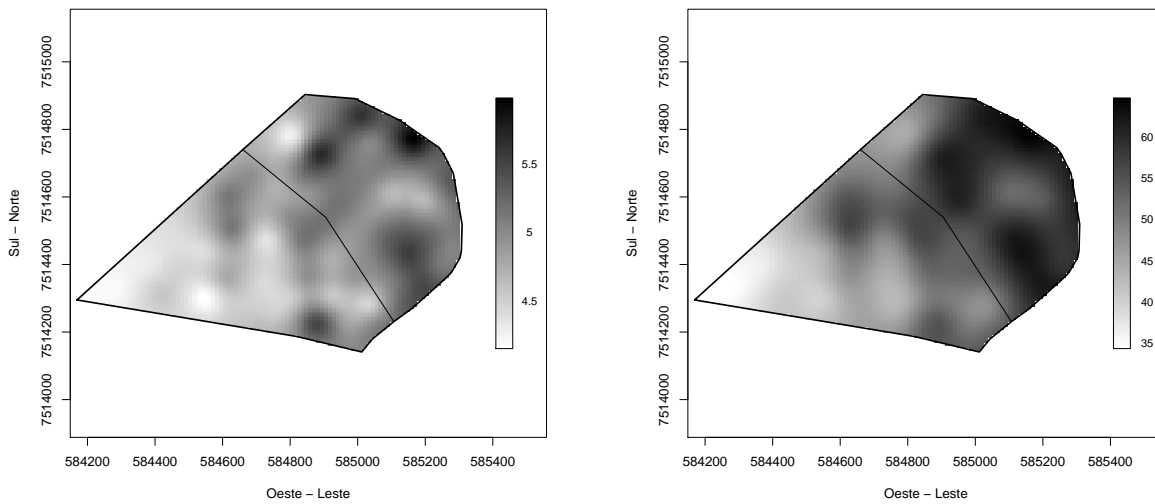


Figura 5 - Gráficos de predições espaciais, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

4.2.5 Análise de resíduos

A Figura 6 mostra que o comportamento dos resíduos do modelo ajustado para o pH aparentemente foge um pouco da suposição de gaussianidade, no entanto, o teste de normalidade de Shapiro-Wilk resultou em um p-valor de 0,22, ou seja, não existe evidências suficientes contra a suposição feita.

A Figura 7 mostra que para a saturação por bases os resíduos do modelos final estão bem próximos da suposição imposta, característica corroborada pelo teste de normalidade de Shapiro-Wilk, que gerou um p-valor de 0,97.

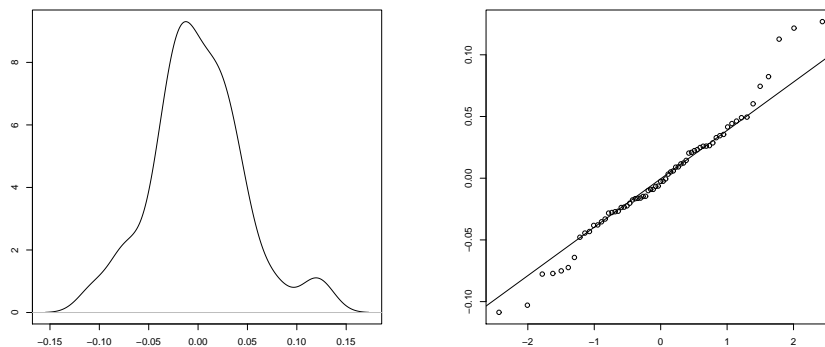


Figura 6 - Densidade e gráfico de quartis dos resíduos para o modelo final do pH

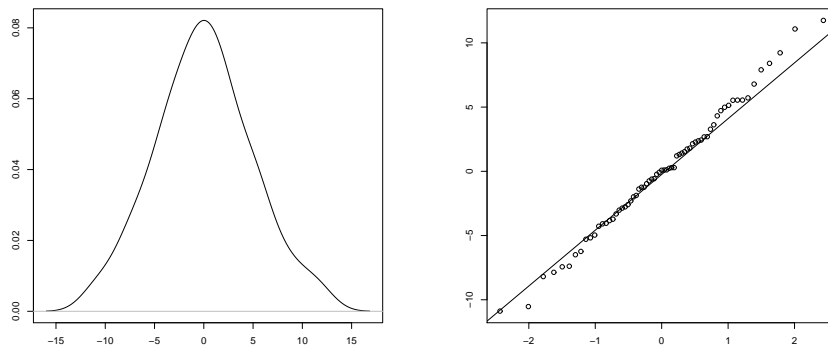


Figura 7 - Densidade e gráfico de quartis dos resíduos para o modelo final da saturação por bases

4.2.6 Modelagem bivariada para o pH e a saturação por bases

Com a modelagem univariada concluída, o próximo passo foi ajustar modelos bivariados para as respostas, em caráter descritivo foi calculada a correlação de Pearson entre os atributos, a estatística resultante ficou em 0,92, o que indica uma correlação positiva e quase perfeita entre as duas variáveis químicas em questão.

Para ajustar o BGCCM e o BCRM, foram utilizadas as estimativas dos modelos univariados para atribuir valores iniciais ao método numérico de maximização. Três abordagens para modelar as médias dos campos aleatórios foram consideradas: médias constantes, médias com tendências induzidas pelas coordenadas oeste-leste e médias com tendências induzidas pela área de manejo do solo. As quais combinadas com diversas escolhas para os parâmetros de suavidade das correlações geraram diversas estimativas diferentes para os parâmetros dos modelos. Inicialmente, foram selecionadas as melhores estimativas em cada abordagem para a média. A Tabela 46 mostra os resultados do BGCCM, onde não há muita diferença entre os parâmetros de variabilidade e correlação estimados para cada abordagem de tendência, a não ser pelo parâmetros ϕ_0 , que para a tendência na área de manejo ficou bem menor do que as demais, fato explicado pela adoção de um κ_0 maior, logo, combinando essas duas informações, tem-se que os alcances de correlações são bem parecidos para as três abordagens de média.

A Tabela 47 mostra os resultados da estimação por máxima verossimilhança utilizando o BCRM. Com esse modelo, a inclusão da covariável coordenada leste-oeste gerou inflacionamento dos parâmetros de variabilidade. Com relação as demais estimativas, os resultados são parecidos com os do BGCCM.

As duas abordagens de modelos bivariados se mostraram estáveis para fazer estimação dos parâmetros, uma vez que, as estimativas dos parâmetros de variabilidade e de correlação apresentaram resultados que geram características similares às encontradas nos modelos univariados.

Nas duas abordagens, o melhor ajuste foi o que considera a área de manejo da propriedade como covariável, os quais geraram AIC iguais a 437,35 e 432,96 para o BGCCM e BCRM, respectivamente. Com as estimativas finais estabelecidas é possível pensar na krigagem ordinária, para tal, foi utilizado o mesmo grid de pontos da krigagem da modelagem univariada.

O processo de krigagem utilizando as estimativas do BGCCM gerou variâncias preditivas relativamente pequenas, para a saturação por bases os valores mínimos e máximos são 0,47

e 58,40, respectivamente, já para o pH esses valores ficaram iguais a 0,0001 e 0,096. Utilizando o BCRM as variâncias preditivas da saturação por bases e do pH ficaram dentro dos intervalos (2;60,35) e (0,002;0,098), respectivamente. As Figuras 8 e 9 ilustram os mapas preditivos utilizando os valores estimados em cada abordagem de modelos bivariados. Não existe muita diferença entre as predições espaciais utilizando modelos univariados ou bivariados, mas o BGCCM gerou menores variâncias de krigagem.

Com essas análises ainda não é possível comparar diretamente os resultados dos modelos geoestatísticos univariados e bivariados. A próxima seção se dedica a essa análise.

Tabela 46 - Estimativas de máxima verossimilhança dos parâmetros associados ao BGCCM para a saturação por bases e o pH

θ	Constantes	Oeste-leste	Área de manejo
β_{01}	52,57	$1,54e - 6$	47,83
β_{02}	4,92	$1,45e - 7$	4,72
β_1	—	$8,98e - 5$	9,46
β_2	—	$8,41e - 6$	0,40
σ_{01}	8,51	8,35	6,85
σ_1	3,47	3,46	3,44
σ_{02}	0,38	0,38	0,31
σ_2	$7e - 5$	$5e - 4$	$1e - 5$
ϕ_0	59,37	57,79	26,73
ϕ_1	46,35	48,15	49,08
ϕ_2	66,82	71,59	86,95
κ_0	1	1	1,5
κ_1	0,5	0,5	0,5
κ_2	0,5	0,5	0,5

Tabela 47 - Estimativas de máxima verossimilhança dos parâmetros associados ao BCRM para a saturação por bases e o pH

θ	Constantes	Oeste-leste	Área de Manejo
β_{01}	51,76	$1,51e - 6$	47,93
β_{02}	4,89	$1,43e - 7$	4,72
β_1	—	$8,85e - 5$	9,12
β_2	—	$8,36e - 6$	0,40
σ_{11}	8,26	99,53	7,82
σ_{12}	0,31	3,67	0,28
σ_{22}	0,13	1,54	0,14
ϕ_1	100,63	100,21	53,32
ϕ_2	21,43	21,38	19,34
κ_1	0,5	0,5	0,5
κ_2	2,5	2,5	2,5

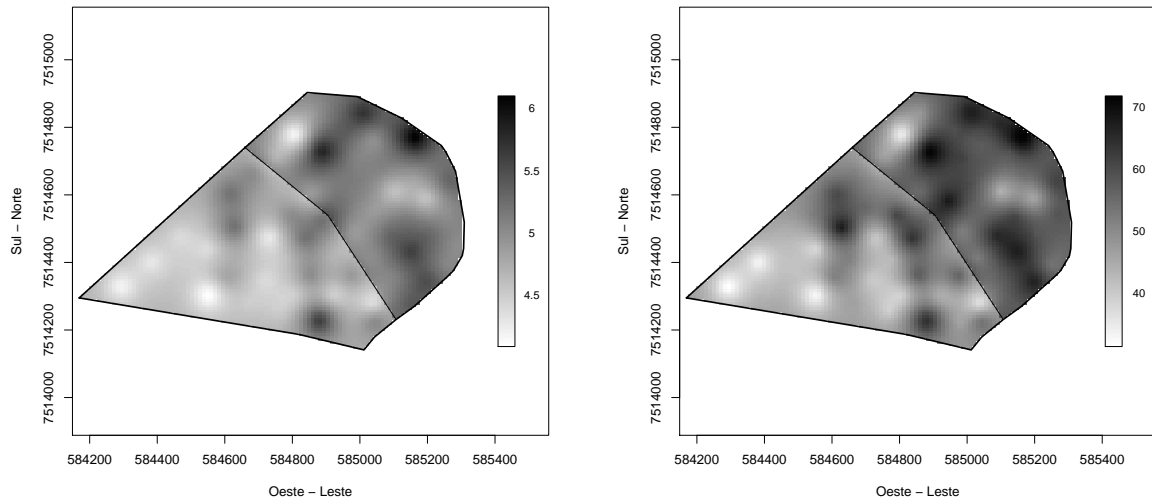


Figura 8 - Gráficos de predições espaciais com o BGCCM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

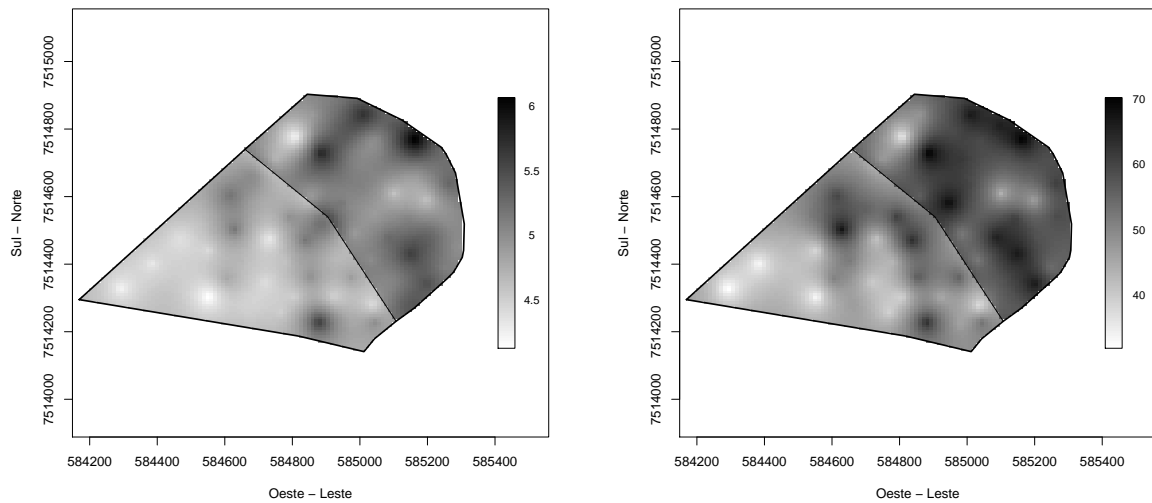


Figura 9 - Gráficos de predições espaciais com o BCRM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

4.2.7 Estudo empírico de comparação entre as abordagens univariada e bivariadas

Além da fundamentação estatística para considerar modelos bivariados para o problema dos dados do solo, existe a justificativa prática para adoção de tal abordagem. A coleta de informação da saturação por bases é mais dispendiosa e como a propriedade agrícola continuará sendo monitorada futuramente, é possível que nas próximas avaliações do solo seja amostrado um número menor de localizações espaciais para observação da saturação por bases, pois pode-se utilizar a informação contida no pH para fazer inferências sobre a saturação por bases. Ou seja, é possível diminuir gastos em monitoramentos de solos se forem utilizados os modelos geoestatísticos bivariados para problemas com dados de solos, onde existem variáveis químicas correlacionadas.

Para ilustrar essa técnica, foram omitidas do conjunto de dados 20 localizações espaciais da saturação por bases. Foram estimados novamente, em todas as abordagens, os parâmetros dos modelos finais selecionados. Ocorreu uma diferença moderada entre os resultados das estimativas calculadas com a omissão de localizações e os das calculadas com a informação completa da saturação por bases. No entanto, o objetivo dessa seção é avaliar a qualidade preditiva dos modelos univariados e bivariados.

Para verificar o ganho com a utilização de modelos bivariados, foi conduzida a krigagem ordinária da saturação por bases nas 20 localizações retiradas do conjunto de dados sob os três modelos reajustados. A Tabela 48 mostra os resultados de predição.

Tabela 48 - Valores observados da saturação por bases e não utilizados na modelagem e suas respectivas predições utilizando os modelo univariado, o BGCCM e o BCRM

Observado	Univariado	BGCCM	BCRM
31	37,97514	36,70374	36,41034
52	50,31664	48,47015	48,75127
41	38,73410	43,28838	42,69730
51	42,32669	50,60234	50,66902
41	56,75724	39,95526	40,90431
43	53,03821	46,20176	46,51603
39	44,36711	43,17711	42,64546
51	56,26729	48,47547	48,79987
64	46,88141	68,33242	68,03238
53	47,06461	53,68137	52,71941
47	46,36761	47,39061	46,98018
67	58,88574	70,07791	69,85622
57	54,87683	59,06358	58,93524
56	57,61760	52,66449	52,93289
59	60,95767	54,67692	55,20092
57	56,82567	56,92622	57,16953
52	57,78833	52,64306	52,94522
61	56,38470	61,31782	61,35151
59	56,07022	54,45870	54,30496
54	62,32295	56,91067	57,15702

A Tabela 49 apresenta os resultados descritivos das diferenças entre os valores observados e os valores preditos nas três abordagens de modelagem. Os modelos bivariados se

mostraram mais eficientes para fazer a krigagem, uma vez que, os desvios padrões dos erros de krigagem são menores do que o valor gerado pela abordagem univariada, ou seja, há evidências de que é melhor utilizar a informação do pH para fazer as inferências sobre a saturação por bases com menos observações.

Tabela 49 - Médias e desvios padrões dos erros de krigagem da saturação por bases nas 20 localizações omitidas no processo de estimação

Estatística	Univariado	BGCCM	BCRM
Média	-0,34	-0,50	-0,50
Desvio Padrão	7,52	3,02	2,84

5 CONSIDERAÇÕES FINAIS

Esse trabalho teve os objetivos de avaliar o comportamento dos modelos geoestatísticos bivariado em conjunto com estimação por máxima verossimilhança e krigagem ordinária, comparar as abordagens de modelos geoestatísticos bivariados, verificar a aplicabilidade dos modelos bivariados com dados observacionais e suas vantagens com relação aos modelos univariados.

O estudo de simulação revelou que para as três configurações amostrais de localização as abordagens de Gelfand et al. (2004) e Diggle e Ribeiro Jr. (2006) estão gerando estimativas precisas para os parâmetros, exceto pelas subestimações moderadas dos parâmetros de variabilidade e correlação. Com relação as predições espaciais, as abordagens se mostram bastante eficientes, pois geraram erros preditivos pequenos. De uma forma geral, os dois modelos bivariados considerados nesse estudo possuem comportamentos parecidos para estimação e krigagem e não há vantagens evidentes com a adoção de algumas das abordagens.

Devido a alta complexidade paramétrica dos modelos geoestatísticos bivariados, existe uma leve instabilidade computacional no processo de estimação por métodos numéricos, problema que pode ser contornado impondo algumas restrições sobre a função alvo do processo de maximização numérica. Além disso, é demandado muito tempo para fazer estimação dos parâmetros do BGCCM e do BCRM.

A análise dos dados observacionais do solo mostrou que os modelos geoestatísticos bivariados são uma ferramenta útil para a agricultura de precisão, pois sua utilização pode baratear os custos dos monitoramentos de solos e ao mesmo tempo gerar mapas preditivos com alta qualidade.

Em trabalhos futuros podem ser investigadas outras configurações paramétricas no estudo de simulação, assim verificando se os modelos geoestatísticos bivariados se comportam bem com diferentes tipos de variabilidade e de padrões de correlação. Outra área que pode ser estudada é a utilização de técnicas bayesianas em conjunto com estimação via MCMC para o BGCCM, assim como a utilização de diferentes reparametrizações para o BCRM.

REFERÊNCIAS

- AZZALINI, A. **Statistical inference: based on the likelihood**, London; New York: Chapman and Hall, 1996. 341 p.
- BANERJEE, S.; GELFAND, A.E. Predict, Interpolation and regression for spatial misaligned data points. **Sankhya**, v.64, p.227-245, 2002.
- BICKEL, P.J.; DOKSUM K.A. **Mathematical Statistics: basic ideas and selected topics**, San Francisco: Holden-Day, 1977. 492 p.
- BROWN, P.J.; LE, N.D.; ZIDEK, J.V. Multivariate spatial interpolation and exposure to air pollutants. **The Canadian Journal of Statistics**, v.22, p.489-509, 1994.
- CRESSIE N.; HUANG, H-C. Classes of Non-Separable, Spatio-temporal stationary covariance functions. **Journal of the American Statistical Association**, Alexandria, v.94, p.1330-1340, 1999.
- CHRISTENSEN, O.F.; DIGGLE, P.J.; RIBEIRO Jr., P.J., Analysing positive-valued spatial data: the transformed gaussian model, **Geostatistics for environmental applications**, Boston: v.11, p.287-298, 2001.
- DIGGLE, P.J.; RIBEIRO Jr., P.J. **Model-Based geostatistics**. New York: Springer, 2006. 230p.
- ELMATZOGLOU, I. **Spatio-temporal geostatistical models, with an application in fish stock**, 2006. 53 p. Submitted for the degree of (Master in statistics) - Lancaster University, Lancaster, 2006.
- FERNANDES, M.V.M. **Modelos para Processos Espaço-Temporais Inflacionados de Zeros**, 2006. 128 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática da Univeridade Federal do Rio de Janeiro, 2006.
- FUENTES, M.; SMITH, R.L. **A new class of stationary spatial models**. North Caroline: Department of Statistics, North Caroline State University, 2001. Technical (Report, 2534).
- GELFAND, A.E; SCHMIDT, A.M.; BANERJEE S.; SIRMANS, C.F. Nonstationary multivariate process modeling through spatially varying coregionalization. **Sociedad Española de Estadística e Investigación Operativa - Test**, Valencia: v.13, p.263-312, 2004.
- GNEITING, T. Nonseparable, Stationary Covariance Functions for Space-Time Data. **Journal of the American Statistical Association**, Alexandria, v.97, p.590-600, 2002.
- HIGDON, D. **Quantitative methods for current environmental issues**, Chichester: Wiley, 2002. 185 p.
- LE, D.N.; ZIDEK, J.V. **Statistical analysis of environmental space-time processes**. New York: Springer, 2006. 327p.

- MARDIA, K.V.; GOODALL, C.R. Spatial-temporal analysis of multivariate environmental monitoring data. In: G. P. PATIL and C. R. Rao, eds., **Multivariate Environmental Statistics**, p.347-386, 1993.
- MATÉRN, B. **Spatial variation**. Verlag, Berlin: Spinger, 1986. 365 p.
- NELDER, J.A.; MEAD R. **A simplex method for function minimization**. *Comput. J.* 7: p.308-313, 1965. Spinger, 1986. 365 p.
- SAMPSON P.D.; GUTTORP, P. Nonparametric estimation of nonstationary spatial covariance structure. **Journal of American Statistical Association**, Alexandria, v.87, p.108-119, 1992.
- SCHABENBERGER, O.; GOTWAY, C.A. **Statistical methods for spatial data analysis**, Boca Raton: Chapman and Hall / CRC, 2005. 488p.
- SCHIMIDT, A.M.; SANSÓ, B. Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço-temporais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14, 2006. Caxambú, **Minicurso...** São Paulo: Associação Brasileira de Estatística, 2006. 151 p.
- SCHMIDT, A.M.; O'HAGAN, A. Bayesian inference for nonstationary spatial covariance structure via spatial deformations, **Journal of Royal Statistical Society**, Oxford: v.65, p.743-758, 2003.
- SCHULLER, J.K. A review and integrating analysis of spatially-variable crop control of crop production. **Fertilizer Research, the Haugue**, v.33, p.1-34, 1992.
- SILVA, A.S. **Modelos gaussianos geoestatísticos espaço-temporais e aplicações**, 2006. 70 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz" da Univeridade de São Paulo, 2006.
- SUN, W.; LE, N.D.; ZIDEK, J.V.; BURNETT, R. Assessment of a bayesian multivariate interpolation approach for health impact studies. **Environmetrics**, Washington, v.9, p.565-586, 1998.
- R: A Language and Environmental for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2006, ISBN 3-900051-07-0, <http://www.R-project.org>.
- geoR: A package for geostatistical analysis. RIBEIRO Jr, P.J. and DIGGLE, P.J., R-NEWS, CRAN.R-project.org/doc/Rnews/.v.1, n.2, p.14-18, 2001.

APÊNDICE

```
#####
## Exemplo de programa utilizado análise dos dados sobre qualidade do solo#
#####
require(geoR)
require(MASS)
require(ASOR)
Attach()

### Análise Descritiva inicial
summary(v67.g$data)
sd(v67.g$data)
boxplot(v67.g$data)
points(v67.g)
plot(v67.g)
require(MASS)
boxcox(v67.g)

### Estimação AD-HOC para os chutes iniciais com tend cte
a<-variog(v67.g)
plot(a)
eyefit(a)

### Estimação de M.V. com tendência constante
estimado1_1<-likfit(v67.g,ini.cov.pars=c(326.29,700),nugget=35,
fix.kappa=TRUE,kappa=0.5)
estimado1_2<-likfit(v67.g,ini.cov.pars=c(326.29,625.58),nugget=42.1,
fix.kappa=TRUE,kappa=1)
estimado1_3<-likfit(v67.g,ini.cov.pars=c(336.81,516.78),nugget=52.62,
fix.kappa=TRUE,kappa=1.5)
estimado1_4<-likfit(v67.g,ini.cov.pars=c(368.39,489.58),nugget=63.15,
fix.kappa=TRUE,kappa=2)
estimado1_5<-likfit(v67.g,ini.cov.pars=c(350,400),nugget=63.15,
fix.kappa=TRUE,kappa=2.5)

### Estimação de M.V. com tendência area
estimado2_1<-likfit(v67.g,trend=~AREA,ini.cov.pars=c(65,300),nugget=45,
fix.kappa=TRUE,kappa=0.5)
estimado2_2<-likfit(v67.g,trend=~AREA,ini.cov.pars=c(68.12,308.79)
,nugget=49.96,fix.kappa=TRUE,kappa=1)
estimado2_3<-likfit(v67.g,trend=~AREA,ini.cov.pars=c(72.66,353.59),
nugget=54.50,fix.kappa=TRUE,kappa=1.5)
estimado2_4<-likfit(v67.g,trend=~AREA,ini.cov.pars=c(63.58,190.39),
nugget=45.41,fix.kappa=TRUE,kappa=2)
estimado2_5<-likfit(v67.g,trend=~AREA,ini.cov.pars=c(65,190.39),
nugget=45.41,fix.kappa=TRUE,kappa=2.5)
```

```
### Estimação de M.V. com tendência nas coordenadas oeste-leste
estimado3_1<-likfit(v67.g,trend=~coords[,1],ini.cov.pars=c(65,300),
nugget=45,fix.kappa=TRUE,kappa=0.5)
estimado3_2<-likfit(v67.g,trend=~coords[,1],ini.cov.pars=c(68.12,308.79),
nugget=49.96,fix.kappa=TRUE,kappa=1)
estimado3_3<-likfit(v67.g,trend=~coords[,1],ini.cov.pars=c(72.66,353.59),
nugget=54.50,fix.kappa=TRUE,kappa=1.5)
estimado3_4<-likfit(v67.g,trend=~coords[,1],ini.cov.pars=c(63.58,190.39),
nugget=45.41,fix.kappa=TRUE,kappa=2)
estimado3_5<-likfit(v67.g,trend=~coords[,1],ini.cov.pars=c(65,190.39),
nugget=45.41,fix.kappa=TRUE,kappa=2.5)
```

```
### Seleção de modelo
```

```
estimado1_1$log
estimado1_2$log
estimado1_3$log
estimado1_4$log
estimado1_5$log
estimado2_1$log
estimado2_2$log
estimado2_3$log
estimado2_4$log
estimado2_5$log
estimado3_1$log
estimado3_2$log
estimado3_3$log
estimado3_4$log
estimado3_5$log
```

```
estimado1_1$AIC
estimado1_2$AIC
estimado1_3$AIC
estimado1_4$AIC
estimado1_5$AIC
estimado2_1$AIC
estimado2_2$AIC
estimado2_3$AIC
estimado2_4$AIC
estimado2_5$AIC
estimado3_1$AIC
estimado3_2$AIC
estimado3_3$AIC
estimado3_4$AIC
estimado3_5$AIC
```

```
### Com o modelo definido, iniciar a krigagem
```

```

### Estabelecendo os extremos para a o Grid de predição
minimo<-min(v67.g$borders[,1])
minimo
maximo<-max(v67.g$borders[,1])
minimoy<-min(v67.g$borders[,2])
maximoy<-max(v67.g$borders[,2])

### Criando Grid
gr<-expand.grid(seq(minimo,maximo,l=100),seq(minimoy,maximoy,l=100))

### Predição
pred1<-krige.conv(v67.g,loc=gr,krige=krige.control(obj=estimado3_5))
par(mfrow=c(1,2),pty="s")
points(v67.g,)
image(pred1,col=terrain.colors(100))
contour(pred1,add=T)
image(pred1,val=pred1$krige.var)
contour(pred1,val=pred1$krige.var,add=T)

### Analisando o ruído branco
estimado3_5$model.components[,3]
plot(density(estimado3_5$model.components[,3]))
qqnorm(estimado3_5$model.components[,3])
qqline(estimado3_5$model.components[,3])
shapiro.test(estimado3_5$model.components[,3])

#####
## Programa para simular dados com os modelos geoestísticos bivariados#
#####

#Simulação com o BGCCM
require(MASS)
require(geoR)
rBGCCM_sim <- function(n1,n2,N,n_arm,perc_co_loc=1,medias=c(0,0),sigmas,phis,
kappas=c(.5,.5,.5),sem_grid=1,sem_cam_ale=1){
set.seed((sem_grid*3331))
grx1 <- runif(n1)
set.seed((sem_grid*3771))
gry1 <- runif(n1)
gr1 <- unname(cbind(grx1[1:n1],gry1[1:n1]))
set.seed((sem_grid*3713))
grx2 <- runif(n2)
set.seed((sem_grid*3317))
gry2 <- runif(n2)
gr2 <- unname(cbind(grx2,gry2))
set.seed((sem_grid*3137))
grx3 <- runif(n_arm)

```

```

set.seed((sem_grid*3337))
gry3 <- runif(n_arm)
gr3 <- unname(cbind(grx3,gry3))
n <- n1+n2
n_co <- round(min(n1,n2)*perc_co_loc,0)
k1 <- c((n1+n_co+1):n)
k2 <- c(1:(n2-n_co))
k1_1 <- c((n_co+1):n1)
k2_1 <- c(1:(n1-n_co))
k3 <- c(1:n)
k4 <- c(1:n1)
k4_1 <- c(1:(n1+n_arm))
k5 <- c((n1+1):n)
k5_1 <- c((n1+n_arm+1):(n2*n_arm))
k5_2 <- c((n1+n_arm+1):(n1+n_arm+n2))
k6 <- c(1:n2)
k6_1 <- c(1:(n2+n_arm))
k7 <- c((n1+n_arm+n_co+1):(n+n_arm))
k8 <- c(1:(n2*n_arm))
if(n1==n2) gr <- rbind(gr1,gr3,gr1,gr3)
if(n1==n2 & perc_co_loc < 1) gr[k7,] <- gr2[k2,]
if(n1>n2) gr <- rbind(gr1,gr3,gr1[k6,],gr3)
if(n1>n2 & perc_co_loc < 1) gr[k7,] <- gr2[k2,]
if(n2>n1) gr <- rbind(gr2[k4,],gr3,gr2,gr3)
if(n2>n1 & perc_co_loc < 1) gr[k1_1,] <- gr1[k2_1,]
m_1 <- medias[1]
m_2 <- medias[2]
s01 <- sigmas[1]
s02 <- sigmas[3]
s1 <- sigmas[2]
s2 <- sigmas[4]
phi0 <- phis[1]
phi1 <- phis[2]
phi2 <- phis[3]
kappa0 <- kappas[1]
kappa1 <- kappas[2]
kappa2 <- kappas[3]
set.seed(sem_cam_ale*31)
R01 <- grf(n=(n1+n_arm),grid=gr[k4_1,],nsim=N,
cov.pars=c(s01^2,phi0),kappa=kappa0)
set.seed(sem_cam_ale*31)
R02 <- grf(n=(n2+n_arm),grid=gr[k5_1,],nsim=N,
cov.pars=c(s02^2,phi0),kappa=kappa0)
set.seed(sem_cam_ale*37)
R1 <- grf(n=(n1+n_arm),grid=gr[k4_1,],nsim=N,
cov.pars=c(s1^2,phi1),kappa=kappa1)
set.seed(sem_cam_ale*71)

```

```

R2 <- grf(n=(n2+n_arm),grid=gr[k5_1,],nsim=N,
cov.pars=c(s2^2,phi2),kappa=kappa2)
y1 <- matrix(0,(n1+n_arm),N)
y2 <- matrix(0,(n2+n_arm),N)
if(N >1){
  y1 <- m_1 + R01$data + R1$data
  y2 <- m_2 + R02$data + R2$data
}
if(N == 1){
  y1[,1] <- m_1 + R01$data + R1$data
  y2[,1] <- m_2 + R02$data + R2$data
}
saida <- list()
geo1 <- cbind(gr[k4,],y1[k4,])
geo2 <- cbind(gr[k5_2,],y2[k6,])
geo_arm1 <- cbind(gr3,y1[(n1+1):(n1+n_arm),])
geo_arm2 <- cbind(gr3,y2[(n2+1):(n2+n_arm),])
saida$geo_arm1 <- as.geodata(geo_arm1,data.col=c(3:(2+N)))
saida$geo_arm2 <- as.geodata(geo_arm2,data.col=c(3:(2+N)))
saida$geo1 <- as.geodata(geo1,data.col=c(3:(2+N)))
saida$geo2 <- as.geodata(geo2,data.col=c(3:(2+N)))
saida$geo1$coords <- unname(saida$geo1$coords)
saida$geo2$coords <- unname(saida$geo2$coords)
saida$geo_arm1$coords <- unname(saida$geo_arm1$coords)
saida$geo_arm2$coords <- unname(saida$geo_arm2$coords)
return(saida)
}
sim1 <- rBGCCM_sim(n1=100,n2=100,N=1100,n_arm=20,
perc_co_loc=1,medias=c(150,60),sigmas=c(8,4,5,2),phis=c(0.25,0.20,0.20),
kappas=c(0.5,0.5,0.5),sem_cam_ale=1)

#Simulação o BCRM
require(MASS)
require(geoR)
rBCRM_sim <- function(n1,n2,N,n_arm,perc_co_loc=1,medias=c(0,0),sigmas,phis,
kappas=c(.5,.5),sem_grid=1,sem_cam_ale=1){
set.seed((sem_grid*3331))
grx1 <- runif(n1)
set.seed((sem_grid*3771))
gry1 <- runif(n1)
gr1 <- unname(cbind(grx1[1:n1],gry1[1:n1]))
set.seed((sem_grid*3713))
grx2 <- runif(n2)
set.seed((sem_grid*3317))
gry2 <- runif(n2)
gr2 <- unname(cbind(grx2,gry2))
set.seed((sem_grid*3137))

```

```

grx3 <- runif(n_arm)
set.seed((sem_grid*3337))
gry3 <- runif(n_arm)
gr3 <- unname(cbind(grx3,gry3))
n <- n1+n2
n_co <- round(min(n1,n2)*perc_co_loc,0)
k1 <- c((n1+n_co+1):n)
k2 <- c(1:(n2-n_co))
k1_1 <- c((n_co+1):n1)
k2_1 <- c(1:(n1-n_co))
k3 <- c(1:n)
k4 <- c(1:n1)
k4_1 <- c(1:(n1+n_arm))
k5 <- c((n1+1):n)
k5_1 <- c((n1+n_arm+1):(n2*n_arm))
k5_2 <- c((n1+n_arm+1):(n1+n_arm+n2))
k6 <- c(1:n2)
k6_1 <- c(1:(n2+n_arm))
k7 <- c((n1+n_arm+n_co+1):(n+n_arm))
k8 <- c(1:(n2*n_arm))
if(n1==n2) gr <- rbind(gr1,gr3,gr1,gr3)
if(n1==n2 & perc_co_loc < 1) gr[k7,] <- gr2[k2,]
if(n1>n2) gr <- rbind(gr1,gr3,gr1[k6,],gr3)
if(n1>n2 & perc_co_loc < 1) gr[k7,] <- gr2[k2,]
if(n2>n1) gr <- rbind(gr2[k4,],gr3,gr2,gr3)
if(n2>n1 & perc_co_loc < 1) gr[k1_1,] <- gr1[k2_1,]
m_1 <- medias[1]
m_2 <- medias[2]
s11 <- sigmas[1]
s21 <- sigmas[2]
s22 <- sigmas[3]
phi1 <- phis[1]
phi2 <- phis[2]
kappa1 <- kappas[1]
kappa2 <- kappas[2]
set.seed(sem_cam_ale*31)
R11 <- grf(n=(n1+n_arm),grid=gr[k4_1,],nsim=N,
cov.pars=c(s11^2,phi1),kappa=kappa1)
set.seed(sem_cam_ale*31)
R21 <- grf(n=(n2+n_arm),grid=gr[k5_1,],nsim=N,
cov.pars=c(s21^2,phi1),kappa=kappa1)
set.seed(sem_cam_ale*37)
R22 <- grf(n=(n2+n_arm),grid=gr[k5_1,],nsim=N,
cov.pars=c(s22^2,phi2),kappa=kappa2)
y1 <- matrix(0,(n1+n_arm),N)
y2 <- matrix(0,(n2+n_arm),N)
if(N >1){

```



```

    y1 <- m_1 + R11$data
    y2 <- m_2 + R21$data + R22$data
  }
  if(N == 1){
    y1[,1] <- m_1 + R11$data
    y2[,1] <- m_2 + R21$data + R22$data
  }
  saida <- list()
  geo1 <- cbind(gr[k4,],y1[k4,])
  geo2 <- cbind(gr[k5_2,],y2[k6,])
  geo_arm1 <- cbind(gr3,y1[(n1+1):(n1+n_arm),])
  geo_arm2 <- cbind(gr3,y2[(n2+1):(n2+n_arm),])
  saida$geo_arm1 <- as.geodata(geo_arm1,data.col=c(3:(2+N)))
  saida$geo_arm2 <- as.geodata(geo_arm2,data.col=c(3:(2+N)))
  saida$geo1 <- as.geodata(geo1,data.col=c(3:(2+N)))
  saida$geo2 <- as.geodata(geo2,data.col=c(3:(2+N)))
  saida$geo1$coords <- unname(saida$geo1$coords)
  saida$geo2$coords <- unname(saida$geo2$coords)
  saida$geo_arm1$coords <- unname(saida$geo_arm1$coords)
  saida$geo_arm2$coords <- unname(saida$geo_arm2$coords)
  return(saida)
}

sim1 <- rBCRM_sim(n1=100,n2=100,N=1100,n_arm=20,
perc_co_loc=1,medias=c(150,60),sigmas=c(9,5,2),phis=c(0.25,0.20),
kappas=c(0.5,0.5),sem_cam_ale=3)
require(ASOR)
Store(sim1)

#####
## Programa para estimação dos parâmetros dos modelos bivariados#
#####

#Estimação para o BGCCM
require(MASS)
require(geoR)
emvBGCCM_sim <- function(geo1,geo2,corr_ini,ruido=F,kappas=c(0.5,0.5,0.5)
,metodo="Nelder-Mead",inferior=-Inf,superior=Inf,
controle=list(),hessiana=F){
n1 <- nrow(geo1$coords)
n2 <- nrow(geo2$coords)
n <- n1+n2
N <- ncol(geo1$data)
if(is.null(N)) N <- 1
k1 <- c(1:n1)
k2 <- c((n1+1):n)
k3 <- c(1:n)
k4 <- c(1:n2)

```

```

dados <- matrix(0,n,N)
if(N > 1) for(i in 1:N) dados[,i] <- c(geo1$data[,i],geo2$data[,i])
if(N == 1) dados[,1] <- c(geo1$data,geo2$data)
dists <- unname(as.matrix(dist(rbind(geo1$coords,geo2$coords)
,diag=TRUE,upper=TRUE)))
X <- matrix(c(rep(1,n1),rep(0,n2),rep(0,n1),rep(1,n2)),n,2)
kappa0 <- kappas[1]
kappa1 <- kappas[2]
kappa2 <- kappas[3]
ctes <- list()
ctes$n1 <- n1
ctes$n2 <- n2
ctes$dados <- dados
ctes$dists <- dists
ctes$X <- X
est_num<-function(p,Ct,rui=F,kap,repet){
  if ( any(p < 1e-20)) return(.Machine$double.xmax^.5)
  if ( any(p[4:6] > 1)) return(.Machine$double.xmax^.5)
  if ( any(p > 100)) return(.Machine$double.xmax^.5)
  R0 <- (((2^(kap[1]-1))*gamma(kap[1]))^(-1)) *
  ((Ct$dists/p[4])^(kap[1])) * besselK(Ct$dists/p[4],nu=kap[1])
  R0[ Ct$dists < 0.005 ] <- 1
  R1 <- (((2^(kap[2]-1))*gamma(kap[2]))^(-1)) *
  ((Ct$dists/p[5])^(kap[2])) * besselK(Ct$dists/p[5],nu=kap[2])
  R1[ Ct$dists < 0.005 ] <- 1
  R2 <- (((2^(kap[3]-1))*gamma(kap[3]))^(-1)) *
  ((Ct$dists/p[6])^(kap[3])) * besselK(Ct$dists/p[6],nu=kap[3])
  R2[ Ct$dists < 0.005 ] <- 1
  V1 <- R0[1:Ct$n1,1:Ct$n1] + p[2]^2 * R1[1:Ct$n1,1:Ct$n1]
  V2 <- p[1]^2 * R0[(Ct$n1+1):(Ct$n1+Ct$n2),(Ct$n1+1):(Ct$n1+Ct$n2)] +
  p[3]^2 * R2[(Ct$n1+1):(Ct$n1+Ct$n2),(Ct$n1+1):(Ct$n1+Ct$n2)]
  V3 <- p[1] * R0 [1:Ct$n1,(Ct$n1+1):(Ct$n1+Ct$n2)]
  V <- cbind(rbind(V1,t(V3)),rbind(V3,V2))
  if(rui == TRUE) {diag(V) <- diag(V) + c(rep((p[7]^2),Ct$n1),
  rep((p[8]^2),Ct$n2))}
  Vinv <- ginv(V)
  tVinvX <- crossprod(Vinv,Ct$X)
  Q_hat <- drop(sum(crossprod(Ct$dados[,repet],Vinv)*Ct$dados[,repet])
  - crossprod(Ct$dados[,repet],tVinvX)%*%
  ginv(crossprod(tVinvX,Ct$X))%*%crossprod(tVinvX,Ct$dados[,repet]))
  lQ_hat <- log(Q_hat)
  ldv <- log(det(V))
  if(is.infinite(lQ_hat) | is.nan(lQ_hat) | is.infinite(ldv) |
  is.nan(ldv)) return(.Machine$double.xmax^.5)
  ll <- drop(-0.5*((Ct$n1+Ct$n2)*(log(2*pi)+lQ_hat-log(Ct$n1+Ct$n2)+1)+ldv))
  if(is.infinite(ll) | is.nan(ll)) return(.Machine$double.xmax^.5)
  return(-ll)
}

```

```

}
resultados <- list()
resultados$betas <- matrix(0,N,2)
colnames(resultados$betas) <- c("mu1","mu2")
resultados$sigmas <- matrix(0,N,4)
colnames(resultados$sigmas) <- c("s01","s1","s02","s2")
resultados$phis <- matrix(0,N,3)
colnames(resultados$phis) <- c("phi0","phi1","phi2")
if(ruido == TRUE){
  resultados$taus <- matrix(0,N,2)
  colnames(resultados$taus) <- c("tau1","tau2")
  resultados$re_parametros <- matrix(0,N,5)
  colnames(resultados$re_parametros) <- c("eta","nu1","nu2","psi1","psi2")
  if(hessiana == TRUE){
    resultados$hessiana <- matrix(0,(8*N),8)
    colnames(resultados$hessiana) <- c("eta","nu1","nu2",
    "phi0","phi1","phi2","psi1","psi2")
    rownames(resultados$hessiana) <- rep(c("eta","nu1","nu2",
    "phi0","phi1","phi2","psi1","psi2"),N)
  }
}
if(ruido==FALSE){
  resultados$re_parametros <- matrix(0,N,3)
  colnames(resultados$re_parametros) <- c("eta","nu1","nu2")
  if(hessiana == TRUE){
    resultados$hessiana <- matrix(0,(6*N),6)
    colnames(resultados$hessiana) <- c("eta","nu1","nu2","phi0",
    "phi1","phi2")
    rownames(resultados$hessiana) <- rep(c("eta","nu1","nu2","phi0",
    "phi1","phi2"),N)
  }
}
resultados$log_lik <- matrix(0,N,1)
resultados$kappas <- c(kappa0=kappas[1],kappa1=kappas[2],kappa2=kappas[3])
resultados$erros <- matrix(0,N,7)
colnames(resultados$erros) <- c("repet","eta","nu1","nu2","phi0","phi1","phi2")
convergencia <- rep(0,N)
for(i in 1:N){
  var1_amos <- var(geo1$data[,i])
  var2_amos <- var(geo2$data[,i])
  var_ini <- NULL
  var_ini[1] <- sqrt(0.8*var1_amos)
  var_ini[2] <- sqrt(0.2*var1_amos)
  var_ini[3] <- sqrt(0.8*var2_amos)
  var_ini[4] <- sqrt(0.2*var2_amos)
  repar_ini <- 0
  repar_ini[1] <- var_ini[3]/var_ini[1]

```

```

repar_ini[2] <- var_ini[2]/var_ini[1]
repar_ini[3] <- var_ini[4]/var_ini[1]
repar_ini[4] <- corr_ini[1]
repar_ini[5] <- corr_ini[2]
repar_ini[6] <- corr_ini[3]
if(ruido == TRUE) {
  repar_ini[7] <- var_ini[5]/var_ini[1]
  repar_ini[8] <- var_ini[6]/var_ini[1]
}

estimativas <- try(optim(p=repar_ini,fn=est_num,Ct=ctes,
rui=ruido,kap=c(kappa0,kappa1,kappa2),repet=i,method=metodo,
lower=inferior,upper=superior,control=controle,hessian=hessiana))
if(class(estimativas) == "try-error"){
  resultados$erros[i,] <- c(i,NaN,NaN,NaN,NaN,NaN,NaN)
  resultados$betas[i,] <- c(NaN,NaN)
  resultados$sigmas[i,] <- c(NaN,NaN,NaN,NaN)
  resultados$phis[i,] <- c(NaN,NaN,NaN)
  if(ruido == TRUE){
    resultados$taus[i,] <- c(NaN,NaN)
    resultados$re_param[i,] <- c(NaN,NaN,NaN,NaN,NaN)
  }
  if(ruido==FALSE) resultados$re_parametros[i,] <- c(NaN,NaN,NaN)
  resultados$log_lik[i,] <- NaN
  if(hessiana == TRUE){
    resultados$hessiana[((length(estimativas$par)*i)-
(length(estimativas$par)-1)):(i*length(estimativas$par)),] <- NaN
  }
}
if(class(estimativas) != "try-error"){
  if(estimativas$convergence == 0){
    convergencia[i] <- 1
  }
  print(i)
  R0 <- (((2^(kappa0-1))*gamma(kappa0))^(-1)) *
  ((dists/estimativas$par[4])^(kappa0)) *
  besselK(dists/estimativas$par[4],nu=kappa0)
  R0[dists < 0.005] <- 1
  R1 <- (((2^(kappa1-1))*gamma(kappa1))^(-1)) *
  ((dists/estimativas$par[5])^(kappa1)) *
  besselK(dists/estimativas$par[5],nu=kappa1)
  R1[dists < 0.005] <- 1
  R2 <- (((2^(kappa2-1))*gamma(kappa2))^(-1)) *
  ((dists/estimativas$par[6])^(kappa2)) *
  besselK(dists/estimativas$par[6],nu=kappa2)
  R2[dists < 0.005] <- 1
  V1 <- R0[k1,k1] + estimativas$par[2]^2 * R1[k1,k1]

```

```

V2 <- estimativas$par[1]^2 * R0[k2,k2] + estimativas$par[3]^2 * R2[k2,k2]
V3 <- estimativas$par[1] * R0[k1,k2]
V <- cbind(rbind(V1,t(V3)),
           rbind(V3,V2))
VinvX <- ginv(V)%*%X
mu <- drop(ginv(crossprod(VinvX,X))%*%crossprod(VinvX,dados[,i]))
dif_y_mu <- dados[,i]-(X%*%mu)
sigmasq <- drop((1/n) * crossprod(dif_y_mu,ginv(V)) %*% dif_y_mu)
#if(sigmasq < 0) sigmasq <- drop((1/n) *
crossprod(dif_y_mu,chol2inv(V)) %*% dif_y_mu)
if(sigmasq < 0) resultados$erros[i,] <- c(i,estimativas$par[1],
estimativas$par[2],estimativas$par[3],
estimativas$par[4],estimativas$par[5],estimativas$par[6])
sigma01 <- sqrt(sigmasq)
sigma02 <- estimativas$par[1] *sigma01
sigma1 <- estimativas$par[2] * sigma01
sigma2 <- estimativas$par[3] * sigma01
if(ruido == TRUE) {
  tau_1 <- estimativas$par[7]*sigma01
  tau_2 <- estimativas$par[8]*sigma01
}
resultados$betas[i,] <- mu
resultados$sigmas[i,] <- c(sigma01,sigma1,sigma02,sigma2)
resultados$phis[i,] <- c(estimativas$par[4],estimativas$par[5],
estimativas$par[6])
if(ruido == TRUE){
  resultados$taus[i,] <- c(tau_1,tau_2)
  resultados$re_param[i,] <- c(estimativas$par[1],estimativas$par[2],
  estimativas$par[3],estimativas$par[7],estimativas$par[8])
}
if(ruido==FALSE) resultados$re_parametros[i,] <- c(estimativas$par[1],
estimativas$par[2],estimativas$par[3])
resultados$log_lik[i,] <- -estimativas$value
if(hessiana == TRUE){
  resultados$hessiana[((length(estimativas$par)*i)-
  (length(estimativas$par)-1)):(i*length(estimativas$par)),] <-
  estimativas$hessian
}
}
}
}
resultados$erros <- subset(resultados$erros,resultados$erros[,"repet"]>0)
print(c("convergência",convergencia))
class(resultados) <- "emvBGCCM"
return(resultados)
}

```

```
require(ASOR)
```

```

Attach()
est1_a <- emvBGCCM_sim(sim1_a$geo1,sim1_a$geo2,corr_ini=c(0.25,0.2,0.2),
kappas=c(0.5,0.5,0.5),hessiana=T,metodo="Nelder-Mead",
controle=list(maxit=10000))
Store(est1_a)

# Estimação para o BCRM
require(MASS)
require(geoR)
emvBCRM_sim <- function(geo1,geo2,corr_ini,ruido=F,kappas=c(0.5,0.5),
metodo="Nelder-Mead",inferior=-Inf,superior=Inf,
controle=list(),hessiana=F){
n1 <- nrow(geo1$coords)
n2 <- nrow(geo2$coords)
n <- n1+n2
N <- ncol(geo1$data)
if(is.null(N)) N <- 1
k1 <- c(1:n1)
k2 <- c((n1+1):n)
k3 <- c(1:n)
k4 <- c(1:n2)
dados <- matrix(0,n,N)
if(N > 1) for(i in 1:N) dados[,i] <- c(geo1$data[,i],geo2$data[,i])
if(N == 1) dados[,1] <- c(geo1$data,geo2$data)
dists <- unname(as.matrix(dist(rbind(geo1$coords,geo2$coords),
diag=TRUE,upper=TRUE)))
X <- matrix(c(rep(1,n1),rep(0,n2),rep(0,n1),rep(1,n2)),n,2)
kappa1 <- kappas[1]
kappa2 <- kappas[2]
ctes <- list()
ctes$n1 <- n1
ctes$n2 <- n2
ctes$dados <- dados
ctes$dists <- dists
ctes$X <- X
est_num<-function(p,Ct,ru=F,kap,repet){
if ( any(p < 1e-20)) return(.Machine$double.xmax^.5)
if ( any(p[3:4] > 1)) return(.Machine$double.xmax^.5)
if ( any(p > 100)) return(.Machine$double.xmax^.5)
R1 <- (((2^(kap[1]-1))*gamma(kap[1]))^(-1)) *
((Ct$dists/p[3])^(kap[1])) * besselK(Ct$dists/p[3],nu=kap[1])
R1[ Ct$dists < 0.005 ] <- 1
R2 <- (((2^(kap[2]-1))*gamma(kap[2]))^(-1)) *
((Ct$dists/p[4])^(kap[2])) * besselK(Ct$dists/p[4],nu=kap[2])
R2[ Ct$dists < 0.005 ] <- 1
V1 <- R1[k1,k1]
V2 <- p[1]^2 * R1[k2,k2] + p[2]^2 * R2[k2,k2]

```

```

V3 <- p[1] * R1 [k1,k2]
V <- cbind(rbind(V1,t(V3)),rbind(V3,V2))
if(rui == TRUE) {diag(V) <- diag(V) + c(rep((p[7]^2),Ct$n1),
rep((p[8]^2),Ct$n2))}
Vinv <- ginv(V)
tVinvX <- crossprod(Vinv,Ct$X)
Q_hat <- drop(sum(crossprod(Ct$dados[,repet],Vinv)*Ct$dados[,repet])
- crossprod(Ct$dados[,repet],tVinvX)%*%
ginv(crossprod(tVinvX,Ct$X))%*%crossprod(tVinvX,Ct$dados[,repet]))
lQ_hat <- log(Q_hat)
ldv <- log(det(V))
if(is.infinite(lQ_hat) | is.nan(lQ_hat) |
is.infinite(ldv) | is.nan(ldv)) return(.Machine$double.xmax^.5)
ll <- drop(-0.5*((Ct$n1+Ct$n2)*(log(2*pi)+lQ_hat-log(Ct$n1+Ct$n2)+1)+ldv))
if(is.infinite(ll) | is.nan(ll)) return(.Machine$double.xmax^.5)
return(-ll)
}
resultados <- list()
resultados$betas <- matrix(0,N,2)
colnames(resultados$betas) <- c("mu1","mu2")
resultados$sigmas <- matrix(0,N,3)
colnames(resultados$sigmas) <- c("s11","s21","s22")
resultados$phis <- matrix(0,N,2)
colnames(resultados$phis) <- c("phi1","phi2")
if(ruido == TRUE){
resultados$taus <- matrix(0,N,2)
colnames(resultados$taus) <- c("tau1","tau2")
resultados$re_parametros <- matrix(0,N,4)
colnames(resultados$re_parametros) <- c("nu1","nu2","psi1","psi2")
if(hessiana == TRUE){
resultados$hessiana <- matrix(0,(6*N),6)
colnames(resultados$hessiana) <- c("nu1","nu2","phi1"
,"phi2","psi1","psi2")
rownames(resultados$hessiana) <- rep(c("nu1","nu2","phi1"
,"phi2","psi1","psi2"),N)
}
}
if(ruido==FALSE){
resultados$re_parametros <- matrix(0,N,2)
colnames(resultados$re_parametros) <- c("nu1","nu2")
if(hessiana == TRUE){
resultados$hessiana <- matrix(0,(4*N),4)
colnames(resultados$hessiana) <- c("nu1","nu2","phi1","phi2")
rownames(resultados$hessiana) <- rep(c("nu1","nu2","phi1","phi2"),N)
}
}
resultados$log_lik <- matrix(0,N,1)

```

```

resultados$kappas <- c(kappa1=kappas[1],kappa2=kappas[2])
resultados$erros <- matrix(0,N,5)
colnames(resultados$erros) <- c("repet","nu1","nu2","phi1","phi2")
convergencia <- rep(0,N)
for(i in 1:N){
  var1_amost <- var(geo1$data[,i])
  var2_amost <- var(geo2$data[,i])
  var_ini <- NULL
  var_ini[1] <- sqrt(var1_amost)
  var_ini[2] <- sqrt(0.8*var2_amost)
  var_ini[3] <- sqrt(0.2*var2_amost)
  repar_ini <- 0
  repar_ini[1] <- var_ini[2]/var_ini[1]
  repar_ini[2] <- var_ini[3]/var_ini[1]
  repar_ini[3] <- corr_ini[1]
  repar_ini[4] <- corr_ini[2]
  if(ruido == TRUE) {
    repar_ini[5] <- var_ini[4]/var_ini[1]
    repar_ini[6] <- var_ini[5]/var_ini[1]
  }

  estimativas <- try(optim(p=repar_ini,fn=est_num,Ct=ctes,
  rui=ruido,kap=c(kappa1,kappa2),repet=i,method=metodo,lower=inferior,
  upper=superior,control=controle,hessian=hessiana))
  if(class(estimativas) == "try-error"){
    resultados$erros[i,] <- c(i,NaN,NaN,NaN,NaN)
    resultados$betas[i,] <- c(NaN,NaN)
    resultados$sigmas[i,] <- c(NaN,NaN,NaN)
    resultados$phis[i,] <- c(NaN,NaN)
    if(ruido == TRUE){
      resultados$taus[i,] <- c(NaN,NaN)
      resultados$re_param[i,] <- c(NaN,NaN,NaN,NaN)
    }
    if(ruido==FALSE) resultados$re_parametros[i,] <- c(NaN,NaN)
    resultados$log_lik[i,] <- NaN
    if(hessiana == TRUE){
      resultados$hessiana[((length(estimativas$par)*i)-
      (length(estimativas$par)-1)):(i*length(estimativas$par)),] <- NaN
    }
  }
  if(class(estimativas) != "try-error"){
    if(estimativas$convergence == 0){
      convergencia[i] <- 1
    }
    print(i)
    R1 <- (((2^(kappa1-1))*gamma(kappa1))^(-1)) *
    ((dists/estimativas$par[3])^(kappa1)) *

```



```

besselK(dists/estimativas$par[3],nu=kappa1)
R1[dists < 0.005] <- 1
R2 <- (((2^(kappa2-1))*gamma(kappa2))^(-1)) *
((dists/estimativas$par[4])^(kappa2)) *
besselK(dists/estimativas$par[4],nu=kappa2)
R2[dists < 0.005] <- 1
V1 <- R1[k1,k1]
V2 <- estimativas$par[1]^2 * R1[k2,k2] + estimativas$par[2]^2 *
R2[k2,k2]
V3 <- estimativas$par[1] * R1[k1,k2]
V <- cbind(rbind(V1,t(V3)),
           rbind(V3,V2))
VinvX <- ginv(V)%*%X
mu <- drop(ginv(crossprod(VinvX,X))%*%crossprod(VinvX,dados[,i]))
dif_y_mu <- dados[,i]-(X%*%mu)
sigmasq <- drop((1/n) * crossprod(dif_y_mu,ginv(V)) %*% dif_y_mu)
#if(sigmasq < 0) sigmasq <- drop((1/n) *
crossprod(dif_y_mu,chol2inv(V)) %*% dif_y_mu)
if(sigmasq < 0) resultados$erros[i,] <- c(i,estimativas$par[1],
estimativas$par[2],estimativas$par[3],estimativas$par[4])
sigma11 <- sqrt(sigmasq)
sigma21 <- estimativas$par[1] *sigma11
sigma22 <- estimativas$par[2] * sigma11
if(ruido == TRUE) {
  tau_1 <- estimativas$par[5]*sigma11
  tau_2 <- estimativas$par[6]*sigma11
}
resultados$betas[i,] <- mu
resultados$sigmas[i,] <- c(sigma11,sigma21,sigma22)
resultados$phis[i,] <- c(estimativas$par[3],estimativas$par[4])
if(ruido == TRUE){
  resultados$taus[i,] <- c(tau_1,tau_2)
  resultados$re_param[i,] <- c(estimativas$par[1],estimativas$par[2],
  estimativas$par[5],estimativas$par[6])
}
if(ruido==FALSE) resultados$re_parametros[i,] <- c(estimativas$par[1],
estimativas$par[2])
resultados$log_lik[i,] <- -estimativas$value
if(hessiana == TRUE){
  resultados$hessiana[((length(estimativas$par)*i)-
(length(estimativas$par)-1)):(i*length(estimativas$par)),] <-
  estimativas$hessian
}
}
}
}
resultados$erros <- subset(resultados$erros,resultados$erros[,"repet"]>0)
print(c("convergência",convergencia))

```

```

class(resultados) <- "emvBGCCM"
return(resultados)
}

```

```

est1_a <- emvBCRM_sim(sim1_a$geo1,sim1_a$geo2,corr_ini=c(0.25,0.2),
kappas=c(0.5,0.5),hessiana=T,metodo="Nelder-Mead",
controle=list(maxit=10000))
Store(est1_a)

```

```

#####
## Programa para predição espacial com os modelos geoestísticos bivariados#
#####

```

```

#Krigagem para o BGCCM
predBGCCM_sim <- function(geos,parametros){
geol <- geos$geol
geo2 <- geos$geo2
geo_a1 <- geos$geo_arm1
geo_a2 <- geos$geo_arm2
if(nrow(parametros$erros)>0){
geol$data <- geol$data[,-c(parametros$erros[,1])]
geo2$data <- geo2$data[,-c(parametros$erros[,1])]
geo_a1$data <- geo_a1$data[,-c(parametros$erros[,1])]
geo_a2$data <- geo_a2$data[,-c(parametros$erros[,1])]
}
grp <- geo_a1$coords
coords1 <- geol$coords
coords2 <- geo2$coords
n1 <- nrow(coords1)
n2 <- nrow(coords2)
n <- n1+n2
npr <- nrow(grp)
nt <- npr + n
k1 <- c(1:n1)
k2 <- c(1:n2)
k3 <- c(1:npr)
k4 <- c((npr+1):(npr+n1))
k5 <- c((npr+1):(2*npr))
k6 <- c((npr+n1+1):nt)
k7 <- c((n1+1):n)
k8 <- c((npr+1):nt)
k9 <- c(((2*npr)+1):((2*npr)+n))
N <- ncol(geol$data)
if(is.null(N)) N <- 1
dados <- matrix(0,n,N)
if(N > 1) for(i in 1:N) dados[,i] <- c(geol$data[,i],geo2$data[,i])

```

```

if(N == 1) dados[,1] <- c(geo1$data,geo2$data)
gr_dados <- rbind(coords1,coords2)
gr <- rbind(grp,gr_dados)
dists <- matrix(0,nt,n)
for(i in 1:nt){
  dists[i,] <- spDistsN1(gr_dados,gr[i,])
}
X <- matrix(c(rep(1,npr),rep(0,npr),rep(1,n1),rep(0,n2),rep(0,npr),
rep(1,npr),rep(0,n1),rep(1,n2)),(2*npr+n),2)
betas <- parametros$betas
sigmas <- parametros$sigmas
phis <- parametros$phis
if(nrow(parametros$erros)>0){
  betas <- matrix(c(parametros$betas[-c(parametros$erros[,1]),]),N,2)
  sigmas <- matrix(c(parametros$sigmas[-c(parametros$erros[,1]),]),N,4)
  phis <- matrix(c(parametros$phis[-c(parametros$erros[,1]),]),N,3)
}
if(!is.null(parametros$taus)) taus <- parametros$taus
kappas <- parametros$kappas
saida <- list()
saida$pred1 <- matrix(0,npr,N)
saida$pred2 <- matrix(0,npr,N)
saida$var_pred1 <- matrix(0,npr,N)
saida$var_pred2 <- matrix(0,npr,N)
for(l in 1:N){
  R0 <- (((2^(kappas[1]-1))*gamma(kappas[1]))^(-1)) *
  ((dists/phis[1,1])^(kappas[1])) * besselK(dists/phis[1,1],
  nu=kappas[1])
  R0[dists < 0.005] <- 1
  R1 <- (((2^(kappas[2]-1))*gamma(kappas[2]))^(-1)) *
  ((dists[k4,k1]/phis[1,2])^(kappas[2])) * besselK(dists[k4,k1]/phis[1,2],
  nu=kappas[2])
  distaux <- dists[k4,k1]
  R1[distaux < 0.005] <- 1
  R2 <- (((2^(kappas[3]-1))*gamma(kappas[3]))^(-1)) *
  ((dists[k6,k7]/phis[1,3])^(kappas[3])) * besselK(dists[k6,k7]/phis[1,3],
  nu=kappas[3])
  distaux <- dists[k6,k7]
  R2[distaux < 0.005] <- 1
  Sig1 <- (sigmas[1,1]^2)*R0[k4,k1] + (sigmas[1,2]^2)*R1
  Sig2 <- (sigmas[1,3]^2)*R0[k6,k7] + (sigmas[1,4]^2)*R2
  Sig12 <- (sigmas[1,1]*sigmas[1,3])*R0[k4,k7]
  Sig <- cbind(rbind(Sig1,t(Sig12)),rbind(Sig12,Sig2))
  if(!is.null(parametros$taus)){
    diag(Sig) <- diag(Sig) + c(rep(taus[1,1],n1),rep(taus[1,2],n2))
  }
}
R1c <- (((2^(kappas[2]-1))*gamma(kappas[2]))^(-1)) *

```

```

((dists[k3,k1]/phis[1,2])^(kappas[2])) * besselK(dists[k3,k1]/phis[1,2],
nu=kappas[2])
distaux <- dists[k3,k1]
R1c[distaux < 0.005] <- 1
Sig_c1 <- (sigmas[1,1]^2)*R0[k3,k1] + (sigmas[1,2]^2)*R1c
Sig_c12 <- (sigmas[1,1]*sigmas[1,3]) * R0[k3,k7]
Sig_c <- cbind(Sig_c1,Sig_c12)
m_p <- as.vector(X%%betas[1,])
m_z_dadoy <- m_p[k3] - (Sig_c%%ginv(Sig)%%(m_p[k9]-dados[,1]))
pondera_var <- Sig_c%%tcrossprod(ginv(Sig),Sig_c)
saida$pred1[,1] <- m_z_dadoy
saida$var_pred1[,1] <- ((sigmas[1,1]^2) + (sigmas[1,2]^2)) -
diag(pondera_var)
R2c <- (((2^(kappas[3]-1))*gamma(kappas[3]))^(-1)) *
((dists[k3,k7]/phis[1,3])^(kappas[3])) *
besselK(dists[k3,k7]/phis[1,3],nu=kappas[3])
distaux <- dists[k3,k7]
R2c[distaux < 0.005] <- 1
Sig_c2 <- (sigmas[1,3]^2)*R0[k3,k7] + (sigmas[1,4]^2)*R2c
Sig_c21 <- (sigmas[1,1]*sigmas[1,3]) * R0[k3,k1]
Sig_c <- cbind(Sig_c21,Sig_c2)
m_z_dadoy <- m_p[k5] - (Sig_c%%ginv(Sig)%%(m_p[k9]-dados[,1]))
pondera_var <- Sig_c%%tcrossprod(ginv(Sig),Sig_c)
saida$pred2[,1] <- m_z_dadoy
saida$var_pred2[,1] <- ((sigmas[1,3]^2) + (sigmas[1,4]^2)) -
diag(pondera_var)
}
return(saida)
}
require(ASOR)
Attach()
pr1_a <- predBGCCM_sim(geos=sim1_a,parametros=est1_a)
Store(pr1_a)

#Krigagem para o BCRM
predBCRM_sim <- function(geos,parametros){
geol <- geos$geol
geo2 <- geos$geo2
geo_a1 <- geos$geo_arm1
geo_a2 <- geos$geo_arm2
if(nrow(parametros$erros)>0){
geol$data <- geol$data[,-c(parametros$erros[,1])]
geo2$data <- geo2$data[,-c(parametros$erros[,1])]
geo_a1$data <- geo_a1$data[,-c(parametros$erros[,1])]
geo_a2$data <- geo_a2$data[,-c(parametros$erros[,1])]
}
grp <- geo_a1$coords

```

```

coords1 <- geo1$coords
coords2 <- geo2$coords
n1 <- nrow(coords1)
n2 <- nrow(coords2)
n <- n1+n2
npr <- nrow(grp)
nt <- npr + n
k1 <- c(1:n1)
k2 <- c(1:n2)
k3 <- c(1:npr)
k4 <- c((npr+1):(npr+n1))
k5 <- c((npr+1):(2*npr))
k6 <- c((npr+n1+1):nt)
k7 <- c((n1+1):n)
k8 <- c((npr+1):nt)
k9 <- c(((2*npr)+1):((2*npr)+n))
N <- ncol(geo1$data)
if(is.null(N)) N <- 1
dados <- matrix(0,n,N)
if(N > 1) for(i in 1:N) dados[,i] <- c(geo1$data[,i],geo2$data[,i])
if(N == 1) dados[,1] <- c(geo1$data,geo2$data)
gr_dados <- rbind(coords1,coords2)
gr <- rbind(grp,gr_dados)
dists <- matrix(0,nt,n)
for(i in 1:nt){
  dists[i,] <- spDistsN1(gr_dados,gr[i,])
}
X <- matrix(c(rep(1,npr),rep(0,npr),rep(1,n1),rep(0,n2),
rep(0,npr),rep(1,npr),rep(0,n1),rep(1,n2)),(2*npr+n),2)
betas <- parametros$betas
sigmas <- parametros$sigmas
phis <- parametros$phis
if(nrow(parametros$erros)>0){
  betas <- matrix(c(parametros$betas[-c(parametros$erros[,1]),]),N,2)
  sigmas <- matrix(c(parametros$sigmas[-c(parametros$erros[,1]),]),N,3)
  phis <- matrix(c(parametros$phis[-c(parametros$erros[,1]),]),N,2)
}
if(!is.null(parametros$taus)) taus <- parametros$taus
kappas <- parametros$kappas
saida <- list()
saida$pred1 <- matrix(0,npr,N)
saida$pred2 <- matrix(0,npr,N)
saida$var_pred1 <- matrix(0,npr,N)
saida$var_pred2 <- matrix(0,npr,N)
for(l in 1:N){
  R1 <- (((2^(kappas[l]-1))*gamma(kappas[l]))^(-1)) *
  ((dists/phis[l,1])^(kappas[l])) * besselK(dists/phis[l,1],nu=kappas[l])

```

```

R1[dists < 0.005] <- 1
R2 <- (((2^(kappas[2]-1))*gamma(kappas[2]))^(-1)) *
((dists[k6,k7]/phis[1,2])^(kappas[2])) *
besselK(dists[k6,k7]/phis[1,2],nu=kappas[2])
distaux <- dists[k6,k7]
R2[distaux < 0.005] <- 1
Sig1 <- (sigmas[1,1]^2)*R1[k4,k1]
Sig2 <- (sigmas[1,2]^2)*R1[k6,k7] + (sigmas[1,3]^2)*R2
Sig12 <- (sigmas[1,1]*sigmas[1,2])*R1[k4,k7]
Sig <- cbind(rbind(Sig1,t(Sig12)),rbind(Sig12,Sig2))
if(!is.null(parametros$taus)){
  diag(Sig) <- diag(Sig) + c(rep(taus[1,1],n1),rep(tau[1,2],n2))
}
Sig_c1 <- (sigmas[1,1]^2)*R1[k3,k1]
Sig_c12 <- (sigmas[1,1]*sigmas[1,2]) * R1[k3,k7]
Sig_c <- cbind(Sig_c1,Sig_c12)
m_p <- as.vector(X%betas[1,])
m_z_dadoy <- m_p[k3] - (Sig_c%ginv(Sig)%%(m_p[k9]-dados[,1]))
pondera_var <- Sig_c%tcrossprod(ginv(Sig),Sig_c)
saida$pred1[,1] <- m_z_dadoy
saida$var_pred1[,1] <- (sigmas[1,1]^2) - diag(pondera_var)
R2c <- (((2^(kappas[2]-1))*gamma(kappas[2]))^(-1)) *
((dists[k3,k7]/phis[1,2])^(kappas[2])) *
besselK(dists[k3,k7]/phis[1,2],nu=kappas[2])
distaux <- dists[k3,k7]
R2c[distaux < 0.005] <- 1
Sig_c2 <- (sigmas[1,2]^2)*R1[k3,k7] + (sigmas[1,3]^2)*R2c
Sig_c21 <- (sigmas[1,1]*sigmas[1,2]) * R1[k3,k1]
Sig_c <- cbind(Sig_c21,Sig_c2)
m_z_dadoy <- m_p[k5] - (Sig_c%ginv(Sig)%%(m_p[k9]-dados[,1]))
pondera_var <- Sig_c%tcrossprod(ginv(Sig),Sig_c)
saida$pred2[,1] <- m_z_dadoy
saida$var_pred2[,1] <- ((sigmas[1,2]^2) + (sigmas[1,3]^2)) -
diag(pondera_var)
}
return(saida)
}
pr1_a <- predBCRM_sim(geos=sim1_a,parametros=est1_a)
Store(pr1_a,file="pr1_a.RData")

```