

UNIVERSIDADE DE SÃO PAULO
ESCOLA SUPERIOR DE AGRICULTURA “LUIZ DE QUEIROZ”
DEPARTAMENTO DE CIÊNCIAS EXATAS

LUIZ RICARDO NAKAMURA

ANÁLISE DE VARIÁVEIS DO SOLO DA REGIÃO DE UNAÍ - MG

Disciplina: Geoestatística
Professor: Paulo Justiniano Ribeiro Jr

Novembro/2012

RESUMO

Compreender a distribuição espacial de dados provenientes de fenômenos ocorridos no espaço constitui um grande desafio para esclarecer questões centrais em diversas áreas do conhecimento, seja na área da saúde, em ambiente, em geologia, em agronomia, entre tantas outras áreas. Uma metodologia bastante utilizada para este fim é a geoestatística, com a qual se analisa a dependência espacial e realiza a interpolação de atributos através da krigagem. Este trabalho teve como objetivo analisar por meio de ferramentas geoestatísticas a variabilidade espacial do cálcio e do pH da água de um estudo realizado no Cerrado (savanas neotropicais brasileiras), como parte da investigação sobre o impacto das práticas de agricultura de conservação da fertilidade do solo e produtividade das culturas. A área em estudo está localizada em Unaí, noroeste de Minas Gerais, Brasil. Para isso, ajustou-se modelos de semivariograma e estimou-se os parâmetros destes modelos. Como critério de escolha do melhor modelo foram utilizados o AIC, BIC e o valor do RD. Após a escolha do melhor modelo, realizou-se a krigagem na área de estudo.

Palavras-chave: Cálcio; Geoestatística; Krigagem; pH da água

1 INTRODUÇÃO

Segundo Druck et al. (2004), compreender a distribuição espacial de dados provenientes de fenômenos ocorridos no espaço constitui um grande desafio para esclarecer questões centrais em diversas áreas do conhecimento, seja na área da saúde, em ambiente, em geologia, em agronomia, entre tantas outras áreas. Tais estudos tornam-se cada vez mais comuns, devido a disponibilidade de sistemas de informação geográfica (SIG). Desta forma, uma área da estatística que vem crescendo muito na atualidade é a estatística espacial.

A estatística espacial tem como intuito mensurar propriedades e relacionamentos, levando em consideração a localização espacial do fenômeno estudado de forma explícita. Essa área da estatística é dividida em três áreas de estudo: geoestatística, dados de área e processos pontuais. Como o que distingue cada uma dessas categorias é o tipo de dado aleatório utilizado, é normal que em cada um destas três áreas, existam métodos estatísticos diferentes para analisar e descrever os dados. Neste trabalho, foi utilizado a geoestatística e seus conceitos.

Geoestatística é um ramo da geografia matemática e da estatística, que une o conceito de variáveis aleatórias com o conceito de variáveis regionalizadas, gerando um novo conceito de funções aleatórias. Neste ramo, as técnicas que mais se destacam são krigagem e a simulação estocástica e, com o auxílio destas e outras técnicas, é possível calcular um valor de uma dada região em estudo, para cada centro da célula de uma malha tridimensional, valor este condicionado aos dados amostrados e a uma função de correlação espacial entre estes dados.

Assim, este trabalho utilizou destas técnicas da geoestatística para analisar a variabilidade espacial do cálcio e do pH da água, em um região localizada em Unaí, noroeste de Minas Gerais, Brasil. Complementarmente, realizou-se uma análise geoestatística com enfoque bayesiana na variável pH da água.

2 MATERIAL E MÉTODOS

2.1 Área de estudo

O estudo foi conduzido no Cerrado (savanas neotropicais brasileiros), como parte da investigação sobre o impacto das práticas de agricultura de conservação da fertilidade do solo e produtividade das culturas. A área em estudo está localizada em Unaí, noroeste de Minas Gerais, Brasil (1). O experimento de campo foi localizado na Faculdade Juvêncio Ferreira Martins Agrícola de Unaí ($16^{\circ} 32' 26''$ S, $46^{\circ} 50' 44''$ W e altitude de 600 m). Essa região é caracterizada pelo clima subúmido tipicamente tropical do Cerrado brasileiro. Segundo a classificação de Köppen, o clima é tropical úmido e seco "Aw" (ou clima de savana).

A área em estudo é um quadrado de 80 m x 80 m. As amostras de solo foram recolhidas, em 2010, por meio de uma grade de 5 x 5 m, a uma profundidade de 0-20 cm, resultando num total de 240 amostras. Das variáveis coletadas, neste trabalho, foram analisados o cálcio (Ca) e o pH da água (pH H₂O).

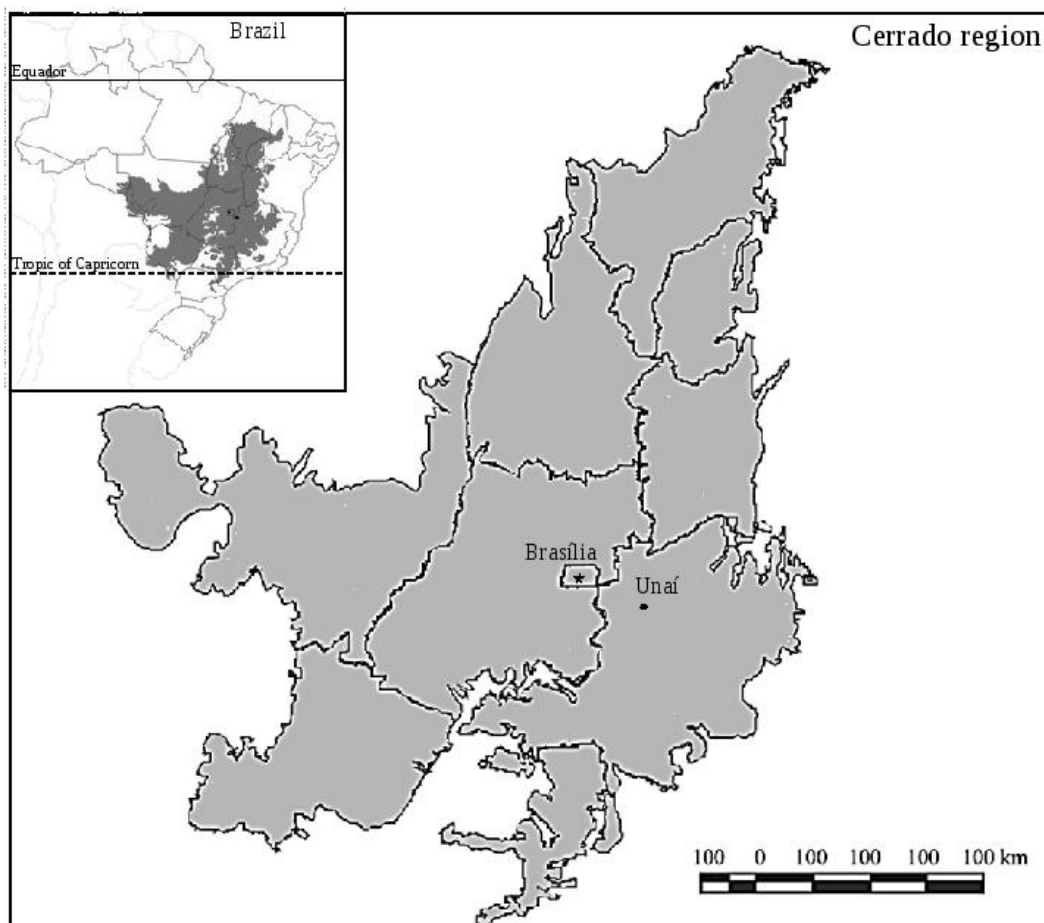


Figura 1: Região de estudo

2.2 Métodos

As análises foram realizadas no software R, utilizando o pacote geoR (RIBEIRO JR. e DIGLLE, 2001).

Primeiramente, verificou-se a normalidade univariada das variáveis e, a um nível de significância de 5%, apenas pH H₂O apresentou normalidade; ao passo que a variável Ca requereu uma transformação raiz quadrada para obter tal distribuição. Posteriormente, realizou-se a construção de um variograma a partir do cálculo da semivariância $\gamma(h)$ dada por:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^N [y(x_i + h) - y(x_i)]^2 \quad (1)$$

O variograma consiste em construir classes de pares de pontos equidistantes e realizar uma espécie de estudo da evolução da correlação entre os grupos conforme se caminha na região em estudo, ou seja, uma função matemática dada pela equação 1 que depende de h, a distância entre pontos de um mesmo grupo, que fornece o efeito de sinal de σ^2 , o qual mede o quanto o padrão espacial está nítido - quanto maior o sinal maior a confiança na existência de dependência espacial -, além do efeito do ruído τ^2 , também chamado efeito pepita, que define a variabilidade não explicada. Para afirmar a dependência espacial dos dados com mais segurança, é de costume construir um envelope de variogramas que funciona como uma espécie de intervalo de credibilidade para a aparente dependência espacial do variograma empírico obtido.

Outro critério utilizado para afirmar essa dependência espacial é denominado como cálculo da razão de dependência espacial (RD). Segundo Cambardella et al. (1994), o valor do RD é dado em porcentagem segundo uma função que depende do efeito pepita do modelo (τ^2) e do sinal de σ^2 . A equação para o cálculo dessa razão é dada por:

$$RD = \frac{\tau^2}{\tau^2 + \sigma^2} \times 100\% \quad (2)$$

A análise dos valores obtidos na equação (2) é dada por:

- $RD \leq 25\%$, indica dependência espacial forte;
- $25\% < RD < 75\%$, a dependência é moderada;
- $RD \geq 75\%$, a dependência é fraca.

Alguns modelos foram ajustados por meio do método da máxima verossimilhança e para a escolha do melhor ajustado, os critérios utilizados foram o valor do RD, o critério de Akaike (AIC) e o critério de Schwarz (BIC), dados pelas equações (2), (3) e (4) respectivamente.

$$AIC = -2\ln(L) + 2p \quad (3)$$

$$BIC = -2\ln(L) + \ln(np) \quad (4)$$

Nas equações (3) e (4), L é a função de verossimilhança e n é o número de parâmetros ajustados. Para ambos os critérios dessas equações, o melhor modelo é aquele com o menor valor obtido de AIC e BIC.

Assim, analisando a existência da dependência espacial, pode-se estimar pontos desconhecidos na área em estudo, pois estes serão semelhantes aos seus vizinhos, aos pontos amostrados. Esta estimação é realizada por meio da krigagem, que estima valores com variância mínima. O uso desta técnica permite a construção de uma média ponderada que atribui pesos aos pontos vizinhos que são conhecidos, sendo que, usualmente, para vizinhos mais próximos, o peso atribuído é maior. Esta interpolação de valores é dada conforme a seguinte equação:

$$Z(x_0) = \sum_{i=1}^N \lambda_i Z(x_i) \quad (5)$$

As construções de mapas com os valores obtidos por meio da krigagem são importantes para a verificação e interpretação da variabilidade espacial dos dados em estudo.

Finalmente, foi realizado também a análise geoestatística com enfoque bayesiano, em que intervalos de credibilidade foram obtidos para os parâmetros e os resultados obtidos foram próximos aos da abordagem clássica.

3 RESULTADOS E DISCUSSÃO

3.1 Variável pH H₂O

Primeiramente, foi realizado o teste de Shapiro-Wilk na variável pH H₂O e constatou-se normalidade ao nível de 5% de significância. Na Tabela 1 são apresentadas estatísticas descritivas desta variável.

Tabela 1: Estatística descritiva da variável pH H₂O

Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
4,150	4,898	5,070	5,059	5,222	6,030

O gráfico superior a esquerda, apresentado na Figura 2, categoriza os dados nos quartis amostrais das observações, em que os símbolos “+”, “Δ”, “o”, “x”, nesta ordem, indicam os quartis amostrais. Essa imagem confirma a idéia da existência de padrão espacial nos dados, uma vez que existem conglomerados das categorias. Essa dependência espacial também é confirmada com a observação do envelope do variograma (Figura 3). Como existem pontos fora do envelope, pode-se afirmar que existe esta dependência. Ainda na Figura 2, pode-se observar o gráfico da coordenada Y pelos dados (canto superior direito) e no canto inferior esquerdo, gráfico dos dados pela coordenada X e, por último, no canto inferior direito tem-se o histograma da variável.

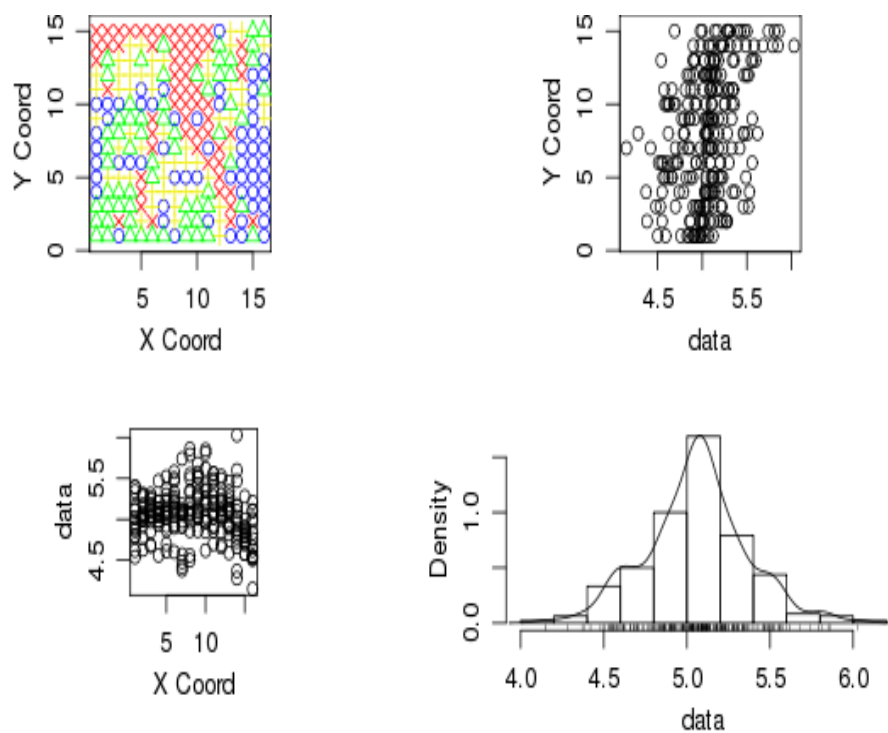


Figura 2: Plot da variável pH H₂O para análise exploratória

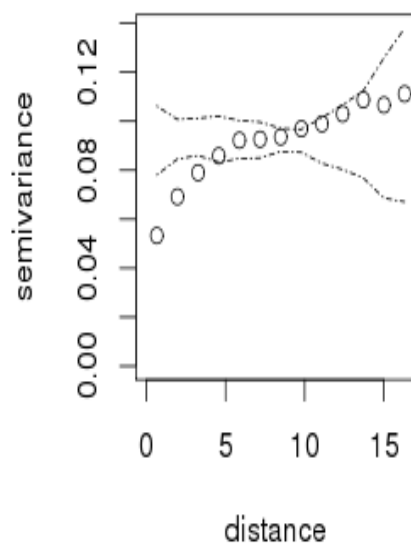


Figura 3: Gráfico dos envelopes do variograma da variável pH H₂O

O próximo passo foi modelar uma função que explique bem a variável estudada. Para ajustar

os modelos por meio do método da máxima verossimilhança, os valores iniciais dos parâmetros foram: $\sigma^2 = 0,9 * \text{Var}(\text{dados}) = 0.0833$, $\tau^2 = 0,1 * \text{Var}(\text{dados}) = 0.0092$ e para ϕ o valor inicial foi a máxima distância entre os pontos; e, assim, ajustou-se dois modelos: um utilizando exponencial e outro a função Matérn.

Na Tabela 2, são apresentados os valores estimados dos parâmetros para os dois modelos e na Tabela 3 apresenta-se os valores do RD, do AIC e BIC, para a comparação dos modelos em estudo.

Tabela 2: Valores dos parâmetros estimados dos modelos ajustado para pH H₂O

Modelo	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\phi}$	$\hat{\beta}_0$
Exponencial	0.0694	0.0392	3.8699	5.0156
Matérn	0.0562	0.0474	2.2218	5.0219

Tabela 3: Valores dos métodos para comparação dos modelos para a variável pH H₂O

Modelo	RD	AIC	BIC
Exponencial	36.10	41.197	55.120
Matérn	45.75	42.231	56.154

Observando a Tabela 3, nota-se que o melhor modelo é aquele com a função exponencial, pois este modelo ao ser comparado com o modelo ajustado com a função Matérn, possui menores valores para AIC e BIC e possui um melhor valor de RD, logo o modelo escolhido é com a função exponencial. Desta forma, realizou-se a krigagem da área em estudo, com o modelo escolhido, para a variável pH H₂O. Esta krigagem pode ser visualizada na Figura 4.

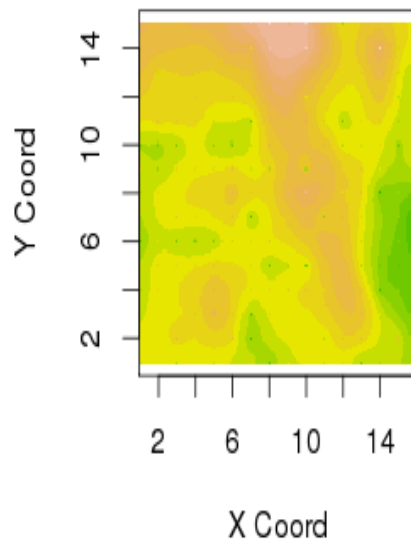


Figura 4: Krigagem da variável pH H₂O

Analisando a Figura 4, percebe-se que os maiores valores da variável estão no canto da área estudada e ao centro da área os valores são menores (em vermelho).

3.2 Análise Bayesiana da Variável pH H₂O

Na análise pela abordagem Bayesiana obteve-se os resultados representados na Tabela 4. Observa-se que os valores são próximos dos resultados obtidos pela análise clássica. A proximidade nas estimativas pode ser atribuída na escolha das funções de distribuição das prioris para os parâmetros do modelo.

defTabelaTabela

Tabela 4: Intervalo de Credibilidade para os parâmetros

Parâmetro	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
β	4,779	5,007	5,042	5,039	5,078	5,312
σ^2	0,062	0,083	0,107	0,105	0,124	0,342
ϕ	0,821	0,821	1,641	1,289	1,641	4,924
τ^2	0	0	0	0	0	0

A Figura 5 representa a krigagem, utilizando-se a função de covariância escolhida no modelo clássico, realizada por meio da abordagem bayesiana.

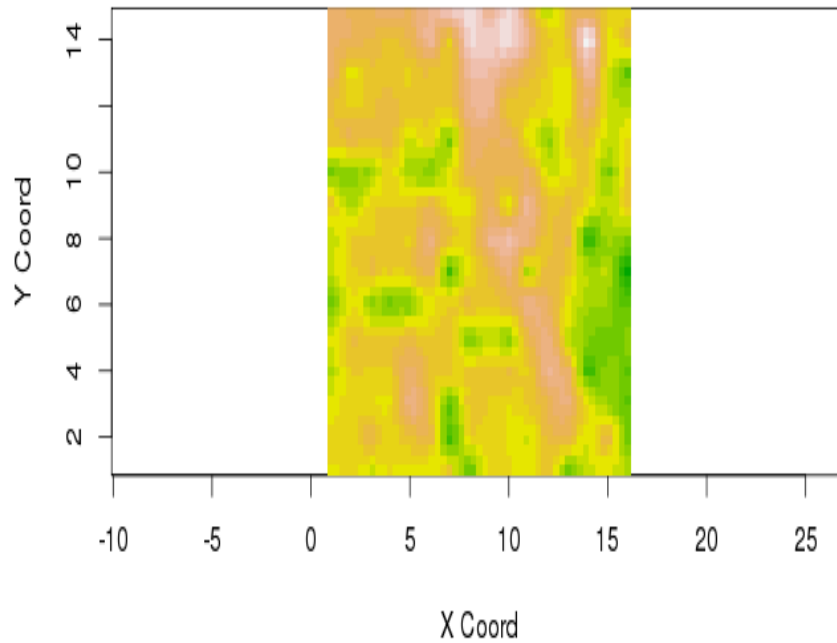


Figura 5: Krigagem da variável pH H₂O realizada pela abordagem bayesiana

Observa-se que a krigagem efetuada utilizando-se a abordagem bayesiana, está muito próxima àquela encontrada do ponto de vista clássico. Isso deve-se principalmente a escolha de prioris não-informativas para os parâmetros do modelo em estudo.

3.3 Variável Ca

Realizou-se o teste Shapiro-Wilk na variável Ca e constatou-se que a mesma não possuía normalidade ao nível de 5% de significância. Desta maneira, se fez necessária a transformação da mesma para atingir este pressuposto. A transformação utilizada foi a raiz quadrada, assim a distribuição dos dados foi a gaussiana. Na Tabela ?? estão as estatísticas descritivas da variável, sem e com transformação.

Tabela 5: Estatística descritiva da variável Ca

	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
Sem Transformação	5,70	12,68	16,75	16,89	20,12	38,70
Com transformação raiz quadrada	2,39	3,56	4,09	4,05	4,49	6,22

O gráfico superior a esquerda, apresentado na Figura 6, categoriza os dados nos quartis

amostrais das observações, em que os símbolos “+”, “Δ”, “o”, “x”, nesta ordem, indicam os quartis amostral. Essa imagem confirma a idéia da existência de padrão espacial nos dados, uma vez que existem conglomerados das categorias. Essa dependência espacial também é confirmada com a observação do envelope do variograma (Figura 7). Como existem pontos fora do envelope, pode-se afirmar que existe esta dependência. Ainda na Figura 6, pode-se observar o gráfico da coordenada Y pelos dados (canto superior direito) e no canto inferior esquerdo, gráfico dos dados pela coordenada X e, por último, no canto inferior direito é o histograma da variável transformada.

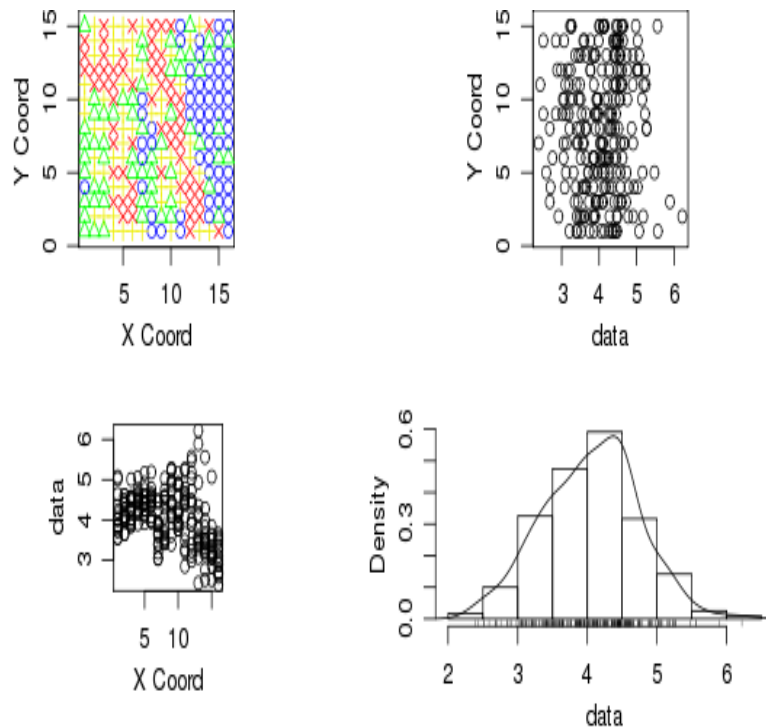


Figura 6: Plot da variável Ca para análise exploratória

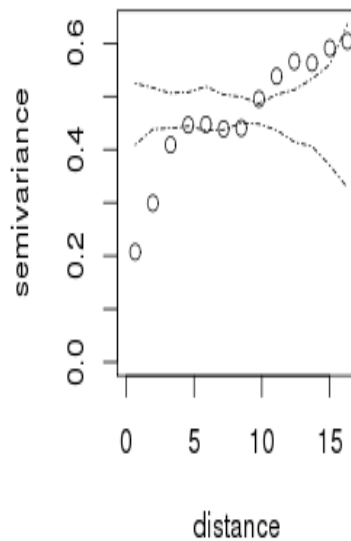


Figura 7: Gráfico dos envelopes do variograma da variável Ca

Em seguida, modelou-se uma função que explicasse bem a variável estudada. Para ajustar os modelos por meio do método da máxima verossimilhança, os valores iniciais dos parâmetros foram: $\sigma^2 = 0,9 * \text{Var}(\text{dados}) = 0,420942$, $\tau^2 = 0,1 * \text{Var}(\text{dados}) = 0,04677133$ e para ϕ o valor inicial atribuído foi a máxima distância entre os pontos. Desta forma, ajustou-se dois modelos: um utilizando a função cúbica e o outro a função Matérn.

Na Tabela 6 apresenta-se os valores estimados dos parâmetros para os dois modelos e na Tabela 7, os valores do RD, do AIC e BIC, para a comparação dos modelos e a escolha do melhor ajustado.

Tabela 6: Valores dos parâmetros estimados dos modelos ajustados para Ca

Modelos	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\phi}$	$\hat{\beta}_0$
Cúbica	0,2814	0,1794	6,1824	4,0098
Matérn	0,3929	0,0986	2,2858	3,9620

Tabela 7: Valores dos métodos para comparação dos modelos para a variável Ca

Modelos	RD	AIC	BIC
Cúbica	38,93	374,4077	388,3303
Matern	20,06	374,7391	388,6616

Observando a Tabela 7 nota-se que os valores de AIC e BIC estão muito parecidos para os dois modelos ajustados; ao analisar o valor do RD, o modelo com a função Matérn se sobressai, indicando uma dependência espacial forte, ao passo que o modelo com a função cúbica indica dependência espacial moderada. Assim, o modelo escolhido foi aquele que utiliza a função Matérn e após essa escolha realizou-se a krigagem na área de estudo (Figura 8).

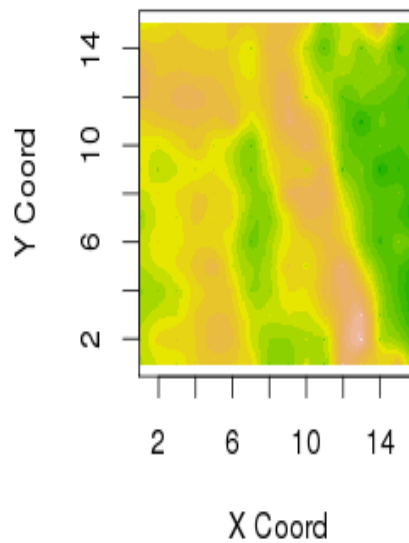


Figura 8: Krigagem da variável Ca

Analisando a Figura 8, observa-se que no canto direito da área esta concentrado maiores valores de cálcio e no canto esquerdo valores mais baixos (vermelho) desta variável.

4 CONCLUSÃO

A partir do semivariograma é possível verificar e modelar a dependência espacial, realizada por meio de funções de correlação. Neste trabalho, para a variável pH H₂O, ajustou-se modelos com funções exponencial e Matérn, sendo que o modelo com a função exponencial foi o melhor ajustado. Para a variável Ca, ajustou-se modelos utilizando as funções cúbica e Matérn, obtendo melhor modelo com a função Matérn. Após o ajuste do melhor modelo, realizou-se a krigagem e, no caso da variável pH H₂O, a krigagem foi realizada também da forma bayesiana, alcançando resultados muito próximos do enfoque clássico. Os resultados obtidos foram satisfatórios para as variáveis estudadas.

REFERÊNCIAS

CAMBARDELLA, C.A.; MOOMAN, T.B.; NOVAK, J.M.; PARKIN, T.B.; KARLEM, D.L.; TURVO, R.F.; KONOPA, A.E. Field scale variability of soil properties in central Iowa soil. **Soil Sci. Am. J.**, 47:1501-1511,1994.

DRUCK, S.; CARVALHO,M.S.; CÂMARA,G.; MONTEIRO,A.M.V. **Análise espacial de dados geográficos**. Brasília, EMBRAPA, 2004 (ISBN: 85-7383-260-6)

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2011. Disponível em: <http://www.R-project.org>.

RIBEIRO Jr, P.J.; DIGGLE, P.J. geoR: A package for geoestatistical analysis. **R-NEWS**, Vienna, v.1, n.2, p.14-18, 2011.