

Geoestatística - Relatório do final do curso

Djair Durand Ramalho Frade

30 de novembro de 2012

Resumo

Na agricultura de precisão as práticas inferências estabelecem condições ideais as espécies cultivadas na agricultura, seja ela química, física ou biológica, utilizando-se da Geoestatística, definida como a análise de dados de amostras georreferenciadas. Esse método parte da premissa de que cada ponto de amostra é único e procura a correlação entre as amostras vizinhas, sendo assim, os métodos geoestatísticos (análise de semivariograma e krigagem) estão sendo utilizados para analisar a dependência espacial e para interpolar atributos de solo através da krigagem. O presente trabalho teve por objetivo modelar geoestatisticamente a variabilidade espacial das propriedades físico-químicas do solo. O composto analisado foi Matéria Orgânica (MO). Foram ajustados modelos de semivariograma aos dados, bem como a estimação dos parâmetros desse modelo. Os modelos escolhidos foram os que obtiveram o menor valor do Critério de Akaike (AIC). A partir dos resultados obtidos, foram construídos mapas de probabilidade fazendo uso do algoritmo de krigagem, em que foi possível concluir que na área estudada, existem regiões onde há a probabilidade da concentração dos compostos ser maior que o limite especificado.

Palavras-Chave: função de correlação modelo de semivariograma, krigagem.

1 Introdução

Com a globalização da economia surge a necessidade no setor agrícola de buscar mais eficiência e melhor controle de campo. Para a agricultura atual, o desafio é aumentar a produção e conservar recursos naturais, por estes motivos, novos métodos e técnicas de produção estão sendo desenvolvidos. Com a evolução da informática e o uso de computadores e satélites, a forma de investigar a agricultura está mudando, o que oferece ao homem informações de como utilizar os recursos do meio ambiente racionalmente. Essa forma de investigar, que considera a posição das amostras relacionadas às propriedades do solo e da produtividade, é conhecida como Agricultura de precisão (FARACO, 2006).

Segundo Balastreire (1998), a Agricultura de precisão é definida como um conjunto de técnicas que permite o gerenciamento localizado de culturas e que se fundamenta na percepção da variabilidade espacial da produtividade e de fatos a ela relacionados. Webster e Oliviera (1990) afirmam que muitas propriedades do solo variam continuamente no espaço e, conseqüentemente, os valores em locais mais próximos entre si tendem a ser mais semelhantes que aqueles tomados mais distantes, até uma distância limite, correspondente ao domínio

dessas propriedades. Caso isso ocorra, os dados não podem ser tratados como independentes e um tratamento estatístico mais adequado é necessário.

Segundo Hamlett (1986), as análises estatísticas “tradicionais”, baseadas na independência das observações, tem sido substituídas por análises espaciais, as quais consideram as correlações entre as observações vizinhas. Ao analisar dados de propriedades físico-químicas do solo mediante os métodos estatísticos clássicos são ignoradas as consequências da heterogeneidade espacial sobre a representatividade dos valores médios das amostras. No entanto, os métodos geoestatísticos (análise de semivariogramas e krigagem) estão sendo utilizados para analisar tanto a dependência espacial como para interpolar atributos de solo através da krigagem (SOUZA, 2004).

O semivariograma é a principal ferramenta geoestatística para diagnosticar a presença. E um método de regressão usado em geoestatística para aproximar ou interpolar dados da correlação entre as unidades amostradas. Segundo Ribeiro Jr. (1995), os semivariogramas são preferidos para caracterizar a estrutura de continuidade espacial da característica avaliada por exigirem hipóteses de estacionariedade menos restritivas (hipótese intrínseca). E esperado que observações mais próximas geograficamente tenham comportamento mais semelhantes entre si do que aquelas separadas por maiores distâncias.

A krigagem também fornece estimadores exatos com propriedades de não tendenciosidade e eficiência (SOUZA et al., 2004). Os métodos de krigagem usam a dependência espacial entre amostras vizinhas, expressas no semivariograma, para estimar valores em qualquer posição dentro do campo, sem tendência e com variância mínima, ou seja, são estimadores ótimos muito usados no estudo da distribuição espacial de atributos do solo (VIEIRA et al., 2000). Segundo Bolfe e Gomes (2011) deve-se analisar e planejar o número de coletas de dados, através do custo/benefício, pois um grande número de pontos amostrais aumentam os custos da operação e podem inviabilizar a implantação do processo. Os métodos geoestatísticos de interpolação em especial o da krigagem, apresentam propriedades ótimas de estimativas obtidas a partir de dados esparsos (BOGNOLA et al., 2008), possibilitando a otimização de grades amostrais o que acarreta em economia de tempo e recursos. Dessa forma, o presente trabalho tem por objetivo modelar os compostos do solo por meio de técnicas geoestatísticas, de forma à contribuir na redução dos custos de produção, na diminuição da contaminação da natureza pelos defensivos utilizados e o aumento da produtividade.

2 Material e Método

2.1 Área de estudo

Os dados analisados neste trabalho foram coletado na área experimental localizada a +14°19'17.11" de latitude e -91°3'57.66" de longitude situada na Estação Experimental Camantulul (Figura 1(a)), município de Santa Lucía Cotzumalguapa, Guatemala. Na área do estudo, foi definida uma grade amostral totalizando 49 posições georreferenciadas (Figura 1(b))



Figura 1: Localização geográfica do estudo (a) e Localização dos 49 pontos amostrais (b)

2.2 Análise exploratória

A análise exploratória dos dados é um procedimento de grande importância na análise estatística aplicando-se a toda metodologia que se queira utilizar. O objetivo desta análise é conhecer e resumir a variável em estudo por meio de interpretações gráficas e das medidas média, mediana, variância, desvio padrão, coeficiente de variação, coeficiente de assimetria e curtose (BUSSAB e MORETTIN, 1987).

Utilizou-se a família de transformações Box-Cox para encontrar a transformação mais adequada para alcançar o comportamento Gaussiano. Essa família depende diretamente de um parâmetro, que é o coeficiente de transformação dos dados.

2.3 Análise geoestatística

A dependência espacial foi analisada por meio de ajuste de semivariograma baseado nas pressuposições da hipóteses intrínsecas (VIEIRA, 2000), estimado pela equação (1):

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

em que, $\hat{\gamma}$ é a estimativa da semivariância experimental, obtida pelos valores amostrados $[Z(X_i + h) - Z(X_i)]$, h é a distância entre pontos amostrais e $N(h)$ é o número total de pares de pontos possíveis, dentro da área de amostragem, com a distância h .

Por meio do semivariograma é possível ajustar modelos matemáticos que expressam a estrutura de dependência espacial das características avaliadas, os principais modelos são o linear com patamar, esférico, exponencial e gaussiano. Com os dados obtidos em campo ajustou-se e selecionou-se o melhor modelo de função de semivariância para a variável em estudo, utilizando o Método da Máxima Verossimilhança. O semivariograma experimental e o ajuste dos modelos foram efetuados no sistema R, utilizando o pacote geoR (RIBEIRO JR. e DIGLLE, 2001).

3 Resultados

3.1 Análise descritiva

Inicia-se a análise exploratória com algumas análises descritivas simples. Para essas análises, foi utilizado o *software* estatístico R. Pode-se observar na Tabela 1, que o coeficiente de variação (CV) é maior que 20% o que segundo Gomes (1987) representa uma baixa homogeneidade dos dados. Pode-se observar ainda que não existe assimetria uma vez que o coeficiente de assimetria esta muito próximo de zero, o que pode-se confirmar observado o gráfico inferior direito da Figura 2.

Tabela 1: Estatísticas descritivas: média, mínimo (Min), máximo (Max), desvio padrão (DP), coeficiente de variação (CV), curtose e assimetria para o composto MO

| | Média | Min. | Max | DP | CV | Curtose | Assimetria |
|------------------|----------|------|------|-----------|----------|----------|-------------|
| Matéria Orgânica | 3,797347 | 1,44 | 5,88 | 0,8264956 | 21,76508 | 3,357404 | -0,09955833 |

Observa-se que o gráfico do canto superior esquerdo da Figura 2 representa os dados dos quantis da variável MO com os dados reais. Os símbolos “+”, “ Δ ”, “O” e “ \times ”, nesta ordem, indicam os quartis amostrais, essa imagem condiz com a ideia de que existe padrão espacial do composto MO, uma vez que existem aglomerados das categorias. Esta inspeção não garante tal pressuposto mas serve para mostrar desvios óbvios, caso existam.

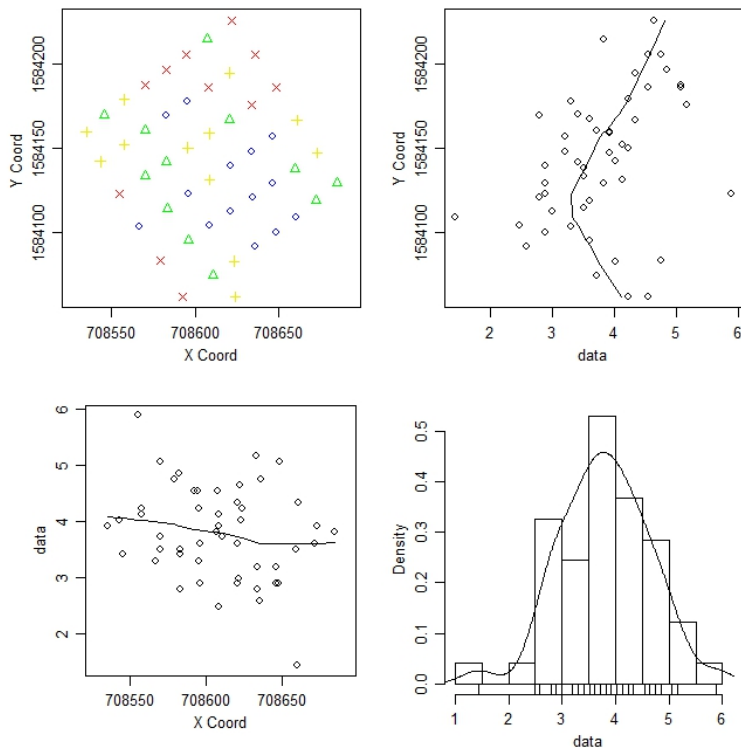


Figura 2: Gráfico descritivo do padrão espacial do composto MO

Para a análise exploratória do pressuposto de normalidade foi usada o gráfico do canto inferior direito da Figura 2, o qual mostra o histograma e a densidade amostral dos dados, desconsiderando o possível padrão espacial.

3.2 Análise geoestatística

A existência de dependência espacial pode ser verificada por meio do diagnóstico gráfico, utilizando-se envelopes simulados e variogramas empíricos (Figura 3) neste caso, foram utilizadas 1000 simulações. Nota-se a existência de pontos fora do envelope simulado, o que segundo Diggle e Ribeiro Jr (2007) indicando uma considerável dependência espacial uma vez que havendo pelo menos um ponto do variograma fora do envelope rejeita-se a hipótese de independência espacial entre as observações.

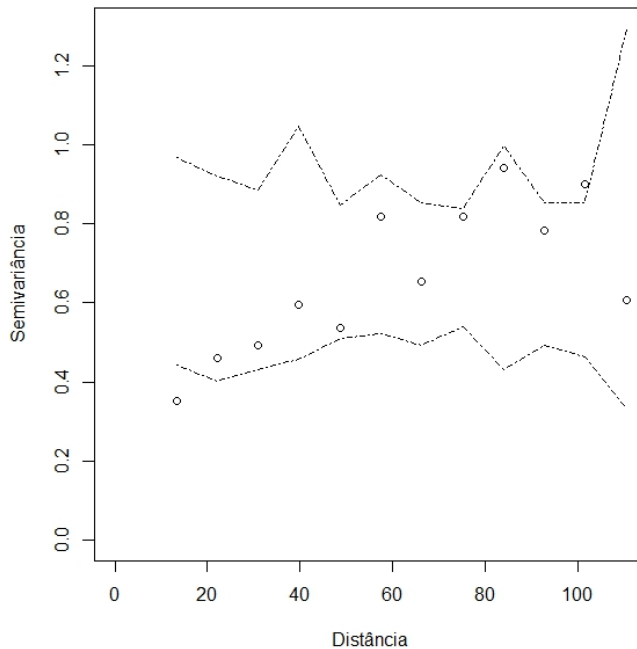


Figura 3: Verificando a dependência espacial da variável MO através do envelope simulado.

A análise variográfica apresentada na Figura 4 apresenta estruturas espaciais bem definidas, ou seja, existe uma função estrutural com semivariância de comportamento modelável. Na Tabela 2 estão apresentados os parâmetros de ajuste dos modelos selecionados calculados pelo método da máxima verossimilhança e os resultados de diferentes critérios de validação de ajuste dos modelos geoestatísticos. Segundo o critério de Akaike, o menor valor de AIC para o composto MO, aponta o modelo esférico.

Tabela 2: Parâmetros dos modelos de semivariograma pelo método da máxima verossimilhança para o composto MO

| Modelo | $\hat{\beta}$ | $\hat{\tau}^2$ | $\hat{\sigma}^2$ | $\hat{\tau}^2 + \hat{\sigma}^2$ | AIC | log-like |
|-------------|---------------|----------------|------------------|---------------------------------|---------|----------|
| Exponential | 4,039 | 0,198 | 0,510 | 0,707 | 115,292 | -53,646 |
| Spherical | 4,074 | 0,284 | 0,456 | 0,740 | 113,635 | -52,818 |
| Normal | 4,127 | 0,375 | 0,414 | 0,789 | 114,210 | -51,738 |

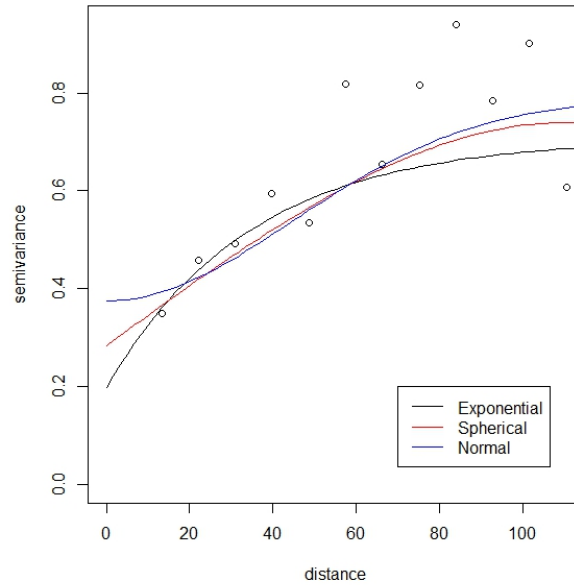


Figura 4: Semivariograma experimental dos dados.

3.3 Krigagem e representação espacial

Para as estimativas de valores em locais não amostrados, foi gerado mapas de distribuição espacial (Figura 5(a)) e os mapas de incertezas (Figura 5(b)), sendo estes obtidos por interpolação por meio da krigagem, a partir dos parâmetros dos modelos ajustados aos semivariogramas.

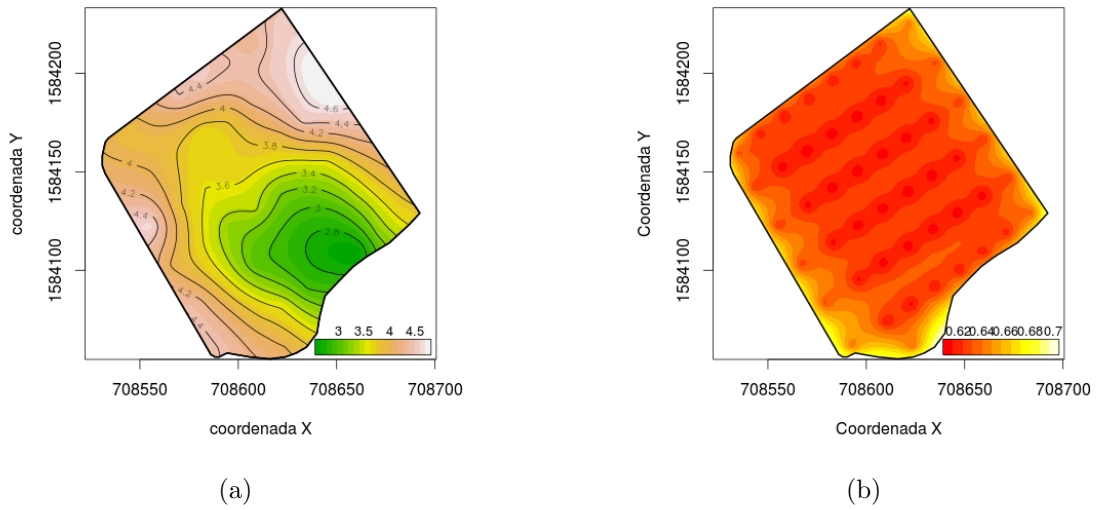


Figura 5: Mapas de valores esperados (a) e mapas dos erros-padrão dos valores esperados (b)

Para a construção de mapas de probabilidades da concentração de MO foi utilizado o critério especificado pelo pesquisador que é de 3,8%. Na Figura 6, observa-se o mapa de probabilidade para predição da concentração de MO, no qual verifica-se uma baixa probabilidade de concentração na região sudeste do mapa.

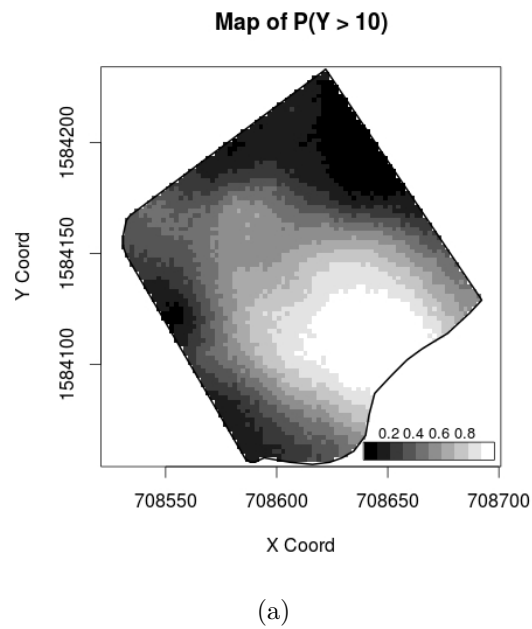


Figura 6: Localização na área com $P[S|y > 3, 8]$.

3.4 Modelagem Baysiana

Quando realizada a análise pela abordagem Bayesiana foram encontrados os resultados reportados na Tabela 3. Observa-se que os valores são próximos das informações obtidas pela abordagem clássica. A proximidade nas estimativas pode ser atribuída a boa escolha das funções de distribuição das prioris para o conjunto de parâmetros do modelo.

Tabela 3: Intervalo de credibilidade para os parâmetros β , σ^2 e ϕ

| Parâmetros | 2,5% | 25% | média | 75% | 97.5% |
|------------|---------|----------|---------|---------|---------|
| β | 3,435 | 4,0160 | 4,290 | 4,576 | 5,166 |
| σ^2 | 0,389 | 0,5578 | 0,992 | 1,299 | 1,846 |
| ϕ | 105,000 | 107,5000 | 110,000 | 113,000 | 115,000 |

4 Conclusão

A partir do semivariograma é possível verificar e modelar a dependência espacial através de funções de correlação que expressam a estrutura de dependência espacial das características avaliadas. As funções de correlação utilizadas nesse trabalho foram: exponencial, gaussiano e o esférica. Esta última função apresentou um melhor ajuste para a variável. A técnica da krigagem é considerada uma boa metodologia de interpolação de dados. Ela utiliza o dado tabular e sua posição geográfica para calcular as interpolações. Utilizando-se o princípio da Primeira Lei de Geografia de Tobler, que diz que unidades de análise mais próximas entre si são mais parecidas do que unidades mais afastadas, a krigagem utiliza funções matemáticas para acrescentar pesos maiores nas posições mais próximas aos pontos amostrais e pesos menores nas posições mais distantes, e criar assim os novos pontos interpolados com base nessas combinações lineares de dados.

Referências

- [1] BALASTREIRE, L.A.; ELIAS, I.A.; AMARAL, J.R. Agricultura de Precisão: Mapeamento da Produtividade da Cultura do Milho. Engenharia Rural, Escola Superior de Agricultura “Luiz de Quiroz”, Universidade de São Paulo, 1998.
- [2] BOGNOLA, I.A.; RIBEIRO JR, P. J.; SILVA, E. A. A.; LINGNAU, C.; HIGA, A. R. Modelagem uni e bivariada da variabilidade espacial de rendimentos de Pinus taeda L. Floresta, Curitiba , v.38, p.373-385, 2008.
- [3] BOLFE, E.L.; GOMES, J.B.V. **Geoestatística como subsídio à - implantação de agricultura de precisão.**[on line]. Disponível em: <<http://www.agronline.com.br/artigos/artigo.php?id=210>>. Acesso em: 03/11/2011.
- [4] DIGGLE, P.J.; RIBEIRO Jr., P.J. **Model-Based geostatistics.** New York: Springer, 2007. 230p.

- [5] FARACO, M.A. **Qualidade do ajuste de modelos Geoestatísticos utilizados na agricultura de precisão**. Cascavel. Dissertação (Mestrado) - Universidade Estadual do Oeste do Paraná, 2006, 135p.
- [6] HAMLETT, J.M.; HORTON, R.; CRESSIE, N.A.C. Resistant and exploratory techniques for use in semivariogram analyses. **Soil Science Society of America Journal**, v.50, p.868-875,1986.
- [7] 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>>. Acesso em 02 jan 2011