

UM ESTUDO COMPARATIVO ENTRE OS PROCEDIMENTOS GEOESTATÍSTICOS DE KRIGEAGEM POR INDICAÇÃO E SIMULAÇÃO ESTOCÁSTICA PARA AMOSTRAS DE ALUMÍNIO

JUSSARA DE OLIVEIRA ORTIZ¹

CARLOS ALBERTO FELGUEIRAS¹

SUZANA DRUCK FUKS²

ANTÔNIO MIGUEL VIEIRA MONTEIRO¹

¹ INPE/DPI - Instituto Nacional de Pesquisas Espaciais

Av. dos Astronautas 1758 Jardim da Granja

CEP 12201-970 São José dos Campos SP Brasil

Fone: (0xx12) 345 6519 Fax: (0xx12) 345 6468

jussara@dpi.inpe.br, carlos@dpi.inpe.br, miguel@dpi.inpe.br

² EMBRAPA/CPAC – Empresa Brasileira de Agropecuária

Br 020 Km 18 Rodovia Brasília Fortaleza

Planaltina Distrito Federal Brasil

Fone: (0xx61) 389 1121 Fax: (0xx61) 389 2953

drucks@ensam.inra.fr

Abstract. This work presents a comparative study between geostatistical methods, the indicator sequential simulation and indicator kriging. Aluminium samples from Santa Catarina State, Brazil, are used to create regular grids in the Geostatistical Library tool (Deutsch and Journel, 1998) and in SPRING, a Geographical Information System environment. This study explores statistics metrics, like standart deviation from estimated values from both methods and from original samples. The results show that indicator simulation better represents the original samples variability.

Keywords: geostatistic, indicator sequential simulation, indicator kriging.

1. Introdução

Em geoprocessamento é usual modelar atributos espaciais, partindo-se do princípio que os valores deste atributo não são conhecidos para qualquer posição do espaço. Os valores não conhecidos são inferidos a partir de um conjunto amostral.

O uso de modelos estocásticos em ciências ambientais, para solucionar questões relacionadas à variabilidade, inerente a um ou mais atributos espacialmente distribuídos, é relativamente novo, apresentando-se, segundo Isaaks (2002), como o estado atual da arte em geoestatística.

A simulação de processos ambientais em Sistemas de Informações Geográficas (SIGs), requer a modelagem de atributos espaciais, através de procedimentos inferenciais onde, segundo Felgueiras et al. (2001), o uso da técnica de simulação estocástica não linear conhecida como simulação sequencial por indicação pode ser usado com vantagens para modelagem de atributos espaciais.

Além da simulação estocástica por indicação, outros procedimentos geoestatísticos como a krigeagem por indicação, ou determinísticos, também possibilitam inferências dos atributos, cada qual com suas particularidades.

Os resultados da krigeagem por indicação tendem a apresentar menor variabilidade, já que é um estimador de variância mínima. Já os modelos estocásticos, por apresentarem um maior variância, traduzida em maior variabilidade dos dados, tendem a reproduzir com mais fidelidade o comportamento dos dados originalmente amostrados.

No contexto deste trabalho, o estudo da variabilidade do atributo é considerado importante, de modo que, além dos resultados obtidos com a simulação não linear, também foram gerados resultados com o procedimento de krigeagem por indicação.

O objetivo deste trabalho, portanto, é efetuar um estudo comparativo entre os procedimentos geoestatísticos de krigeagem por indicação e simulação estocástica por

indicação, através de um estudo de caso para dados de alumínio, no Estado de Santa Catarina (RS), considerados importantes para estudos de fertilidade solo.

Os resultados do trabalho são apresentados em mapas, que permitiram uma análise qualitativa dos atributos espacialmente distribuídos, e também mapas de incertezas, relacionadas ao processo de espacialização, o que permitiu a comparação em termos do desvio padrão obtido por cada processo.

2. Fundamentação Teórica

O procedimento de espacialização consiste em inferir valores do atributo em posições não amostradas, a partir de valores observados. Os algoritmos de inferência de atributos de dados espaciais, segundo Felgueiras (1999), podem ser classificados em dois tipos: interpoladores determinísticos e interpoladores estocásticos. Serão abordados aqui apenas os do segundo tipo.

Os interpoladores estocásticos utilizam ferramentas da geoestatística para inferências de valores dos atributos com estimativas de incertezas. A geoestatística considera que a distribuição espacial do atributo define uma função aleatória, FA, dentro de uma região de interesse A.

Para cada posição não amostrada u , o valor do atributo z em estudo, é considerado uma realização de uma variável aleatória, VA Z , cuja distribuição de probabilidade modela a incerteza local a respeito de z . A função de distribuição acumulada condicionada a (n) amostras, $fdac$, de uma VA contínua, descreve, o comportamento de uma VA $Z(u)$ e é denotada por:

$$F(\mathbf{u}; z / (n)) = Prob \{Z(\mathbf{u}) \leq z / (n)\}$$

2.1. Como a $fdac$ univariada de uma VA numérica é determinada

Uma aproximação da $fdac$, ou o modelo de distribuição de uma VA pode ser obtida utilizando-se procedimentos geoestatísticos não paramétricos. Neste caso, a $fdac$ é obtida por um conjunto de valores estimados, que representam uma aproximação discretizada do modelo de distribuição. Serão abordadas duas metodologias utilizando o enfoque geoestatístico por indicação: a krigagem por indicação e a simulação seqüencial por indicação.

- ✓ *krigagem por indicação* – Deve ser utilizada quando não se pode estabelecer uma função de distribuição a priori, para uma VA. Os valores do atributo são transformados, segundo uma função não linear, a codificação por indicação. A codificação por indicação, sobre um conjunto de dados amostrais numéricos, da VA $Z(u)$, para um valor de corte z_k , gera um conjunto amostral por indicação $I(u; z_k)$ do tipo (Felgueiras, 1999):

$$I(\mathbf{u}; z_k) = \begin{cases} 1, & \text{for } Z(\mathbf{u}) \leq z_k \\ 0, & \text{for } Z(\mathbf{u}) > z_k \end{cases}$$

A esperança condicional da VA numérica por indicação $I(u; z_k)$, fornece uma estimativa da $fdac$ condicionada a (n) amostras, para um valor de corte z_k , e é obtida por:

$$\begin{aligned} E\{I(\mathbf{u}; z_k) / (n)\} &= \\ 1 \cdot Prob\{I(\mathbf{u}; z_k) = 1 / (n)\} + 0 \cdot Prob\{I(\mathbf{u}; z_k) = 0 / (n)\} &= \\ 1 \cdot Prob\{I(\mathbf{u}; z_k) = 1 / (n)\} &= F^*(\mathbf{u}; z_k / (n)) \end{aligned}$$

- ✓ *Simulação por indicação* - Uma simulação estocástica é um processo de construção de realizações alternativas conjuntas, igualmente prováveis, das VAs que compõem um modelo de uma função aleatória. As realizações $\{Z(I) (\mathbf{u}), \mathbf{u} \in A\}$, onde $I = 1, 2, \dots, L$, representam L imagens possíveis da distribuição espacial dos valores de um atributo Z sobre a área A (Felgueiras, 1999).

A simulação seqüencial condicionada, apresentada por Deutsch e Journel (1998), usa a função de distribuição acumulada condicionada às VAs mais correlacionadas, para obter valores $z(\mathbf{u})$, onde uma VA Z em cada posição $\mathbf{u} \in A$.

Uma característica importante da simulação condicionada, em relação à krigeagem por indicação, é que o condicionamento considera os dados amostrais originais e também os valores pré-simulados dentro da vizinhança de \mathbf{u} .

A simulação por indicação utiliza os procedimentos de krigeagem por indicação para estimar as *fdac*'s da VA que representa o atributo considerado, e é utilizada também para obter realizações para o atributo.

O conjunto de realizações, obtidas para um determinado nó dos campos aleatórios, pode ser usado para determinar parâmetros estatísticos da *fdac* local na posição do nó. Dada a *fdac* da VA, torna-se possível definir vários intervalos de probabilidade que podem ser usados para medidas associadas à incertezas (Felgueiras et al., 2001).

3. Área de Estudo

Um estudo de caso foi realizado para área do Estado de Santa Catarina, com amostras de teor de alumínio, utilizadas na dissertação de mestrado de Bonisch (2001).

A Figura 1 apresenta a região escolhida, com as amostras de teor de alumínio sobrepostas à grade gerada por inferência do vizinho mais próximo. Pode-se observar a distribuição espacial do atributo, onde os teores mais altos (14.2) aparecem em tons de cinza mais claro, e os mais baixos (0.0) em preto.

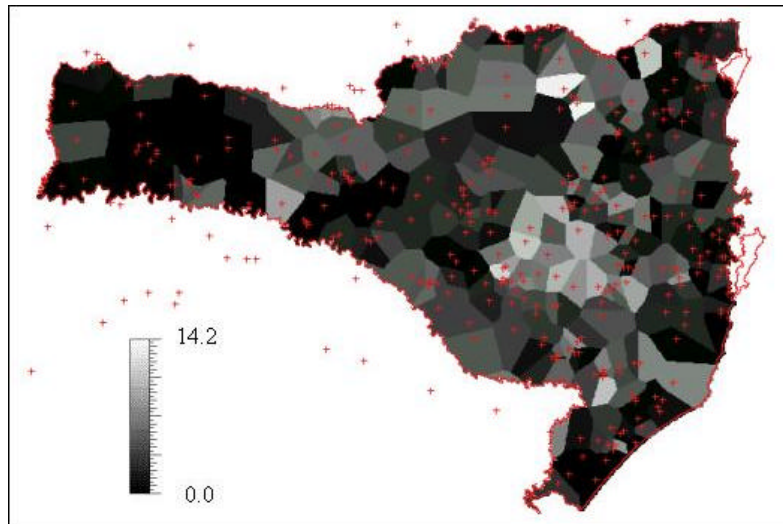


Figura 1 Distribuição das amostras de teor de alumínio superpostas à grade obtida por interpolação por vizinho mais próximo.

4. Resultados e Discussão

Variogramas por indicação foram gerados, no software Spring (SPRING, 2000), para valores de corte determinados por decis: 0.2, 0.6, 1.3, 2.19, 2.9, 3.8, 4.6, 5.4 e 7.10.

Os modelos ajustados no módulo de geoestatística do Spring, forneceram parâmetros para utilizar os programas de geoestatística da GSLIB (Deutsch e Journel, 1998). Foram geradas

grades regulares de 200 linhas x 200 colunas (máximo permitido) com o procedimento de krigagem por indicação e um conjunto de 400 realizações de grade, também de 200 linhas x 200 colunas, utilizando a simulação por indicação, para as amostras de alumínio. Foram geradas também, grades de mesma resolução, com valores do desvio padrão, representando as incertezas para cada procedimento utilizado.

Comparando-se qualitativamente os mapas de média, Figura 2 (a) e (b), obtidos com a krigagem por indicação e com a simulação por indicação, observa-se que:

- a. Em uma análise global, ambos representam os valores do atributo de forma muito similar. Os valores altos e baixos aparecem nas mesmas regiões, coerente com a representação das amostras.
- b. Em uma análise mais detalhada, observa-se que a krigagem por indicação modela o atributo de forma mais homogênea. No mapa de simulação por indicação, pode-se observar pontos com melhor definição de valores mais baixos e também mais altos evidenciando que os valores de média obtidos com a simulação variam de 0 à 13.7. As médias da krigagem variam de 0,15 à 9.99, o que mostra um resultado com menor variabilidade em relação aos valores originais do conjunto amostral que é de 0 à 14.2. Ou ainda, a inferência obtida com a simulação aproxima-se mais do comportamento originalmente amostrado para o alumínio.
- c. Comparando os mapas de desvio padrão (Figura 2 (c e d)), neste trabalho considerados como proporcionais à incerteza das inferências locais, observa-se que estes estão de acordo com os valores estimados. Ou seja, os maiores valores de incerteza estão nas localizações de maior variabilidade de atributo (valores altos próximos a valores baixos)

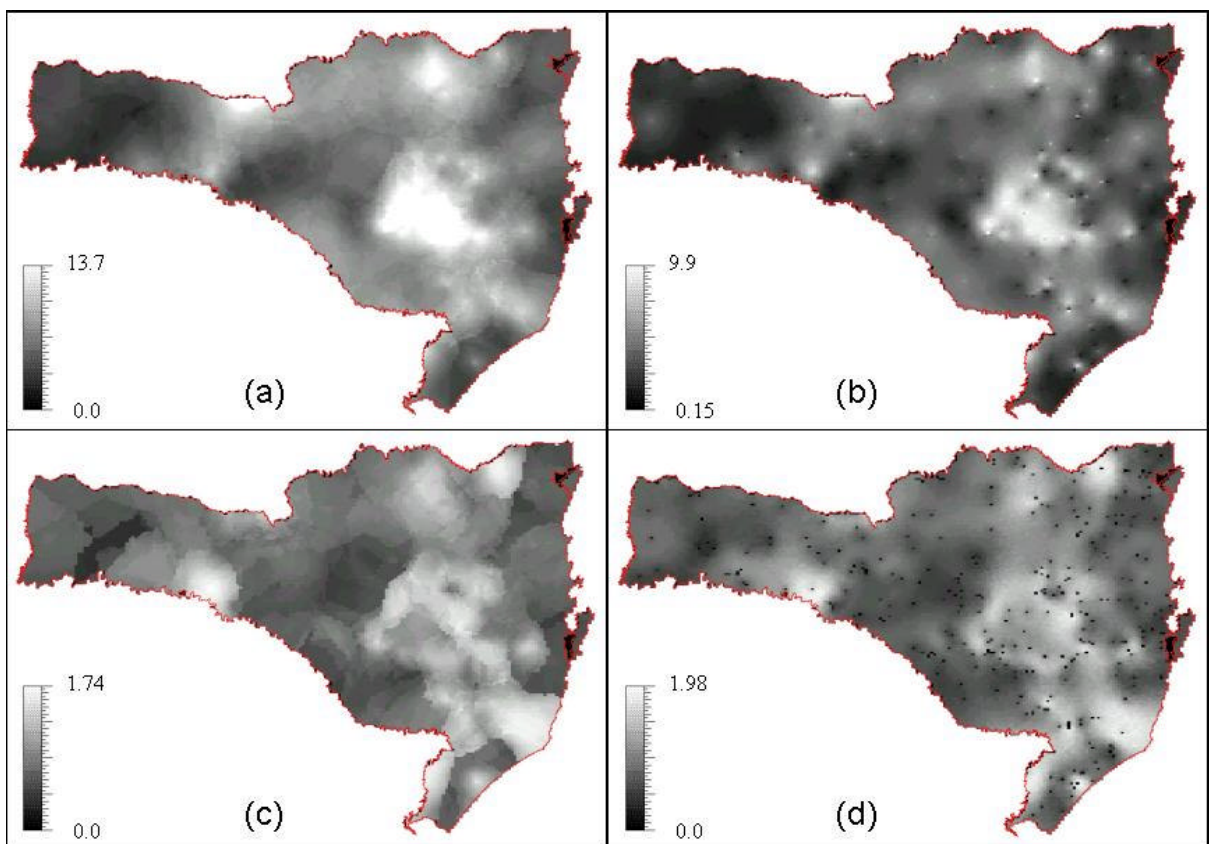


Figura 2 Ilustração dos mapas de: (a) médias krigagem por indicação e (b) médias simulação por indicação; (c) incertezas da krigagem por indicação e (d) incertezas simulação por indicação.

- d. Uma análise visual, baseada na Figura 3, mostra que o histograma dos valores estimados pela simulação está mais próximo do histograma das amostras. Embora os histogramas obtidos pela krigeagem e pela simulação pareçam similares, observa-se pequena assimetria entre os dois. Quantitativamente obteve-se um valor menor de desvio padrão do mapa de krigeagem, 1,74, em relação ao mapa da simulação, com desvio padrão de 1,98. Como o desvio padrão obtido para as amostras é de 2,73, pode-se considerar que a simulação aproximou-se melhor da variabilidade dos dados originais.

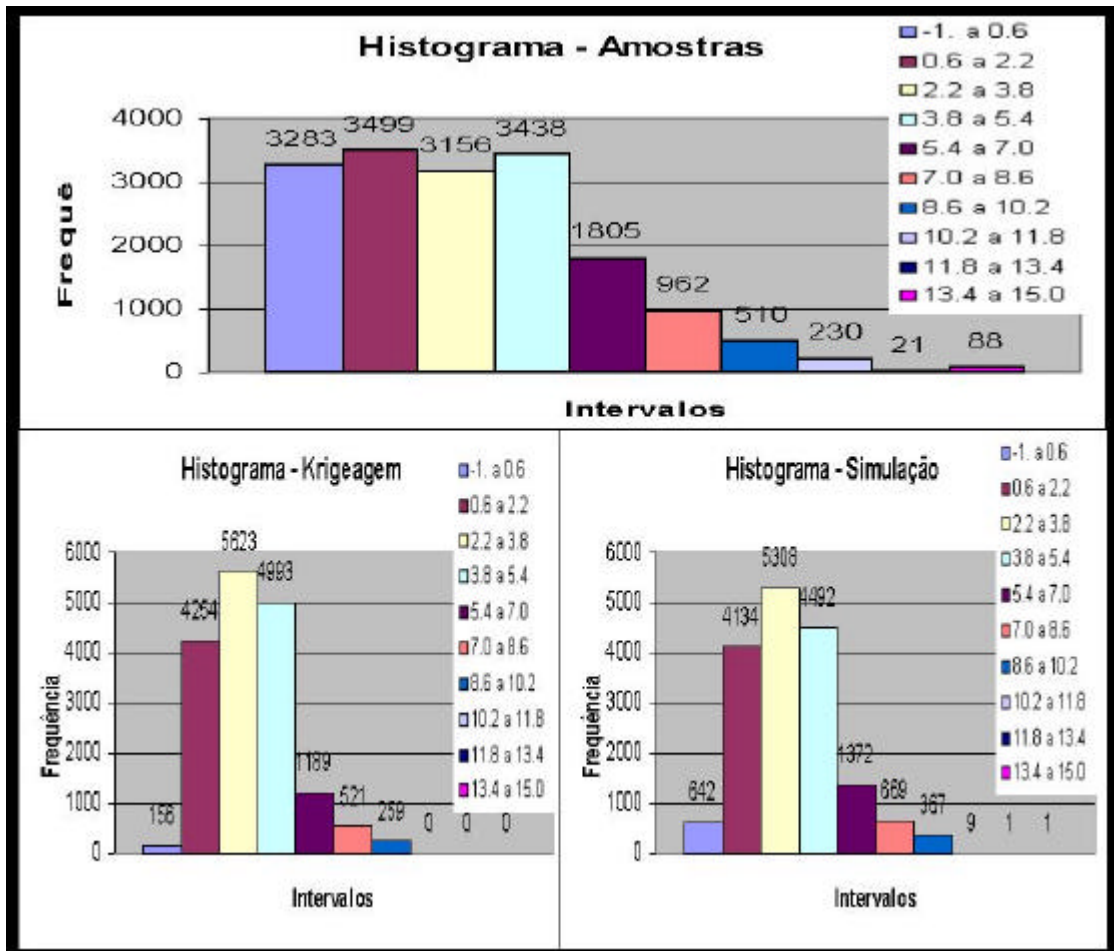


Figura 3 Ilustração dos histogramas: (a) para amostras de teor de alumínio, (b) para os valores estimados por krigeagem por indicação e (c) para os valores estimados por simulação por indicação.

5. Conclusões

A simulação, representou melhor a variabilidade do atributo, pois o desvio padrão dos valores estimados indica melhor aproximação do comportamento originalmente amostrado. Este resultado, confirma a literatura disponível sobre o assunto (Isaaks, 2002), já que a krigeagem é um estimador baseado na minimização da variância, podendo acarretar um efeito de “suavização” da variabilidade original dos dados.

Além disso, as realizações, geradas pela simulação, podem ser usadas em procedimentos de modelagem espacial, através do método de simulação de Monte Carlo (Heuvelink, 1998). Por esse método, realizações que modelam a incerteza do resultado do modelo espacial podem ser geradas, permitindo uma avaliação da propagação das incertezas. A aplicação do método de Monte Carlo caracteriza a continuação deste trabalho.

6. Referências Bibliográficas

Deutsch C. V. and Journel A. G. *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, 1998.

Felgueiras C. A. Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informação Geográfica: O Paradigma Geoestatístico por Indicação. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Publicado em <<http://www.dpi.inpe.br/teses/carlos/>>, 1999.

Felgueiras C. A., Monteiro A. M. V., Fuks S. D. and E. C. G. Camargo. “Uso de Simulação Estocástica não Linear para Inferências de Atributos Espaciais Numéricos” [CD-ROM]. In: X Simpósio Brasileiro de Sensoriamento Remoto, 10., 2001. *Anais. URLib*.

Heuvelink G. B. M. *Error Propagation in Environmental Modeling with GIS*, Bristol, *Taylor and Francis Inc*, 1998.

Isaaks E. H.. *Stochastic Simulation*. Personal Homepage, 2002.

SPRING V.3.4, (DPI/INPE) Sistema de Processamento de Informações Georeferenciadas – Divisão de Processamento de Imagens (DPI) do Instituto Nacional de Pesquisas Espaciais (INPE), <<http://www.dpi.inpe.br/spring/>> , 2000.