

Inferências e Estimativas de Incertezas Utilizando Técnicas de Krigagem Não Linear

Carlos Alberto Felgueiras
Instituto Nacional de Pesquisas Espaciais - INPE/DPI
Av. dos Astronautas 1758 Jardim da Granja
CEP 12201-970 São José dos Campos SP Brasil
Fone: (012) 345 6519 Fax: (012) 345 6468
e-mail: carlos@dpi.inpe.br

Suzana Druck Fuks
Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA/CNPS
Rua Jardim Botânico, 1024 Jardim Botânico
CEP: 22460-000 Rio de Janeiro - RJ - Brasil
Fone: (021) 274 4999 Fax: (021) 274 5291
e-mail: suzana@cnps.embrapa.br

Antonio Miguel Vieira Monteiro
Instituto Nacional de Pesquisas Espaciais - INPE/DPI
Av. dos Astronautas 1758 Jardim da Granja
CEP 12201-970 São José dos Campos SP Brasil
Fone: (012) 345 6474 Fax: (012) 345 6468
e-mail: miguel@dpi.inpe.br

Eduardo Celso Gerbi Camargo
Instituto Nacional de Pesquisas Espaciais - INPE/DPI
Av. dos Astronautas 1758 Jardim da Granja
CEP 12201-970 São José dos Campos SP Brasil
Fone: (012) 345 6524 Fax: (012) 345 6468
e-mail: eduardo@dpi.inpe.br

RESUMO

Propriedades ambientais são dados importantes para modelagens desenvolvidas em ambiente de Sistemas de Informação Geográfica. As técnicas geoestatísticas de inferência, conhecidas por krigagem, consideram os atributos ambientais como variáveis aleatórias o que possibilita a modelagem das incertezas associadas às inferências. Neste contexto, o presente trabalho explora o uso da técnica de krigagem não linear, a *krigagem por indicação*, para inferências a partir de amostras observadas pontualmente. O trabalho apresenta, também, alternativas de cálculos de incertezas para propriedades ambientais de natureza numérica e temática. Por fim, o trabalho mostra e analisa os resultados da aplicação da krigagem por indicação sobre amostras de textura de solo.

ABSTRACT

Environmental properties are very important data for spatial modelling in the Geographical Information Systems context. The geostatistical techniques for estimation, named kriging, consider the environmental attributes as random variables allowing the inference of uncertainties related to the estimated values. This article explores the nonlinear kriging technique, the *indicator kriging*, to estimate values from a set of sample point data. Also, alternatives for uncertainty assessment for numerical and thematic environmental properties are presented. Finally, this works shows and analyzes the results obtained when the indicator kriging approach is used with soil texture samples.

1. Introdução

No contexto de Sistemas de Informação Geográfica – SIGs – muitas propriedades, ou atributos, ambientais são representadas por um conjunto de amostras pontuais. Cada amostra pontual está localizada no espaço geográfico 2D por suas coordenadas (x_i, y_i) e contém o valor z_i da propriedade medida na posição i . Este valor pode representar propriedades numéricas, como altimetria, ou propriedades temáticas, como classes de solo.

As amostras pontuais formam um subconjunto limitado em relação à quantidade de posições e valores possíveis dentro de uma área de interesse. Para se obter valores em posições não amostradas utilizam-se algoritmos inferenciais que se baseiam no conjunto amostral disponível. Os SIGs atuais incorporam vários algoritmos de inferência para espacializar propriedades segundo uma estrutura de grade regularmente distribuída no espaço. Entretanto, raramente essa espacialização é acompanhada de informações sobre a qualidade do dado gerado. Informação sobre a qualidade de um produto é valiosa para apoiar decisões baseadas no mesmo.

As técnicas geoestatísticas de inferência, conhecidas por *krigeagem*, consideram os atributos ambientais como variáveis aleatórias o que possibilita a modelagem das incertezas associadas às inferências. Essas técnicas têm sido utilizadas, com frequência, para espacialização de atributo ambientais amostrados pontualmente (Oliver, 1990 e Camargo, 1997). A *krigeagem* é um estimador de média ponderada local que utiliza o critério de minimização da variância de estimação para definir os pesos associados a cada amostra considerada. Além disso, esse critério permite a criação de um mapa de incertezas de estimação associado ao mapa de atributos espacializados. Estas são as principais vantagens do interpolador de *krigeagem* em relação a outros interpoladores de média ponderada, cujos ponderadores são, em geral, definidos por paradigmas subjetivos. Porém as variâncias de *krigeagem* estão relacionadas a distribuição espacial das amostras e não informam sobre a variabilidade do atributo na região de interesse (Deutsch, 1998).

A técnica de inferência conhecida como *krigeagem por indicação*, que é uma técnica de inferência estatística não linear, permite a estimativa dos valores, e também das incertezas associadas ao atributo, durante o processo de espacialização de uma propriedade amostrada. Essa técnica baseia-se na premissa de que a propriedade em estudo é considerada uma variável aleatória em cada posição não amostrada. As inferências dos valores e das incertezas são obtidos a partir da construção da função de distribuição acumulada, condicionada aos dados amostrados, de cada variável aleatória.

Nesse contexto, os objetivos desse trabalho são: explorar o potencial da *krigeagem por indicação* para espacialização de propriedades ambientais representadas por amostras pontuais; apresentar alternativas para cálculos de incertezas para variáveis aleatórias e; mostrar resultados da utilização da *krigeagem por indicação* para propriedades ambientais de natureza temática.

O presente trabalho contém, após a introdução, uma seção conceitual para formalizar as técnicas de *krigeagem* e *krigeagem por indicação*. Nessa seção também são apresentadas alternativas para cálculo de incertezas baseadas na distribuição dos valores de probabilidades associados à variáveis aleatórias numéricas e temáticas. Na sequência apresenta-se um exemplo onde se explora a técnica de *krigeagem por indicação* sobre um conjunto amostral representativo de classes de textura de solo. Finalmente o trabalho contém uma seção de análise dos resultados obtidos, no exemplo, e uma seção final de conclusões.

2. Conceitos

A geoestatística modela os valores de um atributo, dentro de uma região A da superfície terrestre, como uma função aleatória. Para cada posição $\mathbf{u} \in \mathbf{A}$ o valor do atributo de um dado espacial é modelado como uma *variável aleatória* (VA) $Z(\mathbf{u})$. Isto significa que, na posição \mathbf{u} , a VA $Z(\mathbf{u})$ pode assumir diferentes valores desse atributo, cada valor com uma probabilidade de ocorrência associada. Nas n posições amostradas, \mathbf{u}_α , $\alpha = 1, 2, \dots, n$, os valores $z(\mathbf{u}_\alpha)$ são considerados determinísticos, ou ainda, podem ser considerados VA's cujo valor medido tem uma probabilidade de 100% de ocorrência. A função de distribuição de $Z(\mathbf{u})$ condicionada aos dados amostrado, $F(\mathbf{u}; z|(n))$, é definida por:

$$F(\mathbf{u}; z|(n)) = \text{Prob}\{Z(\mathbf{u}) \leq z|(n)\} \text{ quando o atributo é numérico} \quad (2.1)$$

e

$$F(\mathbf{u}; z|(n)) = \text{Prob}\{Z(\mathbf{u}) = z|(n)\} \text{ quando o atributo é temático} \quad (2.2)$$

A $F(\mathbf{u}; z|(n))$ modela a incerteza sobre os valores de $z(\mathbf{u})$, em posições \mathbf{u} não amostradas, considerando-se as n amostras. Esta seção formaliza os conceitos relacionados aos estimadores geoestatísticos de *krigeagem* e de *krigeagem por indicação*. Também são apresentadas alternativas para o cálculo de incertezas associadas aos atributos inferidos pelo método de *krigeagem por indicação*.

2.1 O Estimador de Krigeagem

A *krigeagem* é uma "coleção de técnicas de regressão linear generalizadas para *minimizar uma variância de estimação* a partir de um modelo de covariância definido a priori", Deutsch, 1998. A *krigeagem* estima um valor de um atributo, em uma posição espacial \mathbf{u} não amostrada, a partir de um conjunto de amostras vizinhas $z(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n$. Esse procedimento é similar àquele usado em interpolação de média móvel ponderada com a diferença que os pesos são derivados de uma análise geoestatística ao invés de um modelo geral e possivelmente inapropriado (Burrough, 1998).

O estimador de *krigeagem simples* é definido por:

$$z_S^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{S\alpha}(\mathbf{u}) \cdot z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^n \lambda_{S\alpha}(\mathbf{u}) \right] \cdot \mu \quad (2.3)$$

onde μ é a média estacionária da função aleatória caracterizada pelas variáveis aleatórias definidas numa região estacionária \mathbf{A} . A hipótese de estacionariedade de segunda ordem implica, também, que a covariância $C(\mathbf{h})$ entre duas variáveis aleatórias só depende do vetor \mathbf{h} determinado pelas localizações espaciais dessas variáveis. Os pesos da equação 2.3, $\lambda_{S\alpha}(\mathbf{u})$ com $\alpha = 1, \dots, n$, são determinados pelos seguintes critérios: o estimador deve ser não tendencioso, ou seja $E\{Z^*(\mathbf{u})\} - E\{Z(\mathbf{u})\} = 0$, e a variância do erro de estimação, $\sigma^2 = \text{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$, é a mínima entre todos os estimadores lineares não tendenciosos. A aplicação dessas premissas possibilita o cálculo dos pesos $\lambda_{S\alpha}(\mathbf{u})$, $\alpha = 1, \dots, n$, pela solução do seguinte sistema de equações de *krigeagem simples*:

$$\sum_{\alpha=1}^n \lambda_{S\alpha}(\mathbf{u}) C(\mathbf{u}_\alpha, \mathbf{u}_\beta) + \phi = C(\mathbf{u}, \mathbf{u}_\beta) \quad \text{para } \beta = 1, \dots, n \quad (2.4)$$

onde: $C(\mathbf{u}_\alpha, \mathbf{u}_\beta)$ é a covariância entre as amostras observadas em \mathbf{u}_α e \mathbf{u}_β ; $C(\mathbf{u}, \mathbf{u}_\beta)$ é a covariância entre a amostra observada em \mathbf{u}_β e o ponto não amostrado \mathbf{u} ; ϕ é o multiplicador de Lagrange necessário para a minimização da variância. As covariâncias são calculadas a partir do *modelo de variograma teórico*, $2\gamma(\mathbf{h})$, ajustado sobre o *variograma experimental* que é determinado, a partir do conjunto amostral, por:

$$2\gamma(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(\alpha, \beta) \in \Omega_{\mathbf{h}}} (z(\mathbf{u}_\alpha) - z(\mathbf{u}_\beta))^2 \quad (2.5)$$

onde $N(\mathbf{h})$ é o número de pares de amostras separadas pelo vetor \mathbf{h} . Sob a hipótese de estacionariedade em \mathbf{A} , a seguinte relação é válida (Isaaks, 1989):

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad (2.6)$$

Uma desvantagem do estimador de *krigeagem simples* é a necessidade de se conhecer a priori a média estacionária da região \mathbf{A} . Uma alternativa é o estimador de *krigeagem ordinária* que possibilita a inferência do atributo, numa posição \mathbf{u} não amostrada, sem a necessidade de se conhecer a média estacionária μ . Sob a condição de que a somatória dos ponderadores da *krigeagem ordinária* $\lambda_{O\alpha}(\mathbf{u})$ é igual a 1, ou seja:

$$1 - \sum_{\alpha=1}^n \lambda_{O\alpha}(\mathbf{u}) = 0 \quad (2.7)$$

obtêm-se a seguinte formulação para o estimador de *krigeagem ordinária*:

$$z_O^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{O\alpha}(\mathbf{u}) \cdot z(\mathbf{u}_\alpha) \quad (2.8)$$

Pode-se mostrar que a formulação do estimador de *krigeagem ordinária* é equivalente a uma *krigeagem simples* com médias estimadas localmente a partir das amostras vizinhas, Deutsch, 1998. A substituição de uma única média estacionária por médias locais, ou tendências locais, faz do estimador de *krigeagem ordinária* uma opção bem mais interessante do que o estimador de *krigeagem simples*. Existem outros tipos de *krigeagem linear*, tais como a *krigeagem*

universal, a krigeagem disjuntiva, a krigeagem probabilística, etc., porém estão fora do escopo deste trabalho.

A variância de krigeagem ordinária σ_O^2 é dada por (Isaaks, 1989):

$$\sigma_O^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{O\alpha}(\mathbf{u}) \lambda_{O\beta}(\mathbf{u}) C(\mathbf{u}_\alpha, \mathbf{u}_\beta) - 2 \sum_{\alpha=1}^n \lambda_{O\alpha}(\mathbf{u}) C(\mathbf{u}, \mathbf{u}_\alpha) \quad (2.9)$$

onde $\sigma^2 = C(\mathbf{h}=\mathbf{0})$ é a variância do atributo e determina o quão errático é o atributo. As covariâncias $C(\mathbf{u}_\alpha, \mathbf{u}_\beta)$ consideram aglomerados de amostras e aumentam a variância σ_O^2 sempre que as amostras estão próximas entre si. A variância σ_O^2 diminui com o aumento da covariância $C(\mathbf{u}, \mathbf{u}_\alpha)$, ou seja, com a proximidade da amostra em relação à posição \mathbf{u} .

Observa-se assim que o valor de variância de krigeagem está relacionada a distribuição espacial das amostras, e não depende diretamente dos valores do atributo. Deutsch, 1998, argumenta que "as variâncias de krigeagem, sendo independentes dos dados, fornecem apenas uma comparação entre diferentes configurações geométricas de dados".

Portanto a variância de krigeagem não pode ser usada como medida de incerteza do atributo em análise a não ser que o atributo possa ser modelado por uma função aleatória multigaussiana. Essa premissa é difícil de ser verificada e raramente ocorre para atributos ambientais. Na seção seguinte descreve-se a krigeagem por indicação que possibilita a estimativa de incertezas sem a necessidade de se definir um modelo de distribuição de probabilidade a priori.

2.2 A Krigeagem por Indicação

Como já citado no início da seção 2, a função de distribuição acumulada, $F(\mathbf{u}; z|(n))$, modela a incerteza sobre os valores de $z(\mathbf{u})$, em posições \mathbf{u} não amostradas, considerando-se as n amostras. Essa função pode ser inferida a partir do procedimento de inferência chamado de *krigeagem por indicação*. Ela é uma técnica de inferência estatística não linear pois é aplicada sobre os valores do atributo transformados por um mapeamento não linear, a *codificação por indicação*.

A codificação por indicação da VA $Z(\mathbf{u}=\mathbf{u}_\alpha)$, em um valor de corte $z = z_k$, gera a VA $I(\mathbf{u}=\mathbf{u}_\alpha; z_k)$ utilizando a seguinte função de mapeamento não linear:

$$I(\mathbf{u}_\alpha; z_k) = \begin{cases} 1, & \text{se } Z(\mathbf{u}_\alpha) \leq z_k \\ 0, & \text{se } Z(\mathbf{u}_\alpha) > z_k \end{cases} \quad \text{para atributos numéricos e,} \quad (2.10)$$

$$I(\mathbf{u}_\alpha; z_k) = \begin{cases} 1, & \text{se } Z(\mathbf{u}_\alpha) = z_k \\ 0, & \text{se } Z(\mathbf{u}_\alpha) \neq z_k \end{cases} \quad \text{para atributos temáticos} \quad (2.11)$$

Os valores de corte, $z_k, k=1,2,\dots,K$, são definidos em função do número de amostras. É necessário que a quantidade de amostras codificadas com valor 1 seja suficiente para se definir, com sucesso, um modelo de variografia para cada valor de corte (Journel, 1983).

A esperança condicional da VA por indicação $I(\mathbf{u}; z_k)$ é calculada por:

$$\begin{aligned} E\{I(\mathbf{u}; z_k) | (n)\} &= 1 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 1 | (n)\} + 0 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 0 | (n)\} \\ &= 1 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 1 | (n)\} = F^*(\mathbf{u}; z_k | (n)) \end{aligned} \quad (2.12)$$

A equação acima representa um resultado muito importante no que diz respeito a inferência da distribuição de probabilidade de uma variável aleatória: "A esperança condicional de $I(\mathbf{u}; z_k)$ fornece, para o valor de corte $z = z_k$, uma estimativa do valor da função de distribuição condicionada, fdc, de $Z(\mathbf{u})$ no caso de atributos temáticos e uma estimativa da função de distribuição acumulada condicionada, fdac, para atributos numéricos".

A *krigeagem por indicação simples* é um procedimento de krigeagem linear simples aplicado ao conjunto amostral codificado por indicação em $z = z_k$, ou seja:

$$F_S^*(\mathbf{u}; z_k | (n)) = \sum_{\alpha=1}^n \lambda_{S\alpha}(\mathbf{u}; z_k) i(\mathbf{u}_\alpha; z_k) + \left[1 - \sum_{\alpha=1}^n \lambda_{S\alpha}(\mathbf{u}; z_k) \right] F^*(z_k) \quad (2.13)$$

onde $F_S^*(z_k)$ é a média da função aleatória da região estacionária e os pesos $\lambda_{S\alpha}(\mathbf{u}; z_k)$ são determinados com o objetivo de minimizar a variância do erro de estimação.

Considerando-se a somatória dos pesos igual a 1 obtêm-se uma variante mais simplificada da krigeagem por indicação simples, a *krigeagem por indicação ordinária*, cuja expressão de estimação se resume a:

$$F_O^*(\mathbf{u}; z_k | (n)) = \sum_{\alpha=1}^n \lambda_{O\alpha}(\mathbf{u}; z_k) i(\mathbf{u}; z_k) \quad (2.14)$$

Os pesos $\lambda_{O\alpha}(\mathbf{u}; z_k)$ são obtidos solucionando-se o seguinte sistema de equações krigeagem por indicação ordinária:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{O\beta}(\mathbf{u}; z_k) C_I(\mathbf{h}_{\alpha\beta}; z_k) + \mu(\mathbf{u}; z_k) = C_I(\mathbf{h}_\alpha; z_k) & \forall \alpha = 1, 2, \dots, n \\ \sum_{\beta=1}^n \lambda_{O\beta}(\mathbf{u}; z_k) = 1 \end{cases} \quad (2.15)$$

onde $\mu(\mathbf{u}; z_k)$ é um parâmetro de Lagrange, $\mathbf{h}_{\alpha\beta}$ é o vetor definido entre posições \mathbf{u}_α e \mathbf{u}_β , \mathbf{h}_α é o vetor definido entre posições \mathbf{u}_α e \mathbf{u} , $C_I(\mathbf{h}_{\alpha\beta}; z_k)$ é a autocovariância definida por $\mathbf{h}_{\alpha\beta}$ e $C_I(\mathbf{h}_\alpha; z_k)$ é a autocovariância definida por \mathbf{h}_α . As autocovariâncias são determinadas pelo modelo de variografia teórico definido pelo conjunto I quando $z = z_k$.

A krigeagem por indicação, simples ou ordinária, fornece, para cada valor k de corte, uma estimativa que é também a melhor estimativa mínima quadrática da esperança condicional da VA $I(\mathbf{u}; z_k)$. Utilizando esta propriedade pode-se calcular estimativas dos valores da fdc de $Z(\mathbf{u})$ para vários valores de z_k , pertencentes ao domínio de $Z(\mathbf{u})$. O conjunto dos valores estimados para das fdc's de $Z(\mathbf{u})$, nos valores de corte, é considerado uma *aproximação discretizada* da fdc real de $Z(\mathbf{u})$. Quanto maior a quantidade de valores de corte melhor é a aproximação.

A krigeagem por indicação é não paramétrica. Não considera nenhum tipo de distribuição de probabilidade a priori para a variável aleatória. Ao invés disso, ela possibilita a construção de uma aproximação discretizada da fdc de $Z(\mathbf{u})$. Os valores de probabilidades discretizados podem ser usados diretamente para se estimar valores característicos da distribuição, tais como: valor médio, variância, moda, quantis e outros.

2.3 Medidas de Incerteza para Variáveis Aleatórias

Como já visto, os procedimentos de krigeagem por indicação possibilitam a estimativa de uma aproximação da fdc, quando $Z(\mathbf{u})$ representa atributos temáticos, ou fdac, quando $Z(\mathbf{u})$ representa atributos numéricos, de uma VA numa determinada localização espacial \mathbf{u} .

Para variáveis aleatórias temáticas estimam-se as probabilidades univariadas $p_j(\mathbf{u})$, $j=1, \dots, L$, das L classes definidas no domínio da VA. Essas probabilidades são utilizadas para inferência do valor, ou da classe, e da incerteza da VA na localização \mathbf{u} não amostrada. O valor estimado para a VA temática é a classe de maior probabilidade. A incerteza em \mathbf{u} , $Inc(\mathbf{u})$, pode ser determinada pelo valor de máxima probabilidade, *incerteza por moda*, ou por uma combinação dos valores de probabilidades das L classes, *incerteza por entropia*.

Dado o conjunto de probabilidades, $p_j(\mathbf{u})$ com $j=1, \dots, L$, a *incerteza por moda* é determinada por:

$$Inc(\mathbf{u}) = 1 - p_{jmax}(\mathbf{u}) \quad (2.16)$$

onde $p_{jmax}(\mathbf{u})$ é a probabilidade da classe mais provável da função de distribuição de probabilidade em \mathbf{u} , ou seja, $p_{jmax}(\mathbf{u}) = \text{Max}(p_j(\mathbf{u}))$ $j=1, \dots, L$,

A *incerteza por entropia* é calculada a partir da medida de entropia proposta por Shannon, 1948. Informalmente, a entropia pode ser entendida com uma medida relacionada a organização espacial de um atributo. Ela mede a confusão, ou a desordem, relacionada aos valores, ou possíveis estados, associados a um atributo. Quando a VA é temática, ou resultado da discretização de uma VA numérica, (Journal, 1993), a entropia de Shannon, de sua distribuição univariada, tem a seguinte formulação:

$$Inc(\mathbf{u}) = H = -\sum_{j=1}^L p_j(\mathbf{u}) \ln(p_j(\mathbf{u})) \quad (2.17)$$

Para variáveis aleatórias numéricas estimam-se as probabilidades acumuladas univariadas $p_k(\mathbf{u})$, $k=1, \dots, K$, dos K valores de corte. Essas probabilidades são utilizadas para inferência do valor e da incerteza da VA na localização \mathbf{u} não amostrada. O valor estimado para a VA numérica pode ser o valor médio ou a mediana da distribuição. A mediana é um estimador mais robusto quando a distribuição tem alto grau de assimetria. A incerteza, $Inc(\mathbf{u})$, pode ser determinada pela variância $\sigma^2 = E\{(Z(\mathbf{u}) - E\{Z\})^2\}$. A variância pode ser usada para se definir intervalos de confiança do tipo:

$$Prob\{Z(\mathbf{u}) \in [z^*(\mathbf{u}) \pm 2\sigma(\mathbf{u})]\} \cong 0.95 \quad (2.18)$$

quando a variável apresenta um nível de simetria que possibilita supor a hipótese de normalidade. Para distribuições altamente assimétricas, uma medida mais robusta é o intervalo interquartil definido como a diferença entre o maior e o menor quartil, onde

$$Q = z_{0.75}^*(\mathbf{u}) - z_{0.25}^*(\mathbf{u}) \text{ onde } z_p^* = F^{*-1}(\mathbf{u}, p | (n))$$

3. Exemplo

Esta seção apresenta resultados da aplicação da krigeagem por indicação para dados amostrais de natureza temática. A área de estudo pertence à fazenda Canchim, base física do Centro de Pesquisa Pecuária do Sudeste (SPPSE – EMBRAPA), localizada no município de São Carlos, estado de São Paulo, Brasil. Essa região cobre uma área de 2660 hectares, entre as coordenadas de 21°55'00" à 21°59'00", latitudes sul, e 47°48'00" à 41°52'00", longitudes oeste. Detalhes das características da região estão descritos em Camargo, 1997.

3.1 Um Exemplo de Inferência para Variáveis Temáticas

Da área de estudo obteve-se um conjunto de amostras de textura do solo (Calderano, 1996) classificados como: solo arenoso, solo médio argiloso, solo argiloso e solo muito argiloso. A Figura 3.1 mostra, no mapa à esquerda, os limites da fazenda Canchim e a distribuição espacial das amostras de textura do solo com seus respectivos valores.

As amostras de textura foram analisadas, utilizando-se o módulo de análise geoestatístico do sistema de informação geográfica SPRING (INPE/DPI, 1999, Câmara, 1996 e Camargo, 1997), com o objetivo de se obter o modelo de variografia representativo do atributo textura do solo. Essa análise mostrou um comportamento anisotrópico para a textura do solo nessa região. A partir das frequências das classes no conjunto amostral, as probabilidades globais, a priori, das classes de textura foram definidas como: $P(1) = 0.2$, $P(2) = 0.34$, $P(3) = 0.38$ e $P(4) = 0.08$.

Foram definidos 4 conjuntos de variáveis por indicação referentes às 4 classes de textura do solo. Para cada um desses conjuntos definiu-se um variograma experimental. Os parâmetros dos modelos de variografia teóricos, ajustados para os variogramas experimentais, estão apresentados na tabela 3.1.

Utilizando-se o conjunto amostral, os modelos teóricos de variografia e o programa de krigeagem por indicação da GSLIB, ik3d.exe, geraram-se grades regulares de 200 linhas por 200 colunas com resolução horizontal, dx , igual a 35 metros e resolução vertical, dy , igual a 50 metros. O retângulo envolvente, em coordenadas UTM e datum SAD69, da área em estudo foi definido como: $xmin = 204000.0$, $xmax = 211000.0$, $ymin = 7565000.0$ e $ymax = 7575000.0$.

Classe	Estrutura	Ângulo	Efeito pepita	Contribuição	Alcance máximo	Alcance mínimo
1	Esférica	135°	0.07	0.126	1795	1380
2	Esférica	135°	0.08	0.06	1753	ϵ

		135 °		0.09	1753	919
3	Esférica	0°	0.092	0.098	3899	ϵ
		0°		0.07	3899	1835
4	Esférica	0°	0.015	0.02	1072	ϵ
	Esférica	90°		0.05	2517	1072

Tabela 3.1: Parâmetros de variografia para as classes de textura do solo ($\epsilon \ll 0.0001$)

As grades regulares foram transferidas para o SPRING que possibilitou a criação dos mapas de valores inferidos e incertezas apresentados na figuras apresentadas nesta seção. O mapas da Figura 3.1 mostram resultados da inferência das classes de textura com interpolador de vizinho mais próximo e com interpolador de krigeagem por indicação (mapa à direita). O mapa inferido por krigeagem foi obtido definindo-se um raio máximo de busca igual a 2000 metros e um número máximo de pontos condicionantes igual a 12. A Figura 3.2 mostra os mapas de incerteza por moda e por entropia referentes ao mapa inferido por krigeagem apresentado na Figura 3.1. As Figuras 3.3 e 3.4 apresentam mapas de valores inferidos e incertezas associadas das classes de textura gerados com 4 e 16 pontos condicionantes.

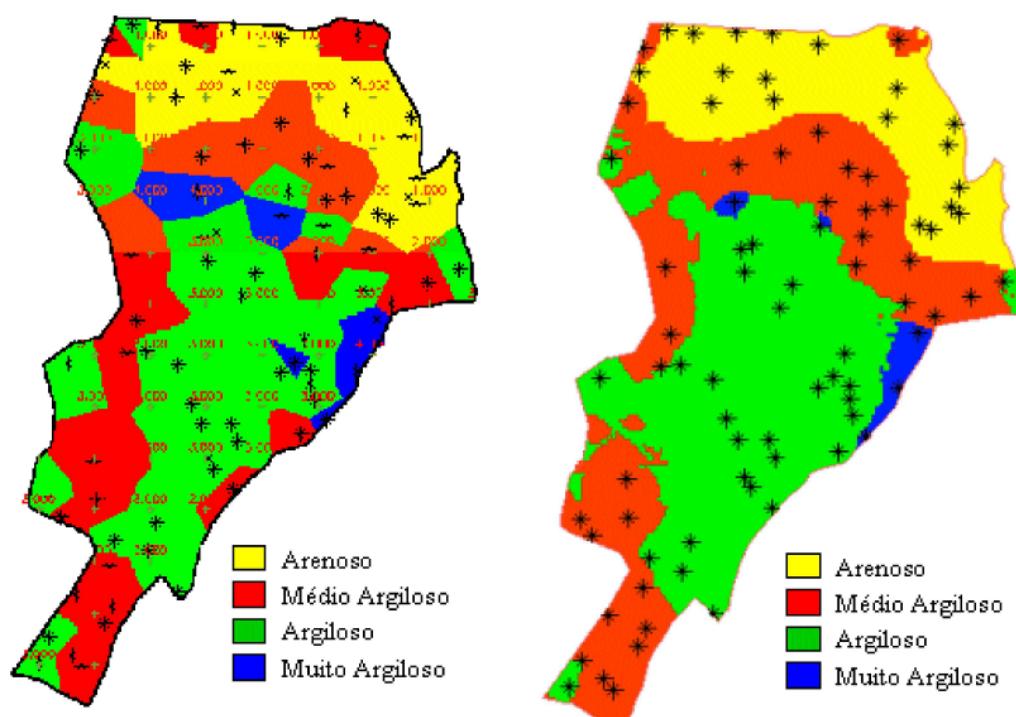


Figura 3.1: Mapa de valores inferidos por vizinho mais próximo (mapa à esquerda) e por krigeagem por indicação utilizando 12 pontos condicionantes (mapa à direita) para a região da Fazenda Canchim

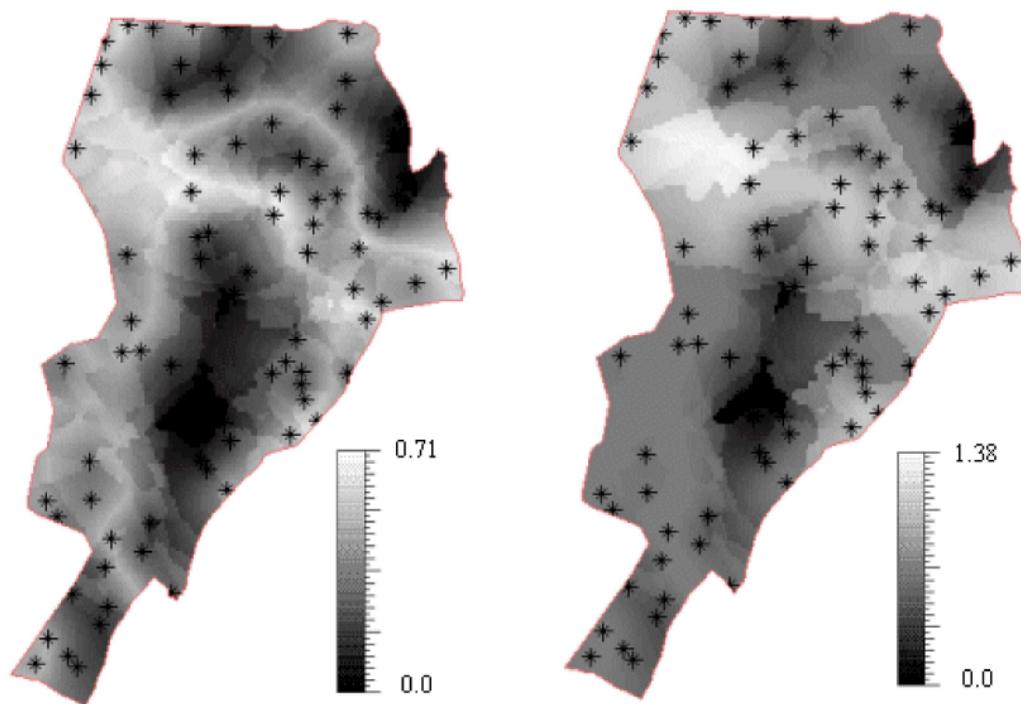


Figura 3.2: Mapas de incerteza por moda (mapa à esquerda) e por entropia de Shannon (mapa à direita) para o mapa de textura inferido por krigeagem por indicação com 12 pontos condicionantes.

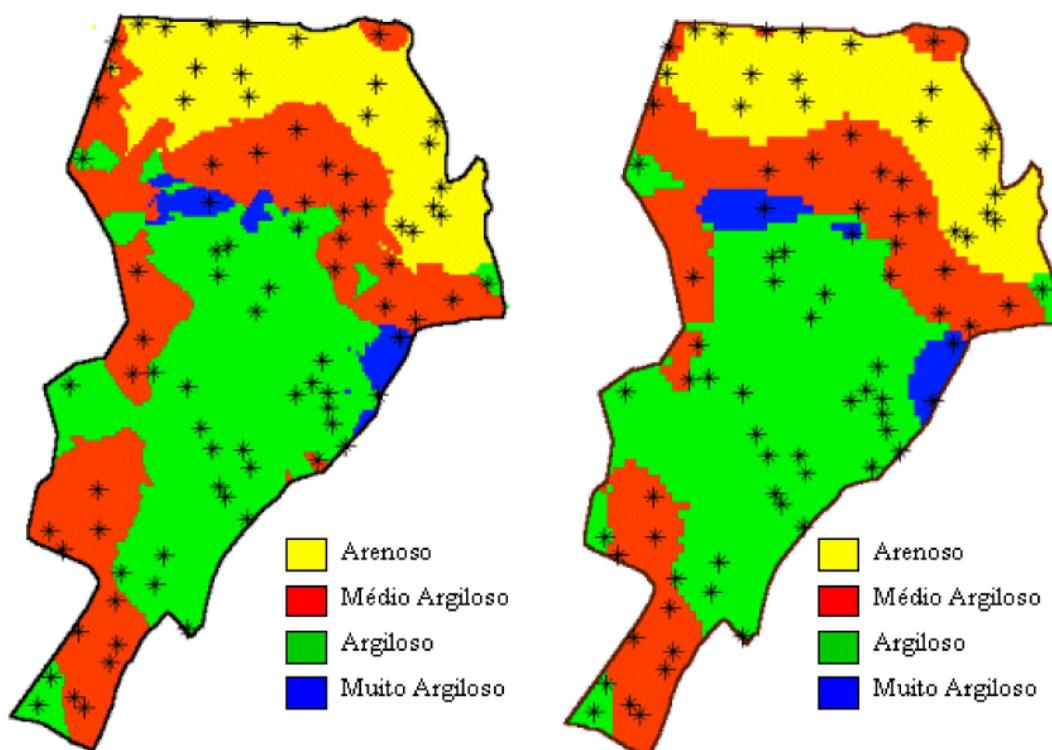


Figura 3.3: Mapas de valores de textura de solo gerados por procedimento de krigeagem por indicação com 4 (mapa à esquerda) e 16 (mapa à direita) pontos condicionantes.

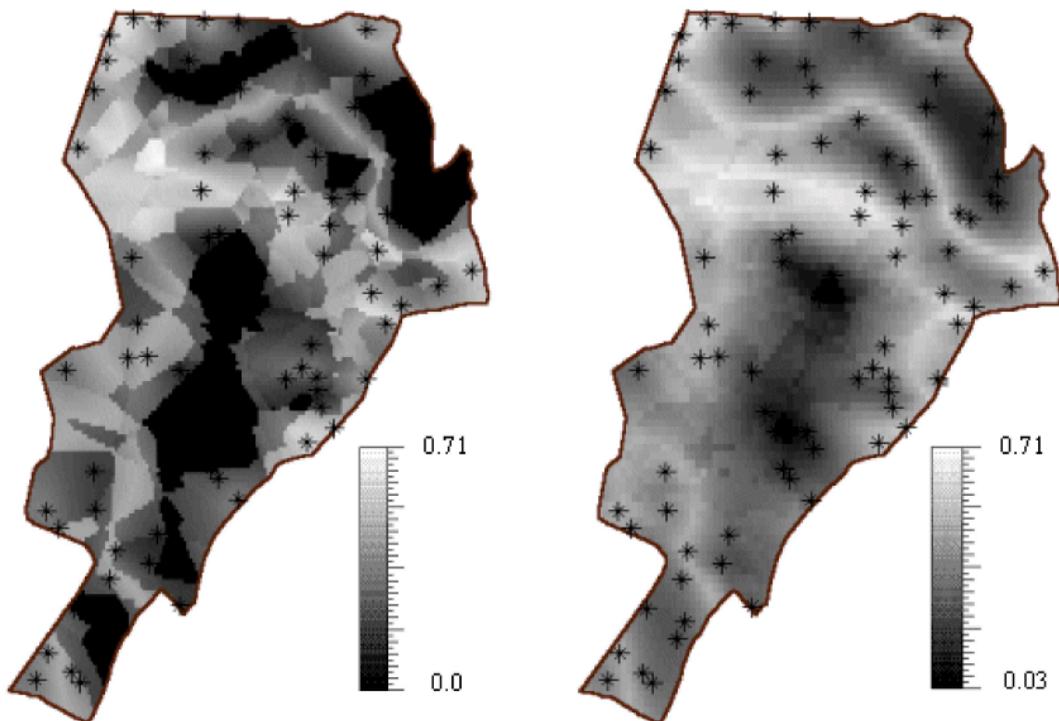


Figura 3.4: Mapas de incerteza por moda para os mapas de textura de solo gerados por procedimento de krigeagem por indicação com 4 (mapa à esquerda) e 16 (mapa à direita) pontos condicionantes.

A tabela 3.2 apresenta os valores mínimo, máximo, médio e a variância dos mapas de incerteza, por moda e por entropia, para valores máximos de pontos condicionantes iguais a 4, 8, 12 e 16.

Máximo de PC's	Incerteza por moda				Incerteza por Entropia			
	Mínima	Máxima	Média	Variância	Mínima	Máxima	Média	Variância
4	0.0	0.729	0.294	0.038	0.0	1.380	0.556	0.120
8	0.0	0.703	0.350	0.024	0.0	1.366	0.716	0.067
12	0.030	0.697	0.372	0.020	0.133	1.340	0.770	0.056
16	0.019	0.690	0.380	0.018	0.097	1.349	0.798	0.050

Tabela 3.2: Valores de incertezas por moda e por entropia obtidos pelo procedimento de krigeagem por indicação para 4, 8, 12 e 16 pontos condicionantes

Os resultados, mapas e tabelas, apresentados nesta seção serão analisados na seção 3.2.

3.2 Análise dos Resultados

Os mapas de valores estimados por krigeagem por indicação, apresentados na seção 3.1, mostram que pode-se aplicar com sucesso esse procedimento estocástico de inferência para espacializar atributos de natureza temática. Essa técnica é equivalente a uma classificação que utiliza apenas amostras pontuais como informação a priori. Isto não é um resultado trivial já que os estimadores determinísticos baseados em médias locais das amostras só podem ser aplicados à atributos numéricos.

A Figura 3.2 ilustra as diferenças existentes entre o mapa de incertezas por moda e o mapa de incertezas por entropia. As diferenças são mais aparentes nas regiões onde várias classes se confundem. Isto já era um resultado esperado, uma vez que, nestas regiões a distribuição de probabilidade das variáveis aleatórias está mais próxima de uma distribuição uniforme. Nessas áreas a incerteza por entropia tem seus valores maximizados. A incerteza por moda mostra um aumento de incerteza a partir do parte central de uma classe em direção as bordas de transição. Os valores máximos de incerteza por moda aparecem nas bordas entre as classes e não tem influência do número de classes próximos as bordas.

As Figuras 3.1 e 3.3 mostram que o estimador de krigeagem por indicação gera mapas de valores inferidos que apresentam diferenças entre si. Isto mostra uma dependência do estimador em relação ao número máximo de pontos condicionantes definidos para cada procedimento. A medida que se aumenta o número de pontos condicionantes o mapa resultante da espacialização tende a ter menor variabilidade do atributo, ou seja, a superfície gerada é mais

suave. Qualitativamente essas diferenças não são muito pronunciadas pois os mapas resultantes mantêm uma certa tendência para cada região classificada.

Comparando-se os mapas de incerteza por moda, apresentados nas figuras 3.2 e 3.4, observa-se que o número de pontos condicionantes também influencia os valores de incertezas estimados. Nos mapas de incertezas as diferenças são bem mais pronunciadas do que nos mapas de valor estimado. A tabela 3.2 apresenta, quantitativamente, esse resultado. Observa-se, desta tabela, que os valores médios da incerteza aumentam com o número de pontos condicionantes. Em contrapartida as variâncias das incertezas diminuem, com o aumento dos pontos condicionantes, caracterizando superfícies de incertezas com variabilidade mais suave.

Conclusões

Este trabalho mostra o potencial da técnica de krigeagem não linear para inferência dos valores e estimativa de incertezas na espacialização de propriedades ambientais amostradas pontualmente. Os resultados obtidos no exemplo mostram que a técnica se aplica com sucesso para propriedades ambientais de natureza temática. Essa técnica se aplica também à propriedades numéricas que são, em geral, tratadas após uma discretização de seu comportamento.

O trabalho apresenta, ainda, diferentes alternativas para estimativa dos mapas de incertezas. Os mapas de incerteza fornecem informação sobre a qualidade dos mapas inferidos. Essas incertezas são propagadas em modelagens ambientais que utilizam o atributo ambiental espacializado como dado de entrada. O mapa final, produto da modelagem, deve ser acompanhado do mapa de incerteza propagado que fornece informação quantitativa sobre os riscos assumidos ao se tomar decisões apoiadas nos produtos inferidos. Infelizmente as incertezas raramente são incorporadas aos produtos gerados pelos SIGs atualmente em uso. Espera-se que, com aumento da conscientização sobre a importância da qualidade da informação, os dados de incerteza passem a fazer parte dos requisitos básicos para projetos de modelagens espaciais desenvolvidas segundo a filosofia de um SIG. Este trabalho é uma contribuição nessa direção.

Referências Bibliográficas

- Burrough, P. A.; McDonnell, R. A., 1998. Principles of Geographical Information Systems. New York, Oxford University Press, 1998. 333p.
- Calderano Filho, B.; Fonseca, O. O. M.; Santos, H. G.; Lemos A. L. Levantamento Semidetalhado dos Solos da Fazenda Canchim São Carlos - SP. Rio de Janeiro, EMBRAPA- CNPS, 1996. 261p.
- Câmara, G., Souza, R. C. M., Freitas U. M., Garrido, J. SPRING: Integrating Remote Sensing and GIS by Object-Oriented Data Modelling. Computer & Graphics, v. 20, n. 3, p. 395-403, 1996.
- Camargo, E. C. G. Desenvolvimento, implementação e teste de procedimentos geoestatísticos (krigeagem) no sistema de processamento de informações georeferenciadas (SPRING). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.
- Deutsch, C. V.; Journel, A. G. GSLIB Geostatistical Software Library and User's Guide. New York, Oxford University Press, 1998. 369p.
- Isaaks, E. H.; Srivastava, R. M. An Introduction to Applied Geostatistics. New York, Oxford University Press, 1989. 561p.
- INPE/DPI - Instituto Nacional de Pesquisas Espaciais. Departamento de Processamento de Imagens. Sistema de Processamento de Informações Georreferenciadas (SPRING). <http://www.inpe.br/spring>. Março, 1999.
- Journel, A. G. Nonparametric Estimation of Spatial Distributions. Mathematical Geology, v. 15, n. 3, p. 445-468, 1983
- Journel, A. G.; Deutsch, C. V. Entropy and Spatial Disorder. Mathematical Geology, v. 25, n. 3, p. 329-355, 1993.
- Oliver, M. A.; Webster, R. Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information Systems, v. 4, n. 3, p. 313-332, 1990.
- Shannon, C. E.; Weaver, W. The Mathematical Theory of Communication. Urbana, The University of Illinois Press, 1949. 117p.