

Avaliação do procedimento geoestatístico de cokrigagem para determinação da distribuição espacial de propriedades de solos⁽¹⁾

Jussara de Oliveira Ortiz⁽²⁾

Carlos Alberto Felgueiras⁽²⁾

Suzana Druck⁽³⁾

Antonio Miguel Vieira Monteiro⁽²⁾

Resumo – Os dados coletados em amostras e perfis de solos, armazenados em banco de dados com registro de localização, constituem uma base para a produção de informação sobre a distribuição e a variabilidade espacial de propriedades que auxiliam a caracterização dos solos. Essa informação possibilita o uso mais racional dos recursos da terra, com real aumento da produtividade. Os dados amostrados tornam-se representativos de toda a área e, nos pontos não coletados, valores são estimados, por procedimentos de inferência, para que possamos ter uma visão completa da distribuição de uma ou mais propriedades dos solos na região de interesse. Neste trabalho, vamos abordar procedimentos inferenciais quantitativos, com base na teoria geoestatística. Em particular, vamos considerar os casos onde o número de amostras mostra-se insuficiente para uma estimativa mais qualificada, e mostrar como utilizar um processo de inferência que pode fazer uso de *informação secundária*, referente a uma medida correlacionada à propriedade em estudo. Este trabalho apresenta, testa e avalia o uso do procedimento geoestatístico de cokrigagem para modelagem de propriedades de dados de solos. Para demonstrar nossa proposta, dados de textura de solo, medidos aqui pelo teor de argila, foram processados e sua distribuição e variabilidade espacial determinadas e espacializadas utilizando-se dos procedimentos inferenciais aqui propostos. Os resultados são comparados aos resultados obtidos por procedimento determinístico de média ponderada e geoestatístico de krigagem,

(1)

(2) INPE – DPI - Divisão de Processamento de Imagens - email: jussara@dpi.inpe.br

(2) INPE – DPI - Divisão de Processamento de Imagens - email: carlos@dpi.inpe.br

(3) EMBRAPA – CPAC – Centro de Pesquisa Agropecuária do Cerrado suzana@cpac.embrapa.br

(2) INPE – DPI - Divisão de Processamento de Imagens - email: miguel@dpi.inpe.br

os quais utilizaram apenas a informação principal (a da propriedade de interesse) em seu processo inferencial. Termos para indexação: modelagem de atributos espaciais, geoestatística, cokrigagem.

Abstract – The data collected as soil samples and soil profiles, stored in database along with the register of their geographic locations, constitute a base for the information production about the variability and the distribution of spatial properties that assist the soil characterization. This information makes possible a more rational use of the earth resources with an undeniable increase of the productivity. The sample become representative of the whole area and, in the locations not collected, attribute values are estimated, for inference procedures, so that let us can have a complete vision of the distribution of the soil attribute in the region of interest. In this work we are going to explore quantitative inferencial procedures, on the basis of the geostatistical theory. In particular, we go to consider the cases where the number of samples reveals insufficient for a more qualified estimate and where a secondary information, referring to a measure correlated to the property in study, is used to improve this quality. This work presents, tests and evaluates the geostatistical procedure, known as cokriging, for spatial modelling of soil properties. To demonstrate our proposal, information of soil texture, represented by the clay concentration, was processed and its spatial distribution and variability was determined and spatialized using the inferencial procedures here considered. The results are compared with those gotten for deterministic procedures and by the geostatistical procedure of kriging, which use only the main information (of the property of interest) in their inferencial process.

Index Terms: modelling of spatial attributes, geostatistic, cokriging .

INTRODUÇÃO

Uma fonte importante de informação para aqueles que planejam a utilização agrícola são os mapas de caracterização da distribuição e variabilidade de certas propriedades físico-químicas dos solos numa área de interesse à produção. Ao longo de anos, trabalhos de campo têm recolhido, analisado, registrado e armazenado, em bancos de dados, uma grande quantidade de informação amostral sobre dados de geologia, de clima e informações geofísicas e geoquímicas relacionadas aos solos. Muitas dessas informações encontram-se disponíveis na forma de amostras pontuais, com registro de localização em coordenadas geográficas. Estas amostras são utilizadas em procedimentos de interpolação espacial, para estimar valores de determinados atributos em uma localização geográfica qualquer não observada, aqui denominada “**u**”. Os resultados destes processos de interpolação são, em geral, grades regulares, que passam a representar a área analisada como um espaço contínuo, onde aqueles atributos se encontram distribuídos. Essas grades são comumente geradas e armazenadas através do uso de Sistemas de Informação Geográfica (SIG). No contexto de um SIG, as grades são usadas em análises e modelagens que buscam integrar estas propriedades com várias outras propriedades e informações, gerando, como produtos, mapas que auxiliam o processo de tomada de decisão sobre problemas agrícolas ou ambientais em geral.

Os algoritmos para inferências de atributos de dados espaciais, representados por amostras pontuais, podem ser classificados em interpoladores determinísticos e interpoladores estocásticos ou probabilísticos. De um modo geral, os interpoladores determinísticos apresentam um valor para o atributo que é considerado exato, ou seja, são tratados como dados sem erros. Já os interpoladores estocásticos ou probabilísticos consideram a incerteza sobre os valores estimados. A abordagem probabilística fornece um conjunto de valores com as respectivas probabilidades de ocorrência. Ou seja, os interpoladores estocásticos, os quais fazem uso das ferramentas geoestatísticas para inferências de valores de atributos, permitem a estimativa das incertezas associadas a estes valores.

Os métodos tradicionais de inferência de um atributo Z utilizam-se somente de valores amostrados do próprio atributo, aqui denominado dado "primário" ou variável primária, para realizar suas estimativas. No entanto, outras informações relacionadas à distribuição do atributo Z, denominados dados "secundários" ou variáveis secundárias, podem estar disponíveis e, em conjunto com as amostras primárias, podem melhorar a qualidade da representação da distribuição espacial do atributo.

O uso de informação secundária é relevante quando, através de sua utilização, podemos reduzir as incertezas em relação à qualidade das informações disponíveis para uso dos tomadores de decisão (Felgueiras, 1999). Em geral, o uso da informação secundária no procedimento inferencial melhora os resultados da interpolação quando a quantidade de amostras da variável primária é escassa, ou insuficiente, para o nível de acurácia esperado, quando o custo de obtenção de mais amostras da variável primária é alto ou inviável e, quando há disponibilidade, a um custo baixo, de informação secundária que tenha um certo grau de correlação com a variável primária.

A informação secundária, mesmo quando não densamente amostrada, pode ser incorporada ao processo inferencial, por exemplo, se usamos a abordagem da cokrigagem, que considera a correlação espacial cruzada entre as variáveis primárias e secundárias. Neste trabalho, desenvolvemos e apresentamos o procedimento de cokrigagem para a determinação da variabilidade e distribuição espacial de propriedades de solos.

Dados de teor de argila, amostrados na bacia do Rio Jardim - DF – Brasil (Reatto, 2000), em dois horizontes diferentes de solo, são utilizados para testar e avaliar a técnica aqui proposta, e os resultados são comparados com os resultados obtidos com o uso de outros estimadores. Este trabalho apresenta uma avaliação e análise das possibilidades do uso da informação secundária no processo de inferência de atributos espaciais, principalmente onde a amostragem do dado primário seja insuficiente. O resultado obtido com o procedimento de cokrigagem é comparado com os resultados obtidos através do uso do estimador geoestatístico de krigagem ordinária, cujas inferências usam somente amostras do atributo em estudo e também do estimador determinístico de média ponderada.

MATERIAL E MÉTODOS

Localização da área de estudo

A bacia do rio Jardim está situada na parte leste do Distrito Federal, entre as latitudes sul 15°40' e 16°02' e longitudes oeste 47°20' e 47°40'. Os dados utilizados neste trabalho são os dados do estudo de caracterização e mapeamento da bacia, realizado pela Embrapa Cerrados. Todo o levantamento do clima e solos pode ser encontrado no boletim de pesquisa número 18 da Embrapa Cerrados (Embrapa, 2000). No levantamento realizado pela Embrapa ênfase, foi dada à classe dos latossolos, definidos como solos altamente intemperizados, com grande parte dos minerais existentes classificados como secundários, ou ainda, constituintes da fração de argila. Para este trabalho o atributo *teor de argila* foi selecionado das amostras coletadas em campo, para os horizontes superficial e subsequente. Para este estudo contamos com 36 amostras, cuja distribuição é mostrada na Figura 1, para as quais exploramos a propriedade de textura do solo, caracterizada aqui pelo teor de argila.

Procedimento Proposto: O Paradigma da Cokrigeagem

O procedimento da cokrigeagem é uma extensão natural do estimador de krigeagem, porém utilizado quando um variograma multivariado ou modelo de covariância e dados multivariados estiverem disponíveis. Uma variável de interesse pode ser cokrigada em uma localização específica, utilizando informações da própria variável e, também, de variáveis auxiliares correlacionadas, que estejam na sua vizinhança.

O interesse do estudo da dependência espacial de diferentes variáveis tem a ver fundamentalmente com a necessidade de que por vezes existe uma variável secundária (normalmente mais abundante) para auxiliar a estimativa de uma variável primária (menos abundante), com a qual está correlacionada.

A abordagem da cokrigeagem ao considerar a informação secundária utiliza-se da correlação espacial cruzada entre as variáveis primária e secundária, e para o caso de uma única variável secundária Y , o

estimador de cokrigagem ordinária, $Z_{Ock}^*(\mathbf{u})$, de $Z(\mathbf{u})$, a variável primária, é definido, segundo Deutsch e Journel, (1998) como:

$$Z_{Ock}^*(\mathbf{u}) = \sum_{\alpha 1=1}^{n1} \lambda_{\alpha 1}(\mathbf{u})Z(\mathbf{u}_{\alpha 1}) + \sum_{\alpha 2=1}^{n2} \lambda'_{\alpha 2}(\mathbf{u})Y(\mathbf{u}'_{\alpha 2}) \quad (1)$$

Onde os valores $\lambda_{\alpha 1}$, são os pesos aplicados às n_1 amostras da variável primária Z , na posição amostrada $u_{\alpha 1}$, e os valores $\lambda'_{\alpha 2}$ são os pesos aplicados às n_2 amostras da variável secundária Y , na posição amostrada $u_{\alpha 2}$. A krigagem requer um modelo para a covariância de Z , em relação a um vetor distância \mathbf{h} , $C_z(\mathbf{h})$. Já a cokrigagem, requer um modelo *conjunto* para a matriz de covariância, incluindo, a covariância de Z , $C_z(\mathbf{h})$, a covariância de Y , $C_y(\mathbf{h})$, a covariância cruzada Z - Y , $C_{zy}(\mathbf{h}) = \text{Cov}\{Z(\mathbf{u}), Y(\mathbf{u}+\mathbf{h})\}$, e a covariância cruzada Y - Z , $C_{yz}(\mathbf{h})$ (pode ou não ser simétrica). A matriz de covariância requer k^2 funções de covariância quando k diferentes variáveis são consideradas na cokrigagem, tornando custoso o processo de ajuste e a junção dos modelos matemáticos utilizados para inferência. Segundo o modelo de tendência $m_i(\mathbf{u})$ adotado, variantes de cokrigagem podem ser distinguidas (Goovaerts, 1997), no entanto, este trabalho trata apenas da cokrigagem ordinária.

A Cokrigagem Ordinária (CKO) considera variações locais das médias, por limitar o domínio da estacionariedade de ambas as médias, primária e secundária, ambas desconhecidas, para a vizinhança W local. Ou seja, $W(\mathbf{u}): m_i(\mathbf{u}') = \text{constante}$, porém desconhecida para $\forall \mathbf{u}' \in W(\mathbf{u}') \quad i = 1, \dots, k$

No caso da cokrigagem ordinária, o estimador linear de krigagem é, então, estendido para incorporar a informação secundária, da seguinte forma:

$$Z_I^*(\mathbf{u}) - m_I(\mathbf{u}) = \sum_{\alpha 1=1}^{n1} \lambda_{\alpha 1}(\mathbf{u})Z_I(\mathbf{u}_{\alpha 1} - m_I(\mathbf{u}_{\alpha 1})) + \sum_{\alpha i=1}^{ni} \lambda_{\alpha i}(\mathbf{u})Z_i(\mathbf{u}_{\alpha i} - m_i(\mathbf{u}_{\alpha i})) \quad (2)$$

Onde $\lambda_{\alpha 1}$ é o peso indicado para o dado primário $Z_I(\mathbf{u}_{\alpha 1})$, e $\lambda_{\alpha i}(\mathbf{u})$, $i > 0$, é o peso indicado para o dado secundário $Z_i(\mathbf{u}_{\alpha i})$. Os valores esperados das variáveis aleatórias $Z_I(\mathbf{u})$ e $Z_i(\mathbf{u}_{\alpha i})$ são denotados $m_I(\mathbf{u})$ e $m_i(\mathbf{u}_{\alpha i})$, respectivamente. Tipicamente, somente os dados primários e secundários, próximos à

localização \mathbf{u} que será estimada, são mantidos, de modo que $n_i(\mathbf{u})$ é usualmente menor que n_i . A quantidade de dados mantida e o tamanho da vizinhança de busca não necessitam ser os mesmos para todos os atributos. Todos os outros estimadores de cokrigagem são variantes da expressão 2. Para todos os estimadores exige-se que a variância do erro $\sigma_E^2(\mathbf{u})$, entre o valor estimado $Z_i^*(\mathbf{u})$ e o valor real $Z_i(\mathbf{u})$, de Z_i em \mathbf{u} , que é dada por:

$$\sigma_E^2(\mathbf{u}) = \text{Var}\{Z_i^*(\mathbf{u}) - Z_i(\mathbf{u})\} \quad (3)$$

seja minimizada, sob a restrição de não tendenciosidade do estimador, ou seja, que o erro esperado seja zero, *ou seja*, $E\{Z_i^*(\mathbf{u}) - Z_i(\mathbf{u})\} = 0$.

Os vários estimadores de cokrigagem diferem no modelo da função aleatória $Z_i(\mathbf{u})$ adotado para as várias variáveis. Tipicamente, cada Função Aleatória (F.A.) é decomposta em uma componente residual $R_i(\mathbf{u})$, e uma componente de tendência $m_i(\mathbf{u})$:

$$Z_i(\mathbf{u}) = R_i(\mathbf{u}) + m_i(\mathbf{u}) \quad \text{para } i = 1, \dots, K \quad (4)$$

A componente residual $R_i(\mathbf{u})$ é então modelada como uma F.A. estacionária com média zero, $E\{R_i(\mathbf{u})\} = 0$, e uma função de covariância, $C_i R(\mathbf{h})$, dada por:

$$\text{Cov}\{R_i(\mathbf{u}), R_i(\mathbf{u} + \mathbf{h})\} = E\{R_i(\mathbf{u})R_i(\mathbf{u} + \mathbf{h})\} = C_i R(\mathbf{h}) \quad (5)$$

A covariância cruzada entre quaisquer dois resíduos das V.A.s $R_i(\mathbf{u})$ e $R_j(\mathbf{u} + \mathbf{h})$ é:

$$C_{ij} R(\mathbf{h}) = \text{Cov}\{R_i(\mathbf{u}), R_j(\mathbf{u} + \mathbf{h})\} \quad (6)$$

Na Cokrigagem Ordinária Tradicional, a soma dos pesos aplicada à variável primária é apontada para um (1), e a soma dos pesos aplicada a qualquer outra variável é apontada para zero. No caso de duas variáveis, como na expressão 1, as condições são:

$$\sum_{\alpha 1=1}^{n1} \lambda_{\alpha 1}(\mathbf{u}) = 1 \quad \text{e} \quad \sum_{\alpha 2=1}^{n2} \lambda_{\alpha 2}(\mathbf{u}) = 0 \quad (7)$$

O problema com este formalismo é que, a segunda condição da expressão 7, tende a limitar severamente a influência da(s) variável(s) secundária(s).

A Cokrigagem Ordinária Padronizada utiliza-se da criação de variáveis secundárias com a mesma média, análogo à variável primária. Assim, todos os pesos ficam restritos a somar um (1). Neste caso a expressão 1 pode ser escrita como:

$$Z_{ock}^*(\mathbf{u}) = \sum_{\alpha 1=1}^{n1} \lambda_{\alpha 1}(\mathbf{u}) Z(\mathbf{u}_{\alpha 1}) + \sum_{\alpha 2=1}^{n2} \lambda'_{\alpha 2}(\mathbf{u}) [Y(\mathbf{u}'_{\alpha 2}) + m_z - m_y] \quad (8)$$

com a condição $\sum \lambda_{\alpha 1}(\mathbf{u}) + \sum \lambda'_{\alpha 2}(\mathbf{u}) = 1$. Os fatores $m_z = E\{Z(\mathbf{u})\}$ e $m_y = E\{Y(\mathbf{u})\}$ são médias estacionárias das variáveis aleatórias Z e Y respectivamente.

O conjunto de dados amostrados pode, ou não, cobrir todas as variáveis em todas as localizações amostradas. Dependendo de como as medidas das diferentes variáveis estiverem distribuídas no espaço, pode-se distinguir entre conjuntos de dados “*isotópicos*” e “*heterotópicos*” (Wackernagel, 1995).

Isotopia e Heterotopia

As medidas disponíveis para diferentes variáveis $Z_i(\mathbf{u})$, em um dado domínio podem estar localizadas nos mesmos pontos amostrais ou em pontos diferentes, para cada variável.

As seguintes situações, segundo Wackernagel (1995), podem ser diferenciadas:

- **dados totalmente heterotópicos:** as variáveis são mediadas em diferentes pontos amostrados, e não há amostras que partilhem da mesma localização;
- **dados parcialmente heterotópicos:** algumas variáveis podem ter sido amostradas na mesma posição;
- **isotopia:** Para cada variável, dados encontram-se disponíveis em todos os pontos amostrados. Ou seja, variáveis diferentes amostradas na mesma posição geográfica.

Consideremos aqueles casos em que na localização espacial de uma mesma amostra, são observados mais do que um atributo - $Z_1(u_i), Z_2(u_i), \dots, Z_N(u_i)$. No caso deste trabalho, a concentração de argila foi medida em dois horizontes (profundidades) diferentes, na mesma localização geográfica.

A correlação entre cada par daqueles atributos $Z_1(u_i), Z_2(u_i), \dots, Z_N(u_i)$ pode ser calculada através do coeficiente de correlação para o conjunto das N amostras, segundo Soares (2000), da seguinte forma:

$$\rho(Z_1, Z_2) = \frac{1}{N\sigma_1 \cdot \sigma_2} \sum_{i=1}^N [(Z_1(u_i) - m_1)(Z_2(u_i) - m_2)]$$

Onde: $m_1, m_2, \sigma_1^2, \sigma_2^2$ são as médias e variâncias de $Z_1(u)$ e $Z_2(u)$, respectivamente.

Pode-se, no entanto, generalizar este quadro de correlações entre as diferentes variáveis e calcular a correlação entre a variável $Z_1(\mathbf{u})$ localizada em \mathbf{u} e a variável $Z_2(\mathbf{u}+\mathbf{h})$ localizada em $\mathbf{u}+\mathbf{h}$. Esta dependência espacial entre cada par de variáveis distanciadas de \mathbf{h} pode ser caracterizada através de variogramas, covariâncias e correlogramas cruzados.

Assim, o modelo de função aleatória univariada pode ser generalizado para casos multivariados. O conjunto de variáveis aleatórias, que define N_V funções aleatórias, $i=1, \dots, N_V$, é também chamado função aleatória multivariada (Druck, 2004):

$$Z_i(\mathbf{u}), i=1, \dots, N_V, \forall \mathbf{u} \in A.$$

A distribuição conjunta de duas variáveis $Z_i(\mathbf{u})$ e $Z_j(\mathbf{u})$ depende do vetor \mathbf{h} :

$$F_{i,j}(h, Z_i, Z_j) = \text{prob}\{Z_i(u) \leq z_i, Z_j(u+h) \leq z_j\}, \forall i, j$$

A dependência espacial entre duas variáveis $Z_i(\mathbf{u})$ e $Z_j(\mathbf{u})$ pode ser medida pela função **covariância cruzada**

$$C_{i,j}(h) = E\{[Z_i(u) - m_i][Z_j(u+h) - m_j]\}, \forall i, j \quad (9)$$

ou, pelo **variograma cruzado**

$$\gamma_{i,j}(h) = \frac{1}{2} E\{[Z_i(u) - Z_i(u+h)][Z_j(u) - Z_j(u+h)]\}, \forall i, j \quad (10)$$

Nota-se que, $\gamma_{i,j}(h) = \gamma_{ji}(h)$

Porém, $C_{ij}(\mathbf{h})$ pode não ser igual a $C_{ji}(\mathbf{h})$, o que significa que não é uma função simétrica em relação à \mathbf{h} . A relação entre o variograma e covariâncias cruzadas pode ser escrita como:

$$\gamma_{ij}(h) = C_{ij}(0) - \frac{1}{2}[C_{ij}(h) + C_{ij}(-h)] \quad (11)$$

Nota-se que ao escrever a covariância cruzada como a soma de dois termos dependentes de \mathbf{h} ,

$$C_{ij}(h) = \frac{1}{2}[C_{ij}(h) + C_{ij}(-h)] + \frac{1}{2}[C_{ij}(h) - C_{ij}(-h)] \quad (12)$$

Pode-se constatar que o variograma cruzado (11) só incorpora o primeiro termo, de modo que a função é simétrica em relação à \mathbf{h} .

Segundo Wackernagel (1995), dados totalmente heterotópicos acarretam um problema para a inferência do variograma cruzado ou modelo de covariância. Os variogramas cruzados experimentais não podem ser computados por dados totalmente heterotópicos. As covariâncias cruzadas, quando podem ser computadas, são também problemáticas quando os valores correspondentes de covariância direta referem-se a diferentes conjuntos de pontos.

Função de covariância cruzada

De acordo com Wackernagel (1995), a covariância cruzada entre duas funções aleatórias pode ser computada não somente nas localizações \mathbf{u} , mas também para pares de localizações separadas por um vetor de distância \mathbf{h} . Sob a suposição de estacionariedade de segunda ordem uma função de covariância cruzada entre duas funções aleatórias é definida como dependente só do vetor \mathbf{h} .

As funções de covariância diretas e cruzadas $C_{ij}(\mathbf{h})$ de um conjunto de p F.As $Z_i(\mathbf{u}), i = 1, \dots, p$ são definidas sob a hipótese estacionária de segunda ordem como:

$$E[Z_i(\mathbf{u})] = m_i, \text{ para todo } \textit{para todo } \mathbf{u} \in D ; i=1, \dots, N$$

$$E[(Z_i(\mathbf{u}) - m_i) \cdot (Z_j(\mathbf{u} + \mathbf{h}) - m_j)] = C_{ij}(\mathbf{h}) \textit{ para todo } \mathbf{u} \textit{ e } \mathbf{u} + \mathbf{h} \in D ; i, j = 1, \dots, N$$

A média de cada variável $Z_i(\mathbf{u})$ em qualquer ponto do domínio é igual a uma constante m_i . A covariância de um par de variáveis depende somente do vetor \mathbf{h} que une um par de pontos e é invariante para qualquer translação do par de pontos no domínio D .

Um conjunto de funções de covariância cruzada é uma função positiva definida, isto é, ω_α^i necessita ser positiva. Para qualquer grupo de pontos $u_\alpha \in D$ e qualquer grupo $\omega_\alpha^i \in \mathfrak{R}$.

$$\begin{aligned} \text{Var} \left\{ \sum_{i=1}^N \sum_{\alpha=0}^n w_\alpha^i Z_i(u_\alpha) \right\} \\ = \sum_{i=1}^N \sum_{j=1}^N i \sum_{\alpha=0}^n \sum_{\beta=0}^n w_\alpha^i w_\beta^j C_{ij}(u_\alpha - u_\beta) \geq 0 \end{aligned} \quad (13)$$

O aspecto interessante da função de *covariância cruzada* é que seus valores podem ser diferentes para valores $+\mathbf{h}$ e $-\mathbf{h}$, ou seja, em geral ela não é uma função par. Isto ocorre quando o efeito de uma variável sobre a outra variável é defasado (“delayed”). Entretanto, o *semivariograma cruzado* é uma função par definida sob a hipótese intrínseca.

Na prática, a componente assimétrica da covariância cruzada é usualmente ignorada, basicamente por duas razões:

- A quantidade de dados disponível raramente permite a compreensão e a conseqüente validação para todo o campo do fenômeno físico, que origina a assimetria da covariância cruzada.
- Modelar a covariância cruzada assimétrica é extremamente complexa.

Assim, os instrumentos geoestatísticos que quantificam a continuidade espacial de um sistema multivariado, são normalmente os variogramas cruzados e as covariâncias cruzadas simétricas:

$$C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h}), \forall i, j, \text{ ou a média de } C_{ij}(\mathbf{h}) \text{ e } C_{ji}(\mathbf{h}).$$

O Semivariograma cruzado

Segundo Wackernagel (1995), o variograma direto e o cruzado são definidos no contexto da hipótese conjunta intrínseca para p F.A's quando \mathbf{u} e $\mathbf{u} + \mathbf{h} \in D$ e todos os pares $i, j = 1, \dots, p$

$$\begin{cases} E[Z_i(\mathbf{u} + \mathbf{h}) - Z_i(\mathbf{u})] = 0 \\ \text{cov}[(Z_i(\mathbf{u} + \mathbf{h}) - Z_i(\mathbf{u})) \cdot (Z_j(\mathbf{u} + \mathbf{h}) - Z_j(\mathbf{u}))] = 2\gamma_{ij}(\mathbf{h}) \end{cases} \quad (14)$$

O variograma cruzado é definido como

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2} E[(Z_i(\mathbf{u} + \mathbf{h}) - Z_i(\mathbf{u})) \cdot (Z_j(\mathbf{u} + \mathbf{h}) - Z_j(\mathbf{u}))] \quad (15)$$

O semivariograma cruzado é uma função par:

$$\gamma_{ij}(\mathbf{h}) = \gamma_{ji}(\mathbf{h})$$

$$\gamma_{ij}(-\mathbf{h}) = \gamma_{ij}(\mathbf{h})$$

A relação entre o semivariograma cruzado e a função de covariância cruzada é definida por:

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(0) - \frac{1}{2}(C_{ij}(-\mathbf{h}) + C_{ij}(\mathbf{h})) \quad (16)$$

o que mostra que o variograma cruzado é a média dos valores da covariância cruzada em $(-\mathbf{h})$ e $(+\mathbf{h})$

Modelos de Variogramas

Uma vez calculados os valores dos variogramas para diferentes passos \mathbf{h} , com base em conjunto de amostras de uma área A , é necessário modelar através de uma função geral e representativa, o que significa ajustá-los por uma curva média. Este ajuste é função de um número reduzido de parâmetros que possibilita quantificar a continuidade espacial de $Z(\mathbf{u})$.

Ajustar o variograma através de uma curva média é inferir uma função de $\gamma(\mathbf{h})$ representativa para toda a área, a partir dos valores conhecidos experimentalmente.

Esta etapa é extremamente importante dentro da metodologia geoestatística, pois é neste ponto que se faz a síntese das características estruturais do fenômeno espacial, ou seja, o grau de dispersão/continuidade, anisotropias, imbricamentos, em um modelo único e coerente de variograma.

Contudo, não se trata de obter um estimador de $\gamma(\mathbf{h})$ de uma FA estacionária $Z(\mathbf{u})$, mas sim de modelar e sintetizar os principais padrões de continuidade espacial, representativos do fenômeno em estudo.

Condição de positividade dos modelos

Dentre as inúmeras funções que podem interpolar os pontos experimentais de um variograma, a escolha deve ser limitada àquelas que forneçam soluções estáveis no posterior cálculo dos estimadores lineares.

Ou seja, os variogramas e covariâncias devem ser positivos-definidos. A condição necessária para que uma covariância seja positiva-definida é:

$$\sum_i \sum_j \lambda_i \lambda_j C(i, j) \geq 0$$

Qualquer combinação linear de covariâncias entre pares de pontos da área A é sempre positiva ou nula.

Ao considerar uma variável $Z(\mathbf{u}_0)$ resultante de uma combinação linear das variáveis aleatórias $Z(\mathbf{u}_1)$,

$Z(\mathbf{u}_2), \dots, Z(\mathbf{u}_N)$, ou seja, $Z(u_0) = \sum_i \lambda_i Z(u_i)$, então uma covariância definida positiva garante que a

variância de $Z(\mathbf{u}_0)$ é sempre positiva:

$$\begin{aligned} \text{var}\{Z(u_0)\} &= E\left\{\sum_i \sum_j \lambda_i \lambda_j Z(u_i) Z(u_j) - m^2\right\} \\ &= \sum_i \sum_j \lambda_i \lambda_j E\{Z(u_i) Z(u_j) - m^2\} \\ &= \sum_i \sum_j \lambda_i \lambda_j C(i, j) \geq 0 \end{aligned} \quad (17)$$

Substituindo a relação $\gamma(h) = C(0) - C(h)$ em (29), tem-se a variância escrita em função do variograma:

$$\text{var}\{Z(u_0)\} = C(0) \sum_i \sum_j \lambda_i \lambda_j - \sum_i \sum_j \lambda_i \lambda_j \gamma(i, j) \geq 0 \quad (18)$$

No caso de $C(0)$ não existir em funções aleatórias não estacionárias (com variogramas sem patamar), a

variância de $Z(\mathbf{u}_0)$ existe se $\sum_i \lambda_i = 0$. Ou seja, a condição necessária de positividade da variância é

garantida se $\sum_i \sum_j \lambda_i \lambda_j \gamma(i, j) \leq 0$, condicionada à soma dos ponderadores ser nula.

Modelando a Corregeionalização

Modelar a corregeionalização implica em inferir $p(p+1)/2$ semivariogramas diretos e cruzados, segundo critérios que determinam os modelos válidos ou permissíveis (Druck, 2004). Ou seja, consiste em

definir o modelo, para duas ou mais variáveis, que fornecerá valores do semivariograma ou covariâncias para qualquer valor de “**h**” necessário no processo de interpolação.

De acordo com Soares (2000), analogamente ao caso univariado, os modelos dos variogramas e covariâncias cruzadas têm de assegurar a positividade da variância de qualquer combinação linear entre as variáveis, conforme em (17). Os modelos permissíveis (válidos) devem, portanto, garantir que a estrutura de correlação para a combinação das variáveis tenha variância ≥ 0 .

Existem vários modelos de correionalização. No entanto, o modelo mais usual é o chamado modelo linear, no qual os variogramas simples (diretos) e cruzados resultam de uma combinação linear de modelos básicos.

Os modelos básicos (“efeito pepita”, primeira estrutura, segunda estrutura etc) têm de fazer parte de todos os variogramas simples e cruzados.

Modelo Linear de Correionalização

Cada variável é caracterizada por seu próprio variograma direto e cada par de variáveis por seus próprios variogramas cruzados. O modelo, para cada um destes variogramas, pode consistir de uma ou mais estruturas (modelos) básicas conforme mostrado na equação (19), contudo, o mesmo modelo básico deve aparecer em cada variograma direto ou cruzado (Isaaks e Srivastava,1998). Em outras palavras, cada modelo de variograma direto ou cruzado deve ser construído usando os mesmos modelos básicos de variograma. Por exemplo, considere duas variáveis aleatórias, Z e Y . Os modelos de variograma direto e cruzado de Z e Y devem ser construídos como:

$$\begin{aligned}\gamma_Z(\mathbf{h}) &= z_0\gamma_0(\mathbf{h}) + z_1\gamma_1(\mathbf{h}) + \dots + z_m\gamma_m(\mathbf{h}) \\ \gamma_Y(\mathbf{h}) &= y_0\gamma_0(\mathbf{h}) + y_1\gamma_1(\mathbf{h}) + \dots + y_m\gamma_m(\mathbf{h}) \\ \gamma_{ZY}(\mathbf{h}) &= w_0\gamma_0(\mathbf{h}) + w_1\gamma_1(\mathbf{h}) + \dots + w_m\gamma_m(\mathbf{h})\end{aligned}\tag{19}$$

Onde: $\gamma_Z(\mathbf{h})$, $\gamma_Y(\mathbf{h})$, $\gamma_{ZY}(\mathbf{h})$ são os modelos para os variogramas direto e cruzado para Z e Y respectivamente; os modelos básicos são dados por $\gamma_0(\mathbf{h})$, $\gamma_1(\mathbf{h}), \dots, \gamma_m(\mathbf{h})$; \mathbf{z} , \mathbf{y} e \mathbf{w} são coeficientes, possivelmente negativos. Pode-se rescrever a expressão (19) na forma matricial, como uma

combinação de cada modelo básico (Goovaerts, 1997). Para assegurar que o modelo linear dado na expressão (19) seja positivo, é suficiente garantir que todas as matrizes dos coeficientes z , y e w , sejam positivas. Isto implica que os coeficientes devem ser escolhidos de modo que:

$$\begin{aligned} z_j > 0 \text{ e } y_j > 0, & \quad \text{para todo } j = 0, \dots, m \\ z_j \cdot y_j > w_j \cdot w_j, & \quad \text{para todo } j = 0, \dots, m \end{aligned} \quad (20)$$

As condições estabelecidas em (20) sugerem uma importante consideração ao modelar-se a correionalização, que qualquer modelo básico que esteja incluído na modelagem do variograma cruzado deve necessariamente estar incluído nos modelos dos variogramas diretos.

O modelo linear de correionalização é muito conveniente para que, as condições de modelagem das variáveis correionalizadas, sejam prontamente verificadas, e com isso o modelo mais adequado ao variograma cruzado seja ajustado. Uma destas importantes condições verificadas ao utilizar o modelo linear de correionalização determina que todos os elementos diagonais das matrizes de correionalização sejam positivos. Todos os sub-determinantes, das matrizes de ordem 2 e 3, também devem ser positivos. Algumas regras para a escolha das estruturas básicas, no modelo de correionalização, podem ser encontradas em (Goovaerts, 1997).

Resumidamente, conforme Drucks (2004), o modelo linear de correionalização é a combinação linear de modelos de covariâncias diretas e cruzadas. Em notação matricial se $C(h)=[C_{ij}(h)]$ é uma matriz de covariância $p \times p$ e similarmente $\gamma(h)=[\gamma_{ij}(h)]$, o modelo pode ser escrito como:

$$C(h) = \sum_{s=0}^S B^s c^s(h) \quad \text{ou} \quad C_{ij}(h) = \sum_{s=0}^S b_{ij}^s c^s(h)$$

$$\gamma(h) = \sum_{s=0}^S B^s g^s(h) \quad \text{ou} \quad \gamma_{ij}(h) = \sum_{s=0}^S b_{ij}^s g^s(h)$$

As condições para que $C_{ij}(h) = \sum_{s=0}^S b_{ij}^s c^s(h)$ seja um modelo de correionalização permissível são:

1. as funções $c^s(h)$ sejam modelos de covariância permissíveis;

2. as S+1 matrizes de correionalização B^s sejam positivas semi-definidas

Modelo Intrínseco de Correionalização

De acordo com Druck (2004), o modelo intrínseco de correionalização é um caso particular do modelo linear uma vez que todos os $p^{(p+1)/2}$ coeficientes b_{ij}^s são proporcionais entre si, isto é,

$$b_{ij}^s = \varphi_{ij} \cdot b^s \quad \forall i, j \quad (21)$$

Todos os semivariogramas e os semivariogramas cruzados são obtidos por um reescalonamento do mesmo modelo linear padronizado de regionalização $\gamma_o(h)$

$$\begin{aligned} \gamma_{ij}(\mathbf{h}) &= \varphi_{ij} \gamma_o(\mathbf{h}) \quad \forall i, j \\ &= \varphi_{ij} \sum_{s=0}^S b^s g^s(\mathbf{h}) \rightarrow \sum_{s=0}^S b^s = 1 \end{aligned} \quad (22)$$

Esse modelo pode ser expresso em termos da matriz de covariância:

$$\begin{aligned} C_{ij}(\mathbf{h}) &= \varphi_{ij} c_o(\mathbf{h}) \quad \forall i, j \\ &= \varphi_{ij} \sum_{s=0}^S b^s c^s(\mathbf{h}) \rightarrow \sum_{s=0}^S b^s = 1 \end{aligned} \quad (23)$$

Considerando como um caso especial do modelo de correionalização linear,

$$Z_i(\mathbf{u}) = \sum_{s=0}^S \sum_{\alpha=1}^n a_{i\alpha} Y_i^s(\mathbf{u}) + m_i$$

$$E[Y_i^s(\mathbf{u})] = 0$$

e

$$Cov[Y_i^s(\mathbf{u}), Y_j^s(\mathbf{u} + \mathbf{h})] = E[Y_i^s(\mathbf{u}) \cdot Y_j^s(\mathbf{u} + \mathbf{h})] = c^s(\mathbf{h})$$

$$Cov[Y_i^s(\mathbf{u}), Y_j^v(\mathbf{u} + \mathbf{h})] = 0, s \neq v$$

O modelo intrínseco de correionalização é fácil de ser inferido, mas muito mais restritivo porque requer que todos variogramas experimentais diretos e cruzados sejam proporcionais entre si.

A condição de proporcionalidade pode ser checada através do formato experimental das curvas ou verificando se a razão entre dois variogramas é constante e independente de h :

$$\frac{\gamma_{ij}(h)}{\gamma_{i'j'}(h)} = c o \forall i, i', j, j' \quad (24)$$

Ou, utilizando o coeficiente de codispersão:

$$cc_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(h)\gamma_{jj}(h)}} \quad (25)$$

Se os coeficientes de codispersão são constantes, o modelo é intrínseco:

$$cc_{ij}(h) = \frac{b_{ij}\gamma(h)}{\sqrt{b_{ii}b_{jj}}\gamma(h)} = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}} = \rho_{ij} \quad (26)$$

Ou seja, a relação entre duas variáveis é invariante a h .

Um outro teste possível para verificar a correlação intrínseca através dos componentes principal das variáveis. Se o semivariograma cruzado entre os componentes principais não é zero em todos os lags então o modelo de correionalização intrínseca não é apropriado.

Cokrigeagem Isotópica

Quando um conjunto de variáveis apresenta-se intrinsecamente correlacionada, a cokrigeagem é equivalente a krigeagem para cada variável, se todas as variáveis tiverem medidas, em todas as localizações amostradas. Por tratar-se do caso deste estudo, será abordado a seguir.

A cokrigeagem com dados isotópicos

Considerando isotopia a cokrigeagem de um conjunto de variáveis tem a vantagem importante, sobre a krigeagem de cada variável, que é preservar a coerência dos estimadores. Pode-se verificar, por exemplo, a estimativa da soma de variáveis:

$$S(u) = \sum_{i=1}^N Z_i(u)$$

A cokrigeagem de $S(u)$ é igual a soma das cokrigeagens das variáveis $Z_i(u)$

$$S^{CK}(u) = \sum_{i=1}^N Z_i^{CK}(u)$$

Contudo, ao efetuar a krigagem de cada termo da soma e posteriormente adicioná-las, geralmente não obtém-se o mesmo resultado da krigagem direta dos dados somados: as duas estimativas não são coerentes.

Conforme ilustra exemplo em Wackernagel (1995): “a espessura $T(\mathbf{u})$ de uma camada geológica é definida como a diferença entre seus limites superior ($Z_S(u)$) e inferior ($Z_I(u)$),

$$T(u) = Z_S(u) - Z_I(u)$$

Os estimadores de cokrigagem de cada termo, usando a informação de duas variáveis é dada por,

$$T^{CK}(u) = Z_S^{CK}(u) - Z_I^{CK}(u),$$

porém, se a krigagem é usada, o termo resultante à esquerda é, em geral, diferente do termo que define a diferença, à direita, conforme segue:

$$T^K(u) \neq ZK_S(u) - ZI^K(u)$$

Em algumas situações a cokrigagem isotópica é equivalente a krigagem. O caso trivial é quando todos os variogramas cruzados ou as funções de covariância cruzadas são zero. Um caso menos evidente é quando os variogramas cruzados ou as covariâncias são proporcionais ao variograma direto ou à função de covariância.

Autokrigeabilidade

Segundo Wackernagel (1995) uma variável é dita ser *autokrigeável* em relação a um conjunto de variáveis, se a krigagem direta desta variável é equivalente à cokrigagem. O caso mais comum é quando todas as variáveis são decorrelacionadas (em qualquer escala) e é fácil verificar que isto implica que todas os pesos da krigagem serão zero, exceto para a variável de interesse.

Em uma situação de isotopia, qualquer variável é autokrigeável se ela pertencer a um conjunto de variáveis intrinsecamente correlacionadas.

A caracterização do modelo intrínseco de correionalização, no caso isotópico, invalida a cokrigagem, pois, demonstra-se que a estrutura de covariância não é adequada para fornecer valores de semivariograma ou de covariância para o processo de interpolação. Neste caso, a krigagem mostra-se mais adequada.

Para uma variável ser considerada autokrigeável é necessário que os variogramas cruzados ou covariâncias com esta variável sejam proporcionais ao variograma direto ou à covariância. Para testar, no caso de situação isotópica, se a variável é autokrigeável, Wackernagel (1995) sugere computar coeficientes de autokrigeabilidade (ac), definido como a razão do variograma cruzado contra o variograma direto:

$$ac_{ij}(h) = \frac{\gamma_{ij}(h)}{\gamma_{ii}(h)}$$

Se os coeficientes (ac 's) forem constantes para qualquer escala de trabalho, para cada variável $j=1, \dots, N$ de um conjunto de variáveis, então a variável de interesse é dita autokrigeável em relação a este conjunto de variáveis.

Avaliando a Influência da Variável Secundária

Estas ponderações são aqui consideradas para a comparação com os resultados obtidos, apresentados e discutidos neste trabalho. Goovaerts (1997) apresenta uma formulação que possibilita medir a influência da informação secundária, através da razão das somas dos valores absolutos, dos pesos atribuídos aos dados primários e secundários, denominada $\Psi(\mathbf{u})$. Esta formulação é usada neste trabalho para as análises dos resultados obtidos.

Antes de apresentar um exemplo, faz-se necessário dizer que dependendo da densidade de amostragem das variáveis primárias e secundárias, pode-se distinguir duas situações de configuração para os dados:

1. Todas as K variáveis estão registradas para cada localização amostrada, situação chamada de igualmente amostrada ou caso isotópico e;

2. A variável primária está pouco amostrada em relação às variáveis secundárias, situação denominada subamostrada ou caso heterotópico, onde, tipicamente poucas medidas do atributo de interesse são suplementadas por dados mais abundantes, de atributos correlacionados, e mais baratos de serem amostrados.

No gráfico mostrado na Figura 2, podemos ver uma situação com uma configuração de dados isotópica, onde, tanto as variáveis primárias quanto as secundárias, foram medidas para quatro localizações, e considerando uma unidade de distância a partir da localização que será estimada, \mathbf{u} . Também são consideradas duas configurações de dados heterotópicas.

Na *configuração 1*, quatro dados secundários estão localizados em um círculo de raio 0,5, para um total de oito dados secundários e, onde \mathbf{u} é a localização espacial do ponto a ser estimado. Na *configuração 2*, um único dado secundário está colocalizado com \mathbf{u} , num total de cinco dados secundários. Para as três configurações, a razão $\Psi(\mathbf{u})$ aumenta com o aumento do coeficiente de correlação.

Na Figura 2 o caso isotópico é representado pela linha sólida. Os dois casos sub-amostrados são representados, para a configuração 1, por pequenas linhas pontilhadas, e para a configuração 2, com pontilhado maior.

Tanto para a configuração 1 quanto para a configuração 2, a contribuição da informação secundária aumenta quando a variável secundária apresenta-se melhor amostrada que a variável primária. Além disso, quanto mais próximos os dados secundários estiverem dos primários. Para Goovaerts (1997), *as estimativas de krigagem e cokrigagem são essencialmente as mesmas, para o caso isotópico. Ainda mais, as diferenças entre as estimativas aumenta quando os dados secundários tornam-se mais numerosos que os dados primários.* E ainda, que no caso isotópico, o ganho da contribuição da informação secundária, só torna-se relevante quando o coeficiente de correlação é maior do que 0,7. Estas últimas observações serviram de referência para a avaliação dos resultados aqui obtidos.

Metodologia

Os modelos de dados geográficos, suas representações computacionais e as operações realizadas sobre elas, foram implementados segundo a modelagem lógica do banco de dados geográfico do Sistema de Processamento de Informações Georeferenciadas, SPRING (SPRING, 2002).

O Experimento

1ª Etapa: - Preparação dos Dados

Fase 1 : As amostras contendo o teor de argila nos horizontes superficial e subsequente, para a bacia do rio Jardim, modeladas como amostras pontuais, foram preparadas em um arquivo texto, formato SPRING padrão ASCII (SPRING, 2002), com a formatação adequada para integrar um banco de dados espaciais dentro do ambiente desse SIG.

Fase 2 : Definição do Modelo de dados do Banco Espacial no ambiente Spring e importação dos dados de teor de argila , modelados como amostras pontuais , do tipo Numérico.

2ª Etapa: Análise Exploratória dos Dados em Ambiente de SIG

Fase 1: *Verificação de Corregionalização:* Já com os dados importados no Spring, com o módulo de *Análise de Geoestatística*, uma *Análise Exploratória* dos dados foi efetuada, para avaliar o grau de correlação entre as variáveis (Figura 3). O diagrama de dispersão das variáveis argila no horizonte superficial e argila no horizonte subsequente, mostrou um alto coeficiente de correlação, 0.86, entre estas variáveis, evidenciando que o processo de cokrigagem poderia ser utilizado com estas variáveis.

Fase 2: *Verificação de Correlação Espacial:* Foram gerados os semivariogramas experimentais para: Argila no horizonte superficial- Argila 1 (Figura 4), Argila no horizonte subsequente- Argila 2 (Figura 5) e o semivariograma cruzado das duas variáveis (Figura 6). A análise dos semivariogramas experimentais, para cada variável separadamente, sugeriu o ajuste com o modelo *exponencial*, considerando a variabilidade do dado como omnidirecional. Não sendo explorado, portanto, padrões de anisotropia.

3ª Etapa: Ajuste dos Modelos

Fase 1: O ajuste final dos semivariogramas experimentais ao modelo exponencial, foi efetuado fora do ambiente SPRING, na plataforma de análise de dados espaciais conhecida como VarioWin (Pannatier, 1996 e VARIOWIN, 1999). Para isso foram utilizados os módulos PREVAR2D, VARIO2D e MODEL do VarioWin. Os modelos ajustados estão apresentados, juntamente com os semivariogramas, nas Figuras 5, 6 e 7.

Fase 2: Modelo Linear de Correogionalização

É importante ressaltar que a cokrigeagem requer, para garantir a adequada modelagem da correogionalização, que os modelos ajustados sigam as indicações apresentadas no item de modelagem da correogionalização deste trabalho. Uma destas importantes considerações é que toda estrutura básica que aparece no variograma cruzado também apareça nos variogramas diretos. Assim, os três modelos foram ajustados por funções exponenciais, conforme é mostrado a seguir.

Os modelos de semivariogramas definidos neste estudo apresentam as seguintes equações:

$$\begin{aligned}\gamma_1(h) &= 1040 + 13730 \exp\left(\frac{h}{3600}\right) \\ \gamma_2(h) &= 1120 + 14480 \exp\left(\frac{h}{3600}\right) \\ \gamma_{12}(h) &= 947 + 13806 \exp\left[\frac{h}{3600}\right]\end{aligned}\quad (27)$$

Aonde γ_1 representa o modelo da variável Argila1, γ_2 o modelo da variável Argila2 e γ_{12} , o modelo de semivariograma cruzado das duas variáveis. Na forma matricial, um ajuste com o modelo linear de correogionalização, incorporando duas estruturas (ou duas escalas de atuação do fenômeno), conforme os modelos definidos em (27), temos:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} = \begin{bmatrix} b_{11}^0 & b_{12}^0 \\ b_{21}^0 & b_{22}^0 \end{bmatrix} g_0(h) + \begin{bmatrix} b_{11}^1 & b_{12}^1 \\ b_{21}^1 & b_{22}^1 \end{bmatrix} g_1(h) \quad (28)$$

$$\begin{matrix} & B^0 & & B^1 \\ \text{Então:} & & & \end{matrix}$$

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} = \begin{bmatrix} 1040947 & \\ & 947 \ 1120 \end{bmatrix} g_0(h) + \begin{bmatrix} 1373013806 & \\ & 1380614480 \end{bmatrix} g_1(h) \quad (29)$$

Onde: B^0 e B^1 devem ser positivas semi-definidas, isto é, são matrizes cujos determinantes 2x2 devem ser positivos definidos ou zero, assim como também seus elementos da diagonal, $b_{ii}^s \geq 0$, para $s=0,1,2$ e $i=1,2$.

$$\text{Ou seja: } |b_{ij}^s| \leq \sqrt{b_{ii}^s b_{jj}^s} \quad (42)$$

Assim, para o caso em estudo, mostra-se que as matrizes B^0 e B^1 são positivas semidefinidas:

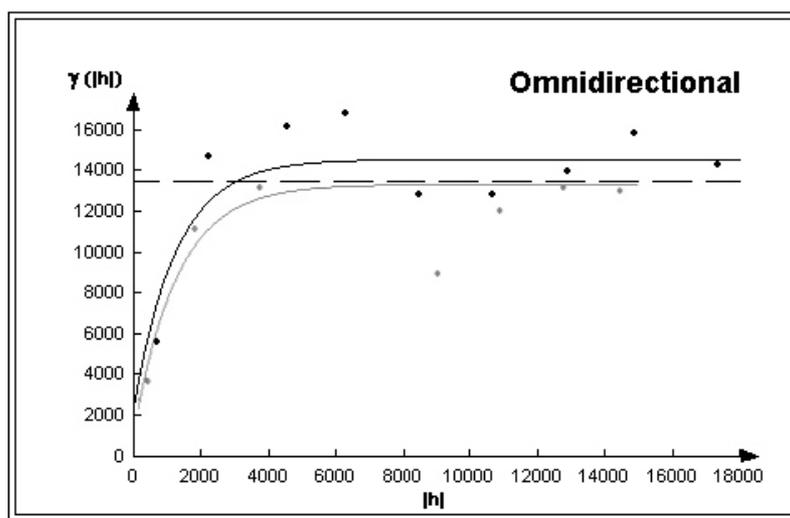
$$B_0 = \begin{bmatrix} 1040947 & \\ & 947 \ 1120 \end{bmatrix} \quad |947| \leq \sqrt{1040 * 1120} \quad \text{ou} \quad 947 \leq 1079$$

$$B_1 = \begin{bmatrix} 1373013806 & \\ & 1380614480 \end{bmatrix} \quad |13806| \leq \sqrt{13730 * 14480} \quad \text{ou} \quad 13806 \leq 14100$$

Com amostragem isotópica, que é o caso deste estudo, o modelo de correção intrínseca, que é um caso particular do modelo linear, não se mostra adequado para a cokrigagem. Portanto, torna-se necessário verificar a condição de correlação intrínseca, antes de considerar o procedimento de cokrigagem isotópica.

Um teste possível é verificar a condição de proporcionalidade dos variogramas experimentais diretos e cruzados, para todas as distâncias (lags). Essa condição pode ser checada através do formato das curvas, ou verificando se a razão entre dois variogramas é constante e independente de h , utilizando o coeficiente de codispersão.

Quanto ao formato das curvas: Conforme pode-se observar no gráfico abaixo, o formato das curvas não caracteriza proporcionalidade.



Quanto ao coeficiente de codispersão: Se os coeficientes de codispersão (cc) são constantes, então o modelo é intrínseco, o que caracteriza a não aplicabilidade do modelo linear de correção regionalização.

Portanto, a seguir, estes coeficientes serão apresentados.

Os coeficientes de codispersão são calculados da seguinte forma:

$$cc_{ij}(h) = \frac{b_{ij}\gamma(h)}{\sqrt{b_{ii}b_{jj}\gamma(h)}} = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}} = \rho_{ij} \quad (31)$$

Onde os valores b 's representam os elementos das matrizes de correção regionalização B 's, considerando a variável "i" (Argila1), a variável "j" (Argila2) e as duas variáveis juntas (ij).

Conforme pode ser observado na tabela abaixo, os valores dos coeficientes não são constantes. Assim, o modelo de correção regionalização linear é considerado modelo permissível para este estudo de caso, que tem amostragem isotópica.

Coeficientes de Codispersão - ccij(h)			
b12	b11	b22	cc12
3600,00	5581,20	4960,00	0,6842233
11073,70	14721,40	14102,50	0,7685465
13150,00	16111,50	16900,00	0,7969194
13943,10	16798,40	18182,60	0,7978062
16978,20	12818,40	16460,70	1,1688279
8932,50	12809,10	12877,60	0,6954986
12034,80	13930,80	13639,10	0,8730879
13137,70	15812,10	14084,50	0,8803469
12998,50	14248,20	16947,40	0,8364918

Tabela 1 – Coeficientes de codispersão

Uma outra estatística que pode ser calculada diz respeito à correlação estrutural. Assim, o coeficiente de correlação na primeira estrutura é obtido ao fazer a razão entre o elemento cruzado (b_{ij}) da matriz de correionalização, para a primeira estrutura, pelo valor resultante do produto dos elementos da diagonal, também na primeira estrutura. O procedimento análogo é feito para as outras estruturas. Ou seja:

$$cc_{\text{estrutural}}(h) = \frac{b^s_{ij}}{\sqrt{b^s_{ii}b^s_{jj}}} = \rho_{ij}$$

Onde “s” representa a estrutura.

Assim, como neste estudo, o modelo de correionalização linear acima definido (29) apresenta duas estruturas diferentes, os coeficientes estruturais correspondem a:

Primeira Estrutura – Pode-se considerar ocorrência em micro escala

$$cc_1(h) = \frac{b^0_{12}}{\sqrt{b^0_{11}b^0_{22}}} = \frac{947}{\sqrt{1040*1120}} = 0.87$$

Segunda Estrutura – Pode-se considerar ocorrência média escala

$$cc_2(h) = \frac{b^1_{12}}{\sqrt{b^1_{11}b^1_{22}}} = \frac{13806}{\sqrt{13730*14480}} = 0.98$$

Com os valores obtidos para os coeficientes de correlação estrutural acima, podemos dizer que as duas variáveis (Argila1 e Argila2) estão atuando juntas tanto em micro, quanto em média escala.

4ª Etapa: O procedimento de Krigeagem

Esta etapa consistiu na realização do procedimento de krigeagem ordinária sobre os dados preparados anteriormente. Este procedimento foi realizado utilizando-se o programa kt3d.exe, disponibilizado na biblioteca GSLIB (Deutsch e Journel ,1998), e as informações da variografia dos dados do horizonte superficial.

5ª Etapa: O Procedimento de CoKrigeagem

A cokrigeagem foi também efetuada pela GSLIB. O procedimento foi muito semelhante ao efetuado para o cálculo da krigeagem, porém, neste caso, utilizou-se o programa cokb3d.exe e os parâmetros dos semivariogramas definidos, para a variável argila, nos dois horizontes considerados e do semivariograma cruzado.

RESULTADOS E DISCUSSÃO

Nesta seção são apresentados e discutidos os resultados obtidos para a estimativa do atributo argila no horizonte superficial, para a área da bacia do rio Jardim, utilizando os procedimentos inferenciais propostos, o uso o interpolador geoestatístico de cokrigeagem, e com a utilização de outros métodos de interpolação como o interpolador determinístico de média ponderada pelo inverso das distâncias às amostras vizinhas selecionadas por quadrantes, e o interpolador geoestatístico de krigeagem

A Figura 7 mostra a imagem (a) gerada com o interpolador de média ponderada pelo inverso das distâncias e a imagem (b) obtida com a krigeagem ordinária. A Figura 8 apresenta as imagens obtidas pelos interpoladores de média ponderada pelo inverso das distâncias em (a), e de cokrigeagem em (b).

Observando as imagens mostradas nas Figuras 8 e 9 pode-se constatar que a imagem gerada pelo interpolador da média ponderada produziu alguns padrões artificiais, gerando uma representação espacial incoerente entre pontos altos e baixos, não capturando portanto aspectos importantes da distribuição e variabilidade espacial do atributo estudado. Já as imagens geradas pelos interpoladores geoestatísticos, de krigeagem e de cokrigeagem, produziram representações espaciais mais coerentes com os valores das amostras e sem os padrões artificiais do modelo determinístico. Comparando-se, os modelos gerados pela krigeagem em relação a cokrigeagem observa-se que eles são muito parecidos. Uma análise qualitativa, visual, mostra o modelo de cokrigeagem com transições um pouco mais suaves, porém com tendência global similar à da krigeagem. Uma análise quantitativa simples, com possibilidade de avaliação de acurácia de estimação, pode ser realizada a partir de um conjunto de amostras de validação, diferente das usadas pelo estimador. Neste trabalho, isto não pôde ser realizado por falta de disponibilidade das amostras de validação.

A principal diferença entre o estimador de cokrigagem em relação ao de krigeagem pode ser observada ao compararmos as imagens com os valores de variância, geradas para cada um desses estimadores, e apresentados na Figura 9 (a) e (b). As imagens de variância de estimação, geradas no processamento da krigeagem e da cokrigagem, apresentaram valores máximos de 26000 e 14000, respectivamente. Estas imagens de variância foram quantizadas em cinco classes: *muito alta*, *alta*, *média*, *baixa* e *muito baixa*. A cada uma dessas classes atribuiu-se uma cor como mostrado na legenda da Figura 9. Observando a Figura 9 pode-se verificar que, no global, a variância resultante do estimador de krigeagem manteve-se alta, com algumas faixas de variância média e baixa em torno dos pontos amostrados e, variância muito baixa nas proximidades das localizações das amostras. A imagem de variância resultante da cokrigagem mostra, no global, uma variância média; pequenas faixas, ao redor dos pontos amostrados, com variância baixa e; muito próxima as amostras, variância muito baixa, como era de se esperar. Portanto, na comparação entre as imagens de variância, geradas pelos dois estimadores, constata-se um ganho do estimador de cokrigagem em termos de diminuição da variância de estimação, o que caracteriza uma estimativa com maior precisão.

CONCLUSÕES

Este trabalho explora a técnica geoestatística de cokrigagem, para o caso de amostragem isotópica, com o uso de informação secundária, para inferência de modelos de atributos espaciais para um conjunto de amostras de teor de argila na região da bacia do Rio Jardim. Os resultados obtidos mostram que essa técnica é bastante promissora e pode efetivamente melhorar a qualidade dos mapeamentos de distribuição e variabilidade para propriedades de solos, quando verificada a hipótese da variável não ser “autokrigeável” (termo definido em Wackernagel, 1995). O experimento demonstra que o procedimento inferencial utilizando a cokrigagem, aqui proposto:

- faz uso das representações de variabilidade espacial, variogramas diretos e cruzados, do atributo envolvido no processo de estimação e da informação secundária que pode ser utilizada, para determinação dos ponderadores das estimativas;

- baseia-se na hipótese de minimização da variância de estimação, gerando modelos espaciais para os parâmetros de solo mais coerentes com a tendência definida pelas amostras, evitando a criação de padrões artificiais, com transições abruptas, como aqueles gerados pelo interpolador de média ponderada pelo inverso das distâncias;
- possibilita a inclusão de informações auxiliares, secundárias, para inferência espacial, o que se torna mais relevante em casos de amostragem escassa do atributo primário.
- cria modelos mais confiáveis pois, em comparação com outras técnicas, como a média ponderada pelo inverso das distâncias e a krigagem, os valores em locais não observados são estimados com maior precisão, ou seja, com menor variância de estimação.

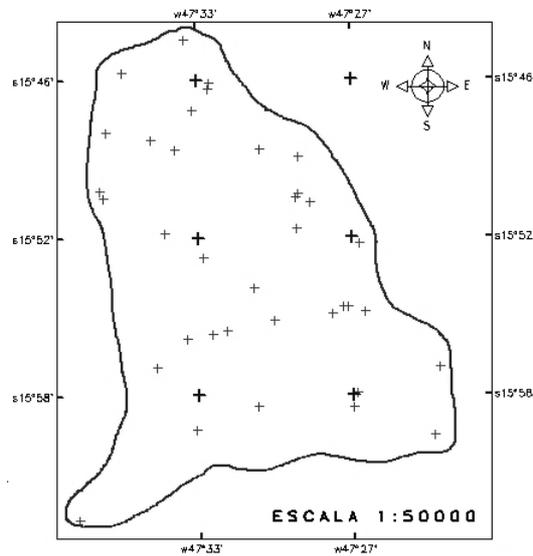
A limitação para o uso do estimador de cokrigagem, em relação ao estimador de krigagem, é o custo de se realizar o estudo de correlação espacial cruzada, entre as amostras. Esta é uma fase crítica do processo, que requer do especialista conhecimento e domínio das técnicas e das ferramentas geoestatísticas, geralmente interativas, para obtenção de um ajuste teórico de variografia, que seja uma boa representação da variabilidade espacial da propriedade em estudo.

Para finalizar, convém ressaltar que o estudo de caso deste trabalho foi realizado com amostragens do tipo isotópica (informação secundária co-localizada com o dado primário). Este é o pior caso de configuração de amostras, ou seja, o caso em que só se tem vantagem com o uso da variável secundária quando esta tem alto grau de correlação com a variável primária. Ainda assim, os resultados obtidos mostram que a variância das estimativas provenientes da cokrigagem diminuem significativamente, ou seja, há um ganho na precisão das estimativas e, portanto, uma melhoria na qualidade das representações espaciais para a distribuição e variabilidade de propriedades de solos, diminuindo as incertezas quando estas representações fazem parte dos conjunto de informações sobre o qual planejadores e tomadores de decisão devem trabalhar.

REFERÊNCIAS

- BORGES, K.A.V , LAENDER, A H. F. and DAVIS, C. A. Spatial data integrity constraints in object oriented geographic data modeling In: Proceedings of the seventh ACM international symposium on Advances in Geographic Information Systems (ACM-GIS'99), Kansas City, Missouri, United States, 1999, p. 1-6.
- BURROUGH, P.A.; McDONELL, R.A. Principles of geographical information systems. Oxford: Clarendon Press, 1998. 327p.
- CÂMARA, G., SOUZA, R. C. M.,FREITAS, U. M., GARRIDO, J., MITSUO II, F. A.. SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling. Computer and Graphics, vol. 15, n.6, p.13-22, July, 1996.
- CAMARGO, E.C.G. Desenvolvimento, Implementação e teste de procedimentos geoestatísticos (Krigagem) no sistema de processamento de informações georreferenciadas (SPRING). Dissertação de Mestrado, São José dos Campos, Instituto Nacional de Pesquisas Espaciais, INPE.1997.
- DEUTSCH, C.V.; JOURNEL, A.G. GSLIB: geoestatistical software library and user's guide. New York: Oxford University Press, 1998. 339p.
- EMBRAPA. Levantamento semidetalhado dos solos da bacia rio jardim. Boletim de Pesquisa, n.18, cdrom, 2000.
- Hans Wackernagel, Multivariate Geostatistics (1995)
- FELGUEIRAS, C.A.. Modelagem ambiental com tratamento de incertezas em sistemas de informações geográficas: o paradigma geoestatístico por indicação. Tese de Doutorado, São José dos Campos, Instituto Nacional de Pesquisas Espaciais, INPE. 1999.
- GOOVAERTS, P. Geostatistics for natural resources evaluation. New York: Oxford University Press, 1997. 476p.
- GUERRA, P. A .G. Geoestatística operacional. Brasília: DNPM, 1988. 145p.
- ISAAKS, E.H.; SRIVASTAVA R.M. An introduction to applied geostatistics. New York: Oxford University Press, 1989. 560p.
- PANNATIER, Y., "VARIOWIN: Software for Spatial Data Analysis in 2D," Springer-Verlag, New York, NY, 1996.
- SPRING. 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/> , Versão 3.6, Julho, 2002.
- VARIOWIN. Geostatistical Software for the PC, <http://www-sst.unil.ch/research/variowin/>, Versão 2.21, November, 1999.
- Amílcar Soares – Geoestatística para as ciências da terra e do ambiente (2000)*

FIGURAS



Distribuição das amostras de argila na bacia do Rio Jardim

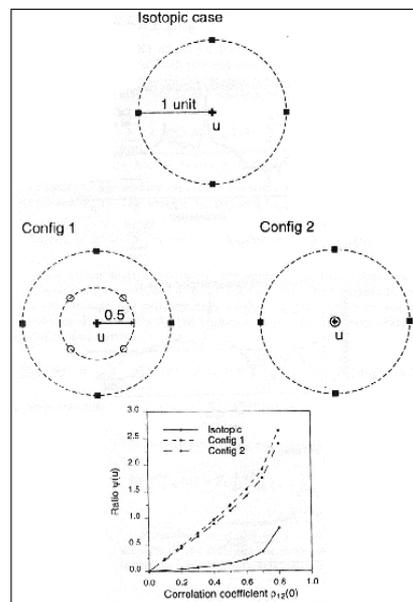


Figura 2. Impacto da densidade da amostragem e configuração de dados na cokrigeagem. O gráfico (parte inferior) mostra que a razão (u) aumenta quando os quatro pares de dados co-localizados primários/secundários são suplementados por quatro dados secundários mais próximos (Config. 1) ou

por um único dado na localização que será estimada (Config. 2). FONTE: Adaptada de Goovaerts, p. 222, 1997.

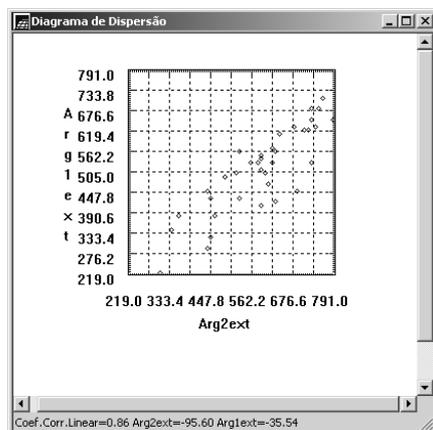


Figura 3. Análise de dispersão das variáveis de argila no horizonte superficial e subsequente, mostrando coeficiente de correlação 0.86, entre as variáveis.

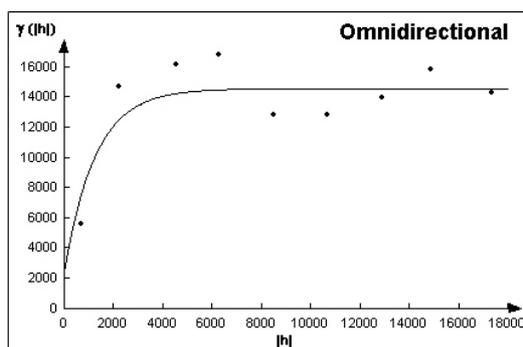


Figura 4. Variograma de amostras de argila, do horizonte superficial, ajustado com o modelo exponencial.

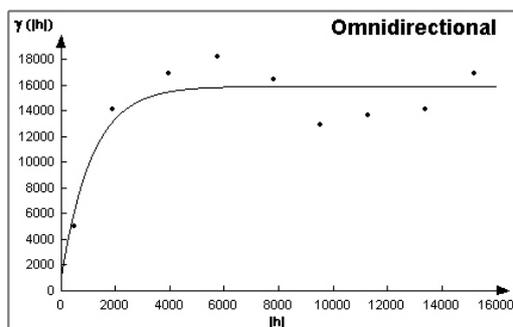


Figura 5. Variograma de amostras de argila, do horizonte subsequente, ajustado com o modelo exponencial.

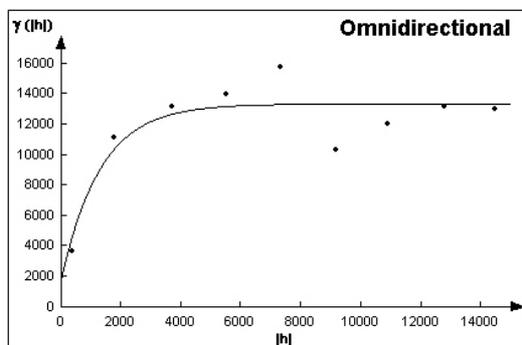
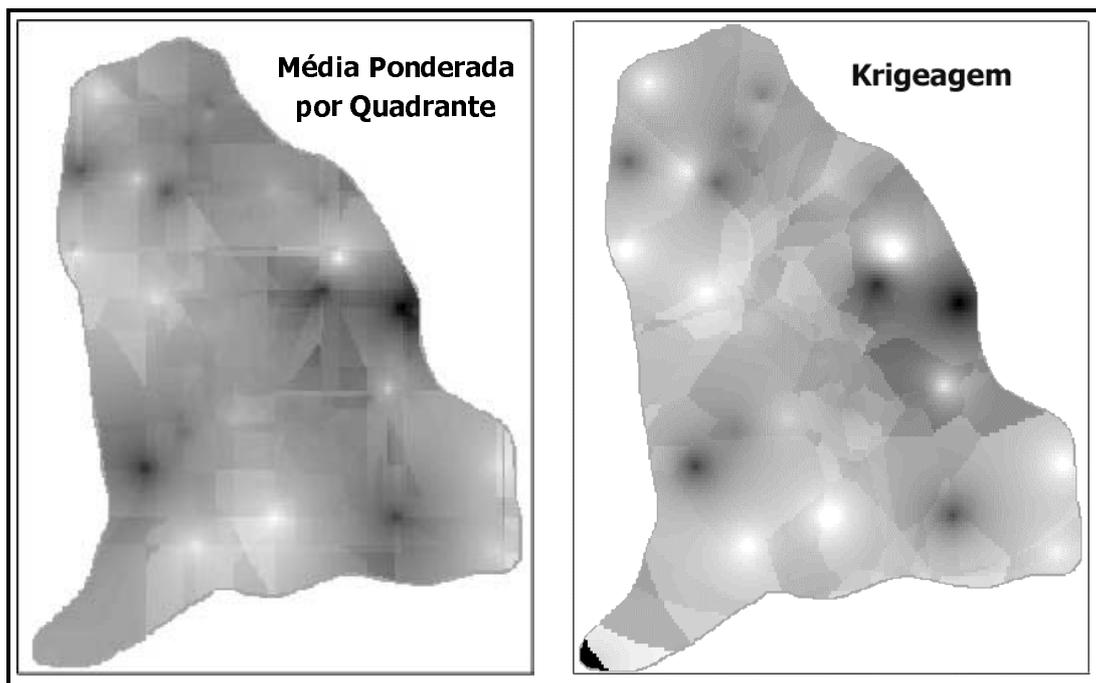


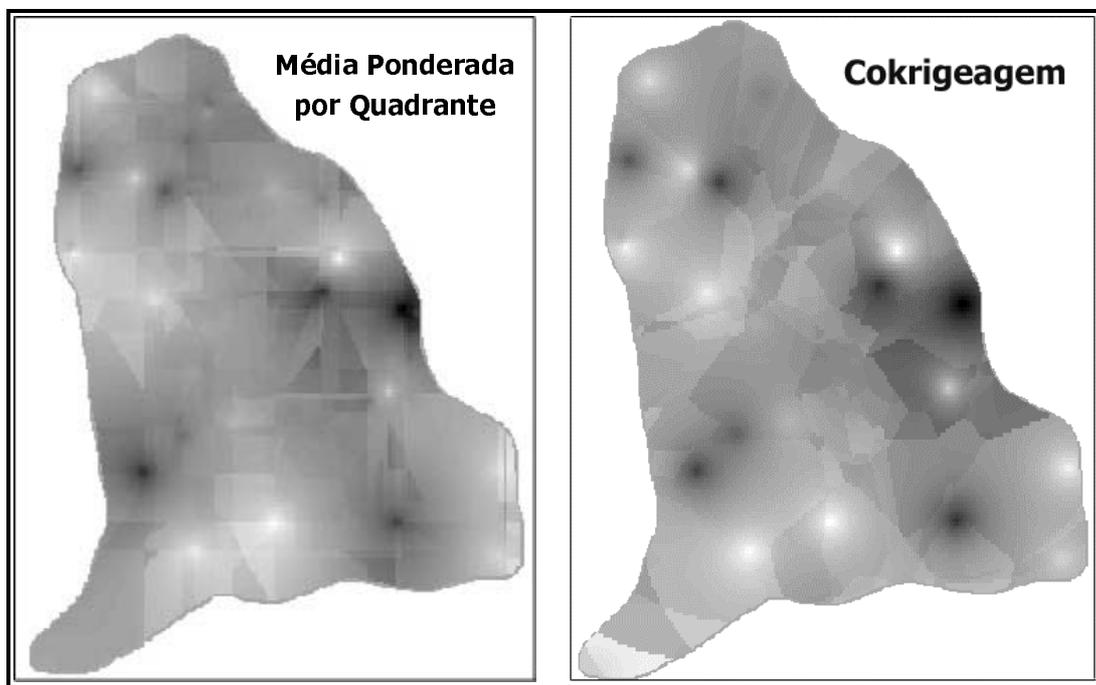
Figura 6. Semivariograma cruzado das duas amostras de argila.



(a)

(b)

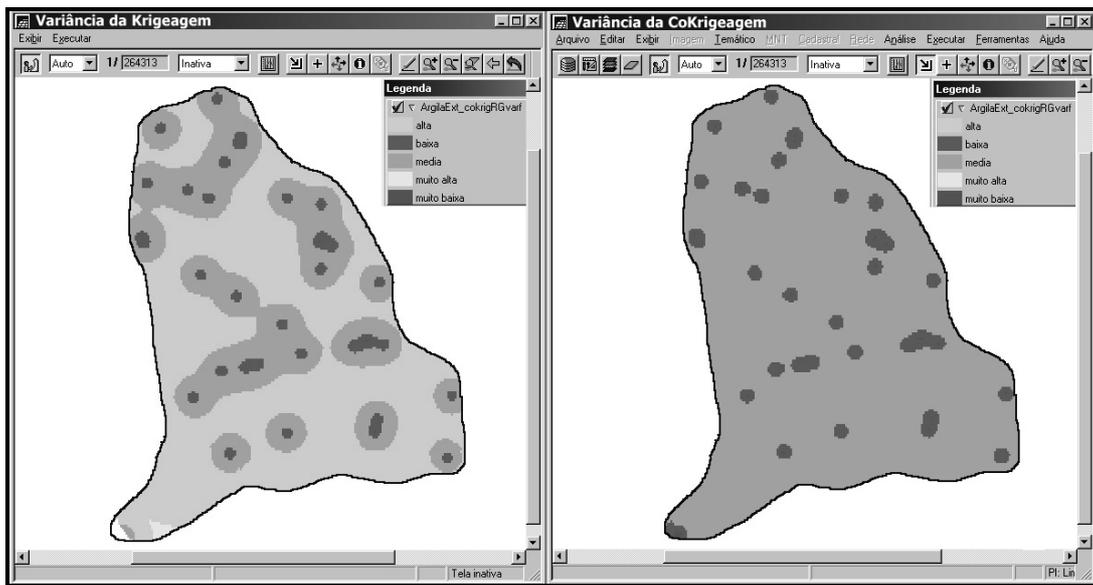
Figura 7 Imagens geradas com os interpoladores: (a) Média ponderada pelo inverso das distâncias e (b) Krigagem Ordinária.



(a)

(b)

Figura 8. Imagens geradas com os interpoladores de: (a) Média Ponderada pelo inverso das distâncias e (b) Cokrigagem



(a)

(b)

Figura 9. (a) Imagem de variância fatiada, resultante da krigeagem ordinária e (b) Imagem de variância fatiada, resultante da cokrigeagem ordinária.

