

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO  
DIVISÃO DE PROCESSAMENTO DE IMAGENS  
SER-300: INTRODUÇÃO AO GEOPROCESSAMENTO

**Laboratório V: Geoestatística**

Henrique Luis Godinho Cassol

São José dos Campos, Maio de 2014.

## 1. INTRODUÇÃO

A maioria das variáveis aleatórias (v.a) ambientais ou naturais é tratada pela estatística clássica como variáveis independentes, ou seja, possuem a mesma probabilidade de ocorrência em certo evento aleatório independente da posição em que estas se encontram no espaço. Este pressuposto, quando aplicado às v.a naturais é importante, pois exclui o efeito bias ou erro sistemático (ANDRIOTTI, 2003).

Porém, sabe-se que além de independência do fenômeno, há uma forte correlação entre as amostras adjacentes. Isto levou Matheron, em 1970, a definir o que chamou de teoria das variáveis regionalizadas ou em outras palavras, a medida da continuidade espacial do fenômeno (SOARES, 2000). Sendo assim, espera-se que próxima às amostras que apresentem altos valores em determinada variável obtenha-se também amostras com valores elevados, que será tanto mais influente quanto maior for a sua continuidade espacial.

Embora os estimadores clássicos como a média local e o inverso do quadrado da distância preveem esta restrição, estes estimadores não levam em consideração a anisotropia espacial das amostras, ou seja, as tendências que certa variável aleatória podem ter em diferentes direções.

Portanto, a teoria das variáveis regionalizadas fundamentada por Matheron em 1970, considera que se devem atribuir pesos às diferentes amostras para estimar os valores desconhecidos, perante análise de variogramas e determinação da anisotropia.

Esta teoria, juntamente com a técnica de Krigeagem, compõe o que se conhece por geoestatística. A Krigeagem é a técnica que serve para predizer os valores de uma variável em locais não amostrados. Embora pareça um método de interpolação, a Krigeagem difere deste por ser o único que confere ao analista a estimativa de precisão e de avaliação do erro.

A geoestatística, contudo, não segue um modelo único, pois varia conforme a interpretação do usuário. O procedimento de ajuste dos modelos

não é direto, como no caso de uma regressão, mas sim iterativo, até que se alcance a precisão desejada.

O objetivo deste trabalho será estimar as concentrações de argila no solo em locais não amostrados utilizando-se dos conceitos de geoestatística e da técnica de Krigeagem, bem como realizar uma avaliação do erro e dos resultados obtidos.

## **2. MATERIAIS E MÉTODOS**

Os dados utilizados, de propriedade do Centro Nacional de Pesquisas de Solos (CNPS - RJ), foram obtidos no levantamento dos solos da Fazenda Canchim, em São Carlos - SP. Estes se referem a uma amostragem de 85 observações georreferenciadas coletadas no horizonte Bw (camada do solo com profundidade média de 1m). Dentre as variáveis disponíveis, selecionou-se para estudo o teor de argila.

Considera-se o teor de argila ao longo do perfil, classificado do seguinte modo (Calderano Filho et al., 1996):

- MUITO ARGILOSO: solos que apresentam 59% ou mais de argila;
- ARGILOSO: solos que apresentam de 35% a 59% de argila;
- MÉDIO: solos que apresentam de 15% a 35% de argila;
- ARENOSO: solos que apresentam menos de 15% de argila.

### **2.1. Metodologia**

Os arquivos contendo os pontos amostrais foram inseridos e processados no SPRING, versão 5.3.2, no formato de dados MNT, disponível pelo software.

As atividades desenvolvidas foram:

- a) a análise exploratória dos dados;
- b) a análise estrutural da variabilidade espacial (cálculo e modelagem do semivariograma);

c) realização das inferências (Krigeagem e Simulação).

### 3. RESULTADOS

Na Tabela 1 estão descritas as estatísticas descritivas das 85 amostras de teor de argila. Nota-se que as amostras seguem uma distribuição próxima a normal, com valores médios e medianos semelhantes.

Tabela 1. Estatísticas descritivas do teor de argila

	N	x	s	CV (%)	Mín	Mediana	Máx
Argila	85	33,05	16,97	51,37	4	33	73

Na Figura 1a e 1b estão apresentados o histograma de frequência e o gráfico de probabilidade normal. Pela Fig 1a e 1b percebe-se que os dados aproximam de uma distribuição normal, porém com uma assimetria acentuada para os valores abaixo da média, à esquerda (coeficiente de assimetria = 0,214).

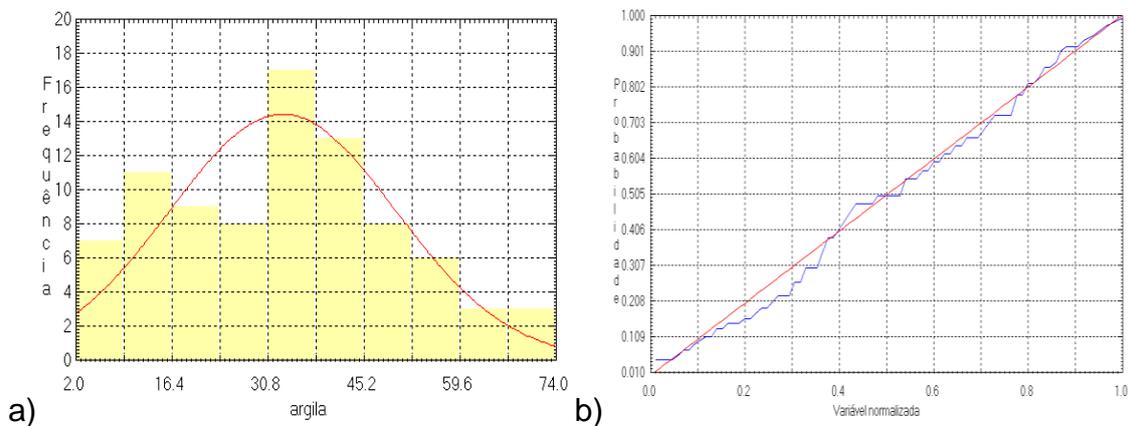


Figura 1. a) Histograma das frequências com 10 classes; b) Gráfico de distribuição de probabilidade normal. Em vermelho a reta de distribuição normal.

#### 3.1. Caso Isotrópico

O semivariograma omnidirecional auxilia o analista a observar se há anisotropia nos dados. O cálculo do semivariograma omnidirecional consiste em usar uma tolerância angular de  $90^\circ$ , ou seja, testa a presença de continuidade espacial em todas as direções (SOARES, 2000).

Em fenômenos aparentemente isotrópicos o semivariograma omnidirecional formado será uma linha reta horizontal (CAMARGO, 2011). Chama-se de efeito pepita puro quando não há continuidade espacial nos dados. Por isto, neste caso, a geoestatística não é a técnica adequada para inferir sobre a variabilidade espacial.

Em caso de presença de anisotropia, o semivariograma omnidirecional terá aspecto de uma função monotônica crescente (SILVA, 2000), aumentando a variabilidade do fenômeno à medida que se aumentam as distâncias entre amostras. O semivariograma omnidirecional também permite ao analista verificar o efeito pepita inicial (SOARES, 2000).

Na Figura 2, observa-se o semivariograma omnidirecional da variável Argila. Pode-se notar que os valores de entrada iniciais não se ajustam bem aos dados, sendo necessário repetir a tarefa com novos parâmetros de entrada.

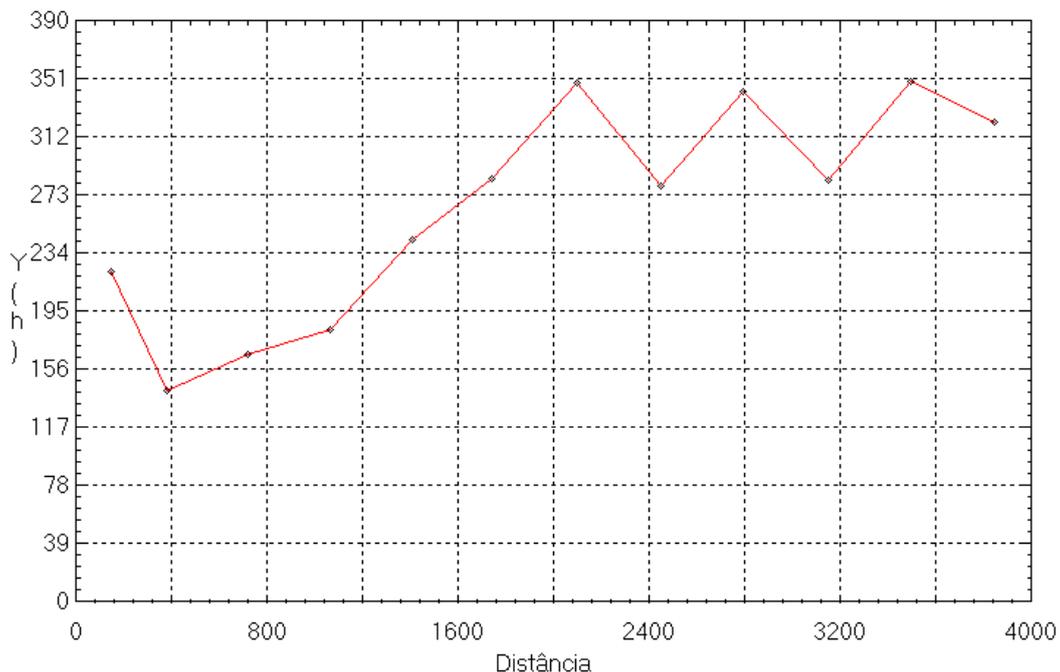


Figura 2. Semivariograma omnidirecional experimental gerado por *default*.

Dessa forma, novos parâmetros de entrada testados foram:

- Número de Lag (4): Representa quantos passos serão usados para calcular os pares de amostras

- Distância Lag (968): Lag ou h é a distância de separação entre amostras. Neste caso utilizou-se o valor inicial de 968m, devido a este ser uma distância média entre amostras;
- Tolerância Lag (484): Usualmente representa a metade de Lag;

O resultado do semivariograma omnidirecional experimental está apresentado na Figura 3.

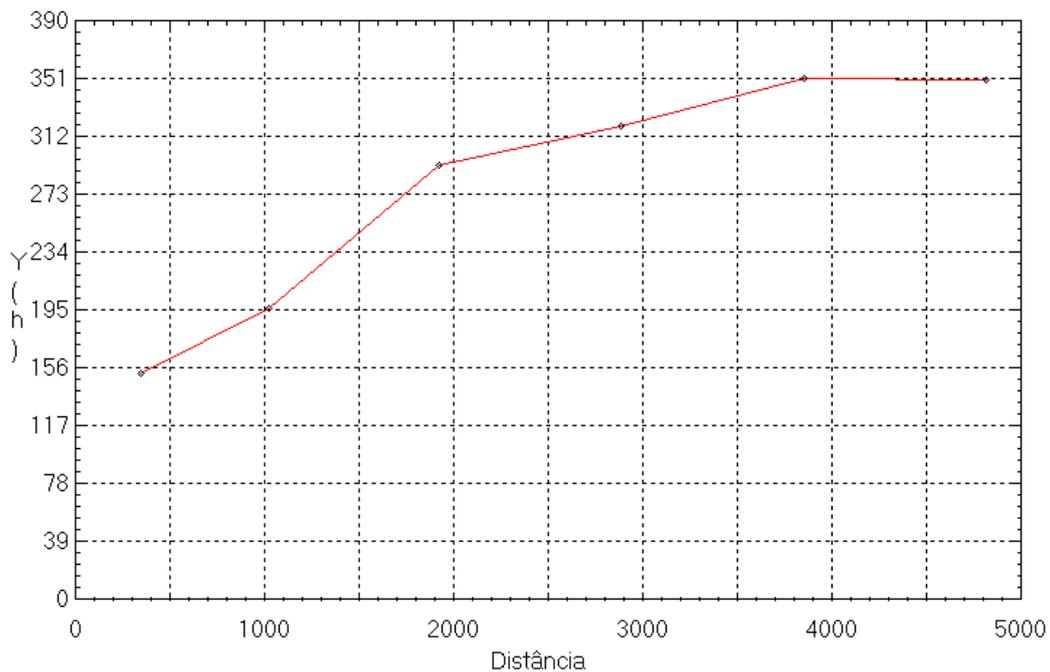


Figura 3. Semivariograma omnidirecional experimental.

O próximo passo é ajustar um modelo teórico que se ajuste ao semivariograma experimental. Isaaks e Srivastava (1989) propuseram uma gama de modelos teóricos que são utilizados segundo o comportamento da variável, a exemplo, tem-se o modelo esférico, exponencial, gaussiano, linear, potencial, etc (Fig. 4).

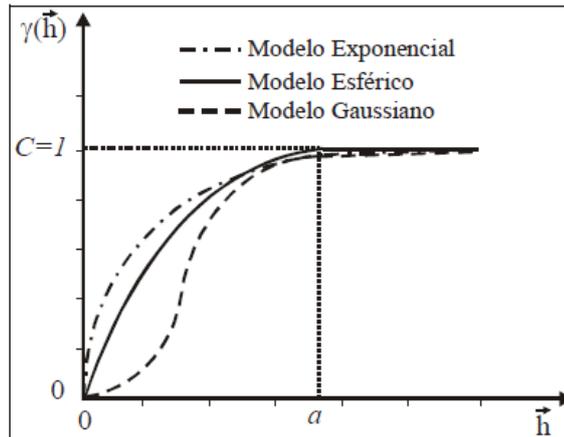


Figura 4 - Representação dos principais modelos de variogramas com o mesmo alcance. Fonte: Camargo (2011).

O ajuste dos variogramas experimentais aos modelos teóricos é feito de modo iterativo, testando-se os melhores parâmetros (efeito pepita, alcance, patamar) até atingir um nível satisfatório. Entretanto, é importante que o modelo ajustado represente a tendência de  $[\gamma(h)]$  em relação ao  $h$  para que as estimativas obtidas pela Krigagem sejam mais exatas e confiáveis.

Assim, o modelo teórico que melhor se ajustou aos dados foi o gaussiano, com os seguintes dados de entrada:

- Efeito pepita (118,854): Valor que se ajusta bem ao modelo e às médias dos pontos entre diferentes valores de lag ( $h$ ). Este valor representa 41% da variabilidade total dos pontos amostrais ( $\sigma^2 = 288,03 \mu\text{M}^2$ ). Isto implica que há uma incerteza significativa nos valores amostrados;
- Alcance (3989,20): Valor referente ao valor de continuidade espacial máxima.
- Contribuição (230,892): A contribuição é a diferença entre o patamar ( $C$ ) e o efeito pepita ( $C = 349,746 - 118,854$ );

### 3.1.1. Validação do modelo

A validação do modelo é feita por meio da análise visual de dispersão dos erros, pelo histograma de frequência de erros e principalmente pela validação cruzada dos erros.

A validação cruzada trata-se de uma técnica da estatística não paramétrica a um conjunto de dados  $Z(x\alpha)$ . O processo consiste em retirar-se uma amostra  $Z(x_0)$  do conjunto de dados e estimar-se mediante a Krigeagem o valor da amostra  $Z(x_0)^*$ . Do conjunto dos valores reais, valores estimados  $[Z(x_0), Z(x_0)^*]$ , obtidos para todas as amostras  $\alpha = 1, 2, \dots, N$ , calculam-se as estatísticas básicas – média e desvios – com o objetivo de aferir a qualidade do modelo escolhido para o semivariograma (DAVIS, 1987).

Espera-se que um bom preditor de Krigeagem contenha uma relação entre os valores estimados e os reais o mais próximo quanto possível de um, estabelecendo uma alta relação. Nota-se pela Figura 5 que há uma boa correspondência entre os valores preditos e os valores observados.

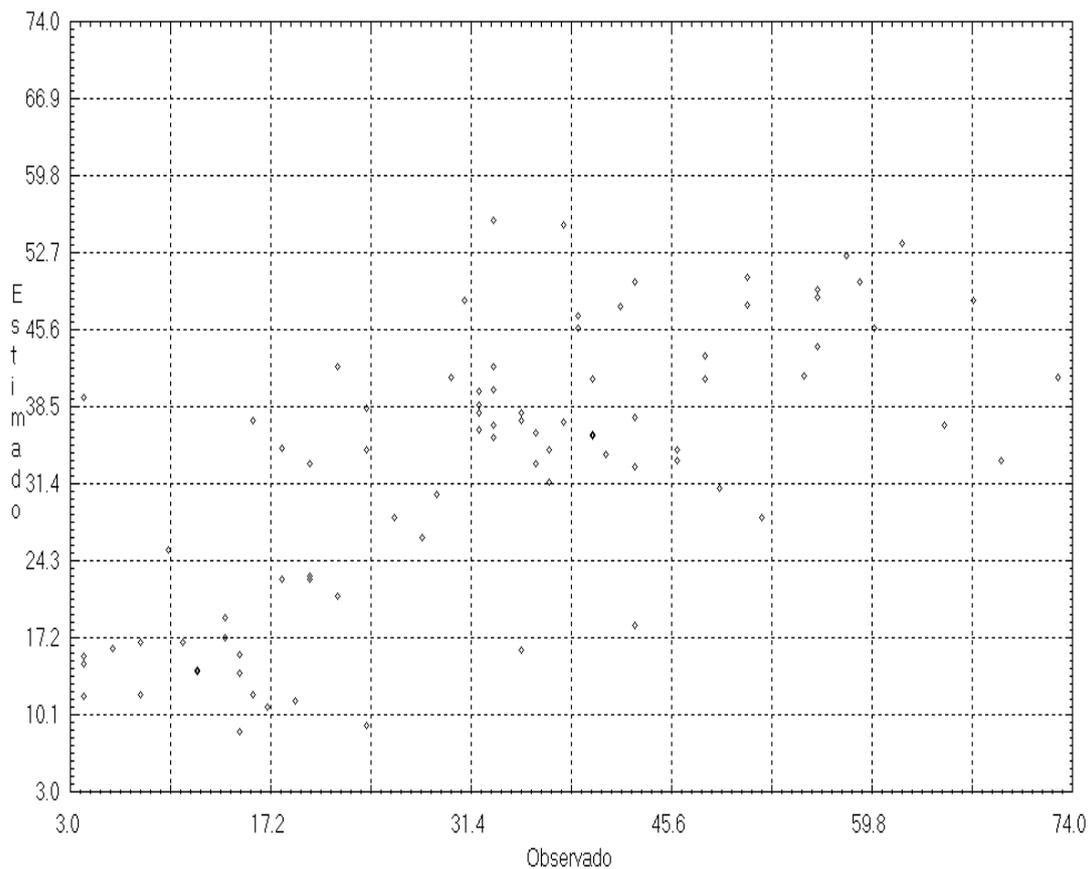


Figura 5 . Validação cruzada da variável argila.

Outra forma de verificar a consistência dos dados é plotando um histograma dos erros. Estes erros serão menores quanto mais próximos de

uma média nula, ou seja, a média da diferença entre os valores estimados e reais tem de ser igual a zero, implicando a não tendenciosidade das estimações (SOARES, 2000). Verifica-se que os erros estão centrados em zero (Figura 6a). Na Figura 6b estão apresentados os erros dispersos por cada ponto amostral. As cruces maiores apresentam os maiores erros. Neste caso não há tendência a concentração de erros por região.

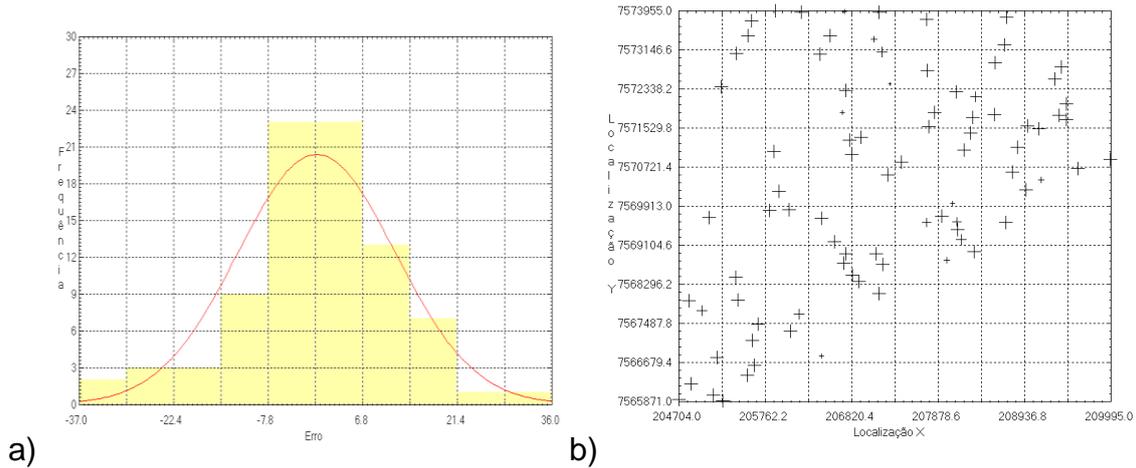


Figura 6. a) Histograma dos erros; b) Dispersão dos erros.

### 3.1.2. Krigeagem Ordinária

O estimador linear geoestatístico  $[Z(x_0)]^* = \sum_{i=1}^N \lambda_i Z(x_i)$ , denominado Krigeagem Normal ou Ordinária é uma combinação linear do conjunto de N variáveis vizinhas de  $x_0 - Z(x_i)$ ,  $i = 1, \dots, N$ , que cumpre com os critérios de não-enviesamento e variância mínima em relação ao erro de estimação  $\varepsilon(x_0)$ .

Segundo Soares (2000), os ponderadores Geoestatísticos  $\lambda_i$  devem inferir sobre dois aspectos espaciais: em primeiro lugar devem refletir sobre a proximidade estrutural das amostras  $Z(x_i)$  em relação ao ponto a se estimar  $Z(x_0)$  e, também, devem ter efeito desagregador em amostras agregadas (clusters) para evitar seu enviesamento. Com isso, a geoestatística está formulada sob duas premissas de qualidade: o não-enviesamento  $[E \{\varepsilon(x_0)\} = 0; E \{Z(x_0)\} = E \{Z(x_i)\} = m]$  e a variância de estimação mínima  $E \{[\varepsilon(x_0)]^2\} = \min$ .

Como visto anteriormente o primeiro critério – não enviesamento – é conseguido pela imposição da condição aos ponderadores de que sua soma seja igual a um. A minimização da variância é assegurada pelo procedimento clássico de se igualar a zero as N derivadas parciais em relação aos pesos e resolver o sistema de N equações e N incógnitas por um método qualquer, a exemplo do Método dos Mínimos Quadrados – MMQ (SOARES, 2000).

No entanto, para cumprir a condição de que a soma dos pesos seja igual a um, a solução de minimização da variância é resolvida por intermédio do formalismo de Lagrange, que implica adicionar mais uma equação e uma incógnita – o parâmetro de Lagrange  $\mu$  – à equação.

Neste método, o procedimento de interpolação refere-se aos pesos dados às amostras segundo análise do variograma. Conforme Soares (2000), quanto mais próximas as amostras estiverem do ponto a estimar maiores serão seus pesos no estimador e; quanto mais correlacionadas estiverem as amostras maior será seu efeito de agrupamento e menor será seu peso no estimador:

$$[\lambda] = [K]^{-1} \cdot [M]$$

$[K]^{-1}$  – é o fator de desagregação originado pela matriz de covariâncias entre as amostras (primeiro membro do sistema de Krigeagem);

$[M]$  – é a distância estrutural entre as amostras (segundo membro do sistema de Krigeagem);

O resultado da Krigeagem ordinária está representado na Figura 7. Nota-se que os maiores valores estão concentrados em torno dos pontos amostrais. Há que se considerar que os dados foram modelados conforme um modelo isotrópico, que foi visto pelo semivariograma experimental que não é verdadeiro. Assim, os dados foram testados para um modelo anisotrópico dos dados.

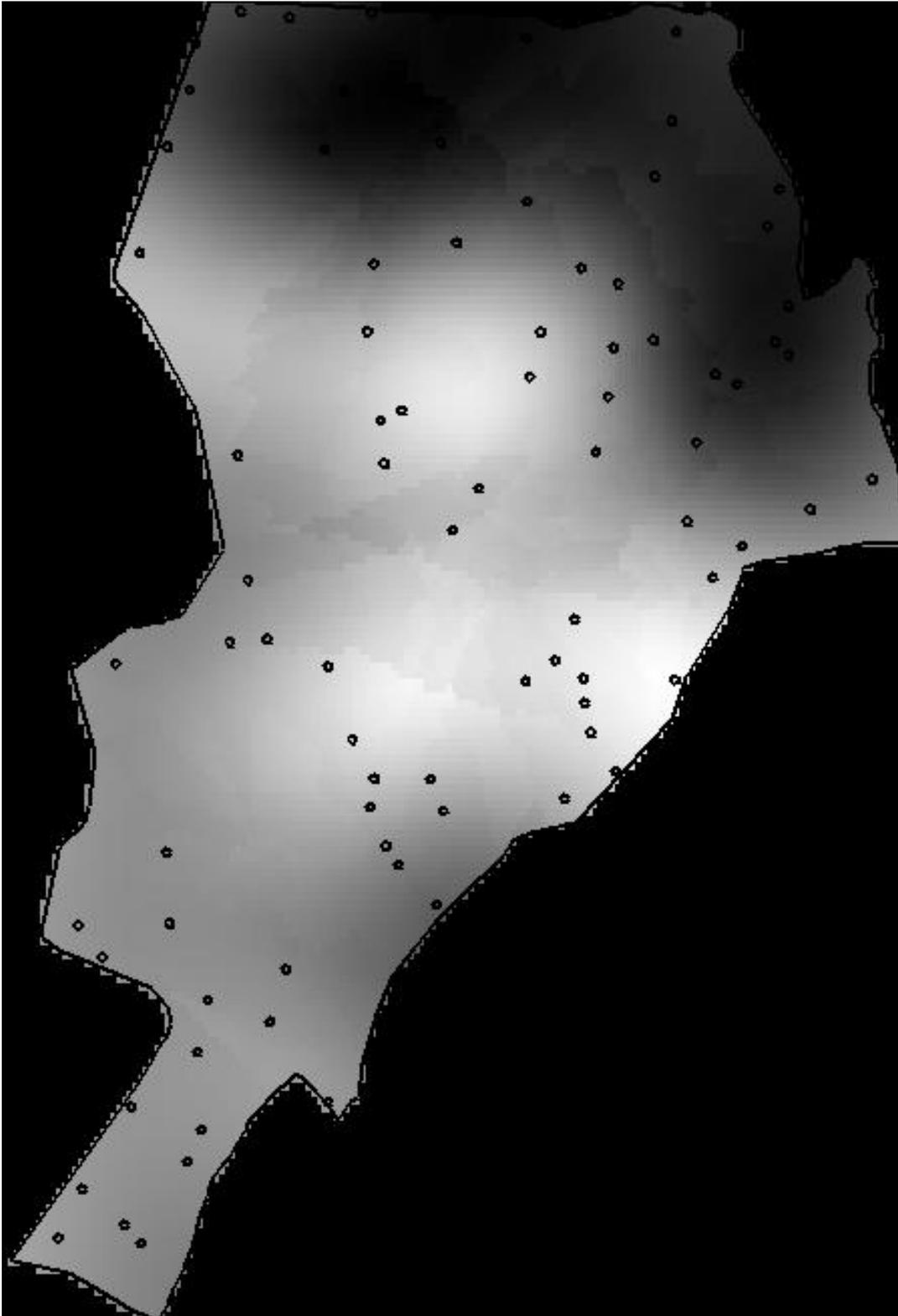


Figura 7. Mapa de teor de argila gerado pela krigagem ordinária do caso isotrópico.

### 3.2. Caso Anisotrópico

A continuidade espacial do fenômeno estudado pode variar em diferentes direções. Segundo Soares (2000), uma dada característica de um recurso natural é considerada isotrópica quando o variograma tem o mesmo comportamento em todas as direções, ou seja,  $\gamma(h)$  depende somente do módulo do vetor  $h$ .

Na prática, tal situação é rara, senão inexistente. Dessa forma, a anisotropia é definida como a função de continuidade espacial de uma variável regionalizada que se estende para uma determinada direção particular. Guerra, (1988); Soares (2000) descrevem dois modelos básicos de anisotropia. A anisotropia geométrica e a anisotropia zonal.

A anisotropia geométrica é um modelo no qual a continuidade espacial apresenta o mesmo patamar quando o variograma for constituído em diferentes direções, porém com diferentes alcances. Estes alcances são máximos e mínimos em direções perpendiculares. A anisotropia geométrica significa que os alcances nas várias direções podem ser modelados por uma elipse.

A anisotropia zonal, por outro lado, ocorre quando o patamar é diferente para cada uma das distintas direções. Esta anisotropia é típica de fenômenos estratificados.

Quando o fenômeno é anisotrópico, o aumento da tolerância angular para o cálculo do semivariograma pode resultar na atenuação do alcance e implica a sobrestimação dos alcances das direções de menor continuidade. De modo idêntico, altas tolerâncias de lag ( $h$ ) para as primeiras distâncias podem conduzir à sobrestimação do efeito pepita (SOARES, 2000).

Na Figura 8 tem-se o semivariograma de superfície. O semivariograma de superfície serve para nos indicar os eixos de maior e menor descontinuidade espacial do fenômeno.

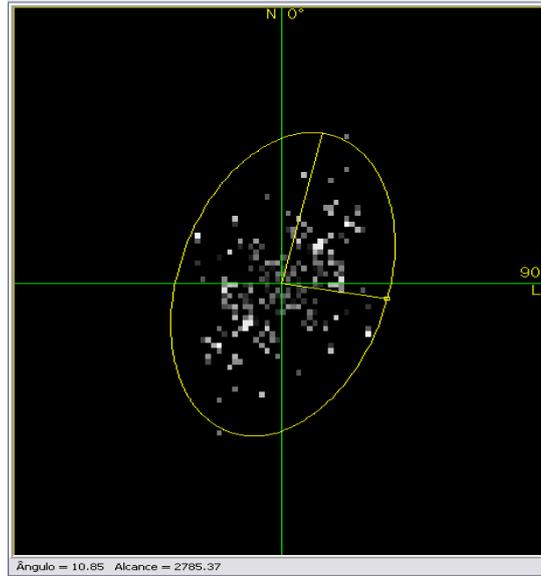


Figura 8. Semivariograma de superfície.

Nota-se pela Fig 8 que há dois eixos de variabilidade espacial, uma maior variabilidade em torno de  $17^\circ$  e uma menor, ortogonalmente em  $107^\circ$ . Na Figura 9 tem-se os semivariogramas direcionais com os respectivos ângulos de anisotropia. Em verde, o eixo de maior variabilidade ( $17^\circ$ ), em vermelho, o semivariograma omnidirecional ( $90^\circ$ ) e em azul o semivariograma do eixo de menor variabilidade ( $107^\circ$ ).

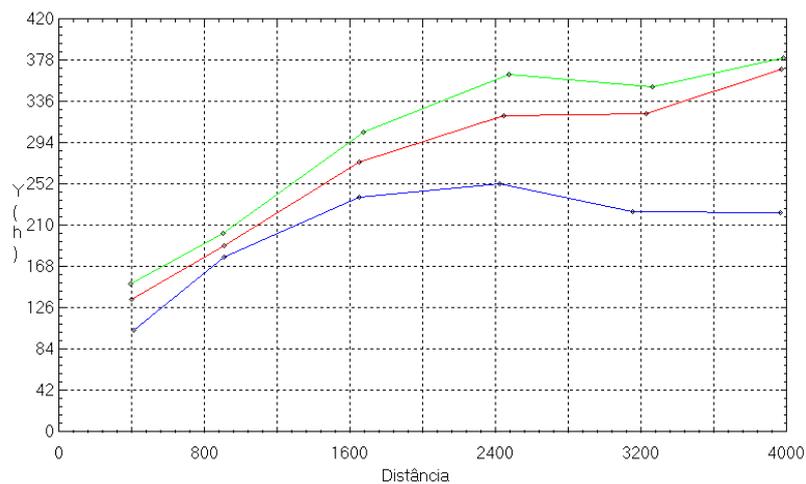


Figura 9. Semivariogramas direcionais, com anisotropia. Em verde o eixo de maior variabilidade ( $17^\circ$ ), em vermelho o semivariograma omnidirecional ( $90^\circ$ ) e em azul o semivariograma do eixo de menor variabilidade ( $107^\circ$ ).

O modelo de ajuste consistiu em três estruturas esféricas conforme a Figura 10, retirados dos ajustes de anisotropia dos dados. Neste caso, denomina-se anisotropia combinada a idéia básica para modelar este tipo de anisotropia é dividir em faixas convenientes o gráfico de semivariogramas, conforme ilustra a Figura abaixo, de maneira que, em cada faixa reste somente a anisotropia geométrica.

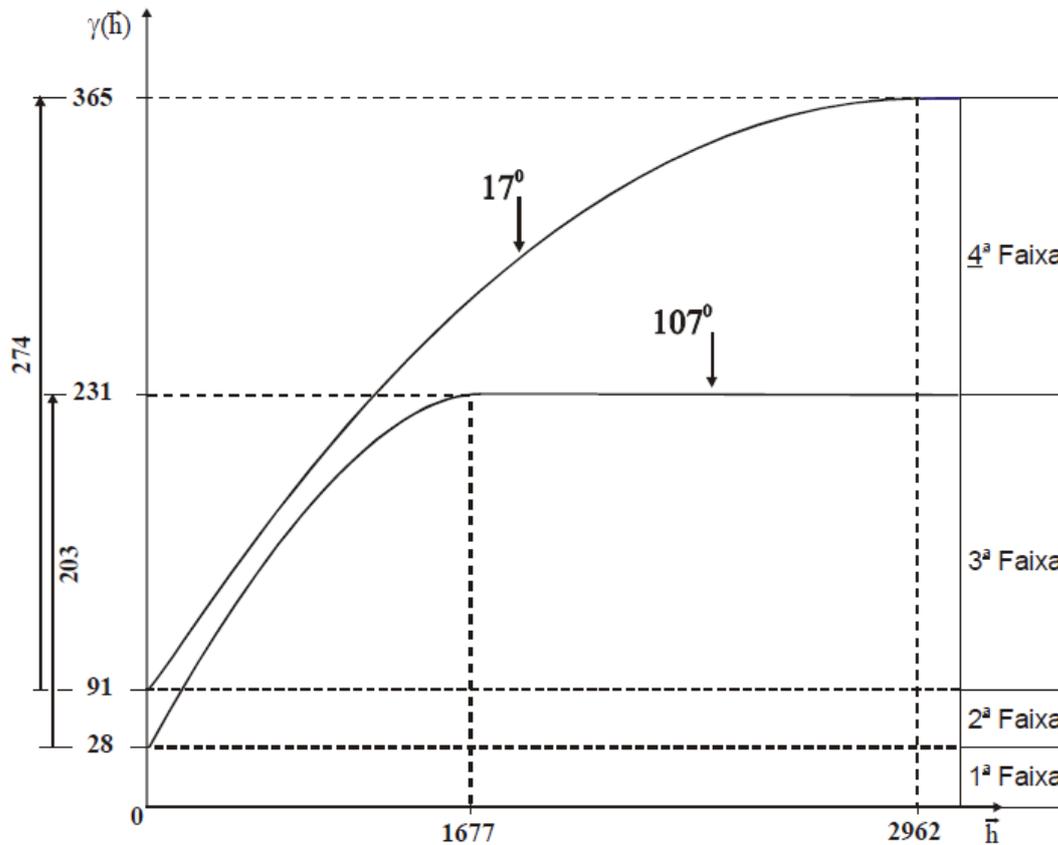


Figura 10. Estrutura anisotrópica do semivariograma direcional.

### 3.2.1. Erros de estimação

Na Figura 11 está apresentado o histograma de frequências dos erros gerados pelo modelo anisotrópico. Comparando-o ao histograma dos erros do caso isotrópico, verifica-se que este segue uma normal mais comportada e com os erros mais centrados em zero (Tabela 2).

Entretanto, a variância e, principalmente, o coeficiente de variação foram superiores ao do caso isotrópico.

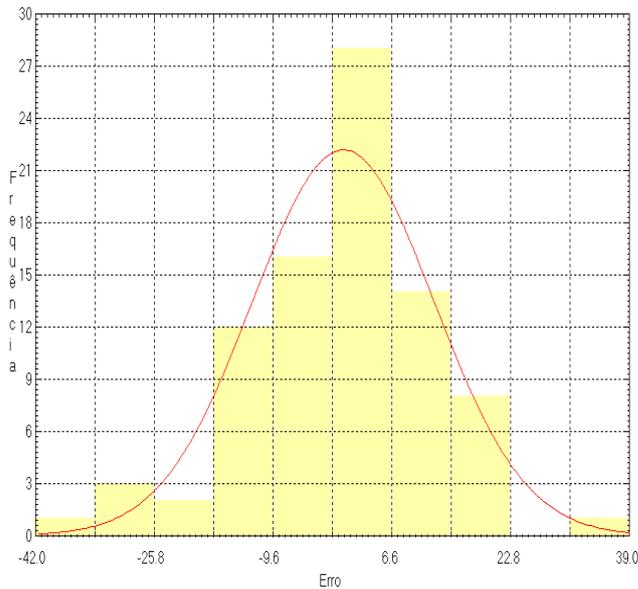


Tabela 2. Estatísticas descritivas dos erros

	Erros_Iso	Erros_Anisi
N	85	85
Média	-0,31	-0,02
Variância	147,93	153,72
Desvio P.	12,16	12,40
CV (%)	39,22	676,71
Assimetria	-0,31	-0,42
Curtose	3,83	4,26
Mínimo	-35,48	-40,57
Máximo	35,28	37,42

Figura 11. Histograma dos erros da krigeagem anisotrópica.

Nas Figuras 12a e 12b estão destacados a dispersão dos erros e a validação cruzada do caso anisotrópico, respectivamente. No caso do diagrama de dispersão dos erros, verifica-se que os erros foram levemente inferiores, sendo a validação cruzada bastante semelhante ao do caso isotrópico.

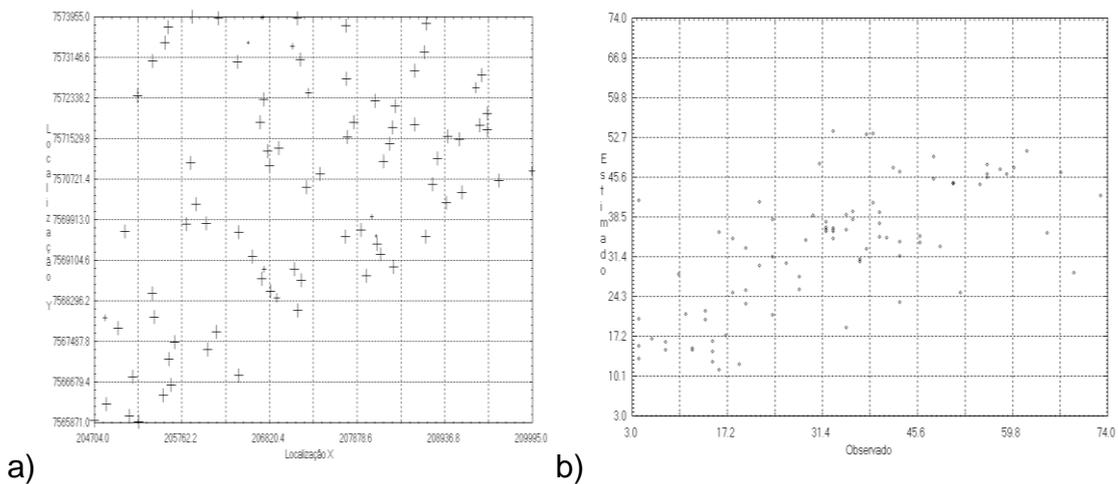


Figura 12. a) Dispersão dos erros; b) Validação cruzada do caso anisotrópico.

Na Figura 13 tem-se o resultado da krigeagem ordinária com três estruturas de semivariograma e dois eixos de anisotropia. Em Fig 13a tem-se o mapa de teor de argila grado pela KO; em Fig 13b, a variância da KO e em Fig

13c, o fatiamento das classes de teor de argila, variando de arenoso (marrom-claro) a muito argiloso (marrom-escuro).

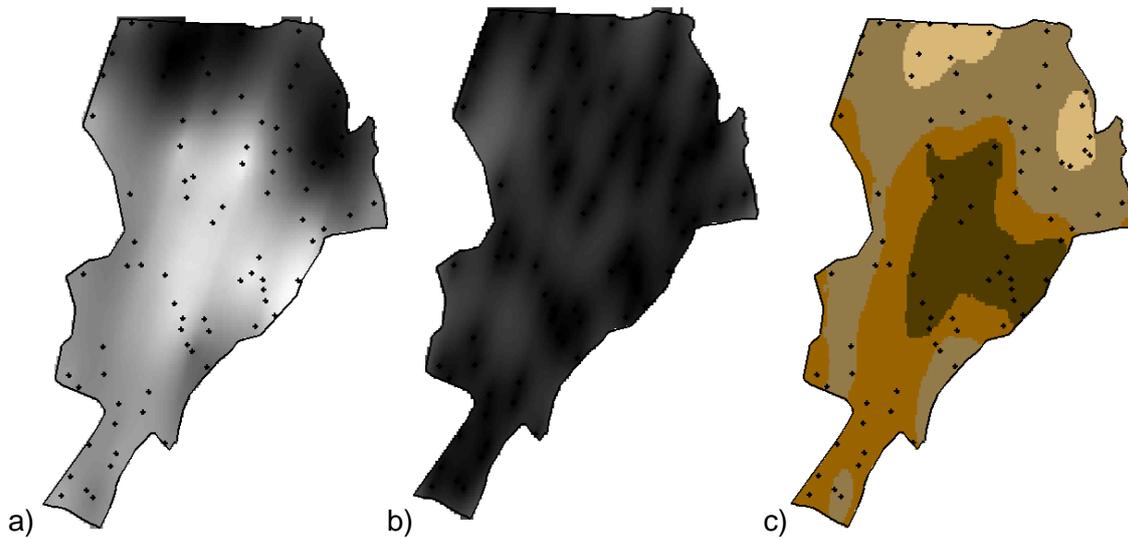


Figura 13. a) Mapa de teor de argila gerado pela KO; b) Variância da KO; c) Classes de teor de argila obtidos a partir do mapa de teor da KO.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRIOTTI, J. L. S. 2003. **Fundamentos de Estatística e Geoestatística**. Editora Unisinos, São Leopoldo, 165 p.

CAMARGO, E. C. G. **Geoprocessamento para projetos ambientais**. Instituto Nacional de Pesquisas Espaciais – INPE. Capítulo 5 – Geoestatística: Fundamentos e Aplicações. 2ª Edição revisada e ampliada. São José dos Campos, 1998. Disponível em: [http://www.dpi.inpe.br/gilberto/tutoriais/gis\\_ambiente/5geoest.pdf](http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf). Acesso em 27 de Junho de 2011.

ISAAKS, E. H.; SRIVASTAVA R. M. **An Introduction to Applied Geostatistics**. New York, Oxford University Press, 1989. 560p.

SOARES, A. **Geoestatística para ciências da terra e do ambiente**. Lisboa: IST Press. 2000. 206p.