Geoestatística:

Considerações Iniciais

É impossível saber, antes de amostrar, de que maneira os valores das variáveis irão se comportar: se dependente ou independente uma da outra.

Devido as limitações da estatística clássica e pelo fato de que muitas variáveis são heterogêneas (os atributos variam no espaço e no tempo), torna-se necessária a utilização de procedimentos estatísticos adicionais, que considerem e reflitam essa variações.

Assumindo as hipóteses exigidas pela estatística clássica, pode-se dizer que:

um **valor medido** é em parte explicado por uma **média** e

em parte pela **variação do acaso**.

Geoestatística:

Considerações Iniciais

Os desvios dos valores em torno da média são assumidos como sendo **independentes** e com **distribuição normal de média zero** e **variância** σ^2

ou seja

a **média aritmética** dos dados amostrais é adotada como sendo **bom estimador da posição central dos valores da população**.

A **média** é então tomada como estimativa da propriedade em **locais não amostrados**, tornando necessário identificar o nível de precisão dessa média como estimador, o que, na estatística clássica, é realizado por meio das **medidas de dispersão**.

Quando é verificado que **a componente residual da variância é relativamente muito grande**, o que normalmente é indicado por um **alto CV**, o experimento ficar prejudicado, sendo que **a causa pode ser a variabilidade do conjunto de dados**, assumido como homogêneo no início.

Geoestatística:

Considerações Iniciais

Se a **distribuição espacial das amostras é considerada**, em muitos casos, **será possível tirar vantagem da variabilidade espacial**.

Portanto a estatística clássica e a geoestatística, ou estatística espacial, se complementam. Uma não exclui a outra, e perguntas não respondidas por uma, muitas vezes podem ser respondidas pela outra.

Quando os **dados são abundantes**, os **métodos de interpolação**, em geral produzem valores semelhantes.

Os métodos tradicionais de interpolação espacial como **triangulação** e **método do inverso do quadrado da distância**, no caso de **dados esparsos**, possuem **limitações** na representação da variabilidade espacial, porque **desconsideram a anisotropia** e a **continuidade do fenômeno** que se quer observar.

Geoestatística:

Considerações Iniciais

A sequência de passos em um estudo geoestatístico envolve:

1) Análise exploratória dos dados para que se possa ter a compreensão da natureza espacial da variável;

2) Análise estrutural do conjunto de dados para determinação da correlação espacial ou continuidade dos dados;

3) Elaboração de estimativas por krigagem para estimar valores em pontos não amostrados.

Antes de selecionar a técnica de estimativa mais apropriada devemos responder a algumas questões básicas:

- Queremos uma estimativa global ou local?
- Nosso objetivo é estimar apenas a média ou a distribuição completa dos dados?
- Queremos estimar valores pontuais ou blocos?

O objetivo é determinar o valor do atributo z em posições u do espaço "não amostradas" → z(u).

Geoestatística:

Interpolação

- Atribuição de pesos para as amostras (*weighted average interpolation algorithms*).

- Pesos entre 0 e 1, sendo que o somatório dos pesos deve ser 1.

- Quanto mais perto um dado nó do grid, maior seu peso.

Interpoladores Clássicos

Polígonos: define zonas/áreas de influência, sendo o peso de acordo com a área que corresponde a cada amostra;

Inverso da distância: valor estimado a partir de combinações lineares dos dados vizinhos, com o peso dado pela distância que separa as amostras;

Spline: polinomiais de uma dada ordem que melhor se ajustam aos dados de forma suavizada.

Geoestatística:

Desvantagem dos métodos clássicos

- Não consideram o suporte amostral;
- Não consideram o padrão de variabilidade espacial;
- Não fornecem medida de erro da estimativa.

Vantagem dos métodos clássicos

- Simples;
- Intuitivos;
- Facilmente implementáveis em rotinas computacionais.

Geoestatística:

Estimativa por combinação linear ponderada

A ideia básica é estimarmos um atributo qualquer em uma posição usando:

$$Z^*(u) = \sum_{i=1}^{n} \lambda_i Z(u_i)$$

Em que: *u* refere-se a uma localização qualquer
Z(u)* é o valor estimado nessa localização u
existem *n* dados $Z(u_i)$, i = 1,...,n na circunvizinhança de u
λ_i refere-se aos pesos calculados.

Quais os fatores que devem ser considerados para calcular os pesos?

- Proximidade das amostras;
- Redundância entre dados amostrais;
- Anisotropia;
- Magnitude da continuidade.

Geoestatística:

Estimativa por combinação linear ponderada

A krigagem é um estimador linear a partir da informação disponível, em que temor *z(u)* posições desconhecidas e dados nas posições $z(u_\alpha)$ como realizações da variável regionalizada em estudo.

$$Z^*(u) - \bar{X}(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u)[Z(u_\alpha) - (u_\alpha)]$$

Em que: $Z^*(u)$ = é o valor estimado;

$\bar{X}(u)$ = média do atributo *z(u)*;

$\lambda_\alpha(u)$ = pesos a serem determinados.

Dessa forma, nosso objetivo é determinar os pesos de krigagem $\lambda_\alpha(u)$, tal que:

$$\sigma_k^2(u) = Var\{Z^*(u) - Z(u)\}$$

Geoestatística:

Estimativa por combinação linear ponderada

Essa variância deve ser minimizada sob a condição de não-tendenciosidade.

$$E\{Z^*(u) - Z(u)\} = 0$$

Assim a geoestatística é um conjunto de técnicas estatísticas utilizadas para analisar a variabilidade e estimar valores de uma variável de interesse.

A metodologia proposta pela geoestatística difere da proposta pela estatística clássica, basicamente, na forma de avaliar a variação dos dados.

Enquanto a estatística clássica pressupõe não haver relação entre a variação e a distância entre pontos de amostragem, isto é, as variações são aleatórias no espaço, a geoestatística considera dependência de variação com o espaço de amostragem e que, em parte, essas variações são sistemáticas.

Geoestatística:

Estimativa por combinação linear ponderada

A regionalização da variável é a maneira de identificar e avaliar a estrutura espacial dos atributos estudados.

Uma premissa básica é que em todas as áreas existem regiões mais ricas do que outras, para uma determinada variável.

Ou seja, o valor da variável regionalizada **f(x)** depende da posição espacial **x**.

Na teoria das variáveis regionalizadas, **Z(x)** pode ser definida como uma variável aleatória que assume diferentes valores **Z** em função de **x**.

O que diferencia a **krigagem** de outros métodos de interpolação é a **estimação de uma matriz de covariância espacial** que determina os **pesos** atribuídos às diferentes amostras, o **tratamento da redundância dos dados**, **a vizinhança a ser considerada** no procedimento de inferência e o **erro associado ao valor estimado**.

Geoestatística:

Estimativa por combinação linear ponderada

O conjunto de amostras deve ser representativo do fenômeno que se pretende estudar.

Se **x** representa uma posição em uma, duas ou três dimensões, então o valor da variável **Z**, em **x**, é, dada por (Burrough, 1987):

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon''$$

onde: **m(x)** é uma função determinística que descreve a componente estrutural de **Z** em **x**;

ε'(x) é um termo estocástico, que varia localmente e depende espacialmente de **m(x)**;

ε'' é um ruído aleatório não correlacionado, com distribuição normal com **média zero** e **variância** σ^2.

Geoestatística:

Estimativa por combinação linear ponderada

A hipótese mais comum é chamada ***estacionaridade de 2ª ordem***:

⇒ A componente determinística, **m(x)**, é constante (não há tendências na região).

⇒ A variância das diferenças entre duas amostras depende somente da distância **h** entre elas, isto é:

$$Var[Z(x) - Z(x+h)] = E\{[Z(x) - Z(x+h)]^2\} = 2\gamma(h)$$

γ**(h)** é chamado de semivariância.

Para mostrar a contribuição da semivariância, podemos reescrever a equação como:

$$Z(x) = m(x) + \gamma(h) + \varepsilon''$$

Em outras palavras, como supomos **m(x)** ser constante, a variação local das amostras (e seu relacionamento espacial) pode ser caracterizado pela semivariância γ**(h)**.

Geoestatística:

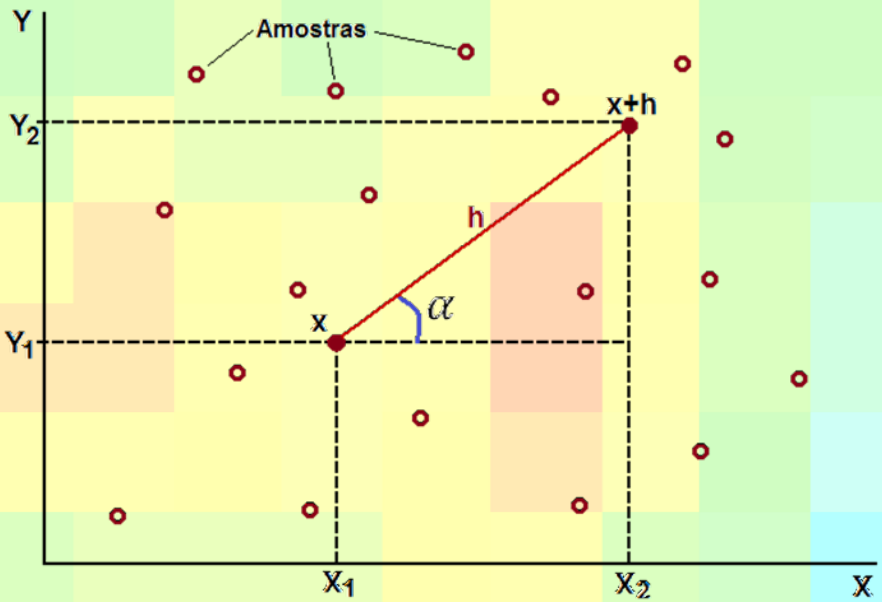
VARIOGRAMA (SEMIVARIOGRAMA)

Quando a amostragem envolve duas direções (x,y) o instrumento mais indicado na estimativa da dependência entre amostras é o variograma (Silva, 1988).

O variograma é uma ferramenta básica de suporte às técnicas de *krigagem*, que permite representar, quantitativamente, a variação de um fenômeno regionalizado no espaço (Huijbregts, 1975).

Considere duas variáveis regionalizadas, X e Y, onde X=Z(x) e Y=Z(x+h).

Neste caso, referem-se ao mesmo atributo (por exemplo, teor de zinco no solo) medido em duas posições diferentes, onde **x** denota uma posição de duas dimensões, com componentes (x,y), e **h** um vetor distância (módulo e direção) que separa os pontos.



Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

O **nível de dependência** entre essas duas variáveis regionalizadas, X e Y, **é representado pelo variograma**, 2γ**(h)**, o qual é definido como a **esperança matemática do quadrado da diferença entre os valores de pontos no espaço**, **separados pelo vetor de distância h**, isto é,

$$2\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} = Var[Z(x) - Z(x+h)]$$

Através de uma amostra $z(x_i)$, i = 1,2,...,n, o variograma pode ser estimado por:

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

onde: $2\hat{\gamma}(h)$ é o variograma estimado;

N(h) e o número de pares de valores medidos, $z(x_i)$ e $z(x_i+h)$, separados por um vetor distância h;

$z(x_i)$ e $z(x_i+h)$ são valores de i-ésima observação da variável regionalizada, coletados nos pontos x_i e x_i+h (i = 1,2,...,n), separados pelo vetor h.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

O variograma é função do vetor h (módulo e direção).

$$\gamma(h) = \frac{1}{2N(h)} \sum \left[v_i - v_j\right]^2$$

Essa equação é também definida como *semivariograma*, porque é dividida por 2.

Para efeito prático, tratar-se-á *variograma* e *semivariograma* com o mesmo significado.

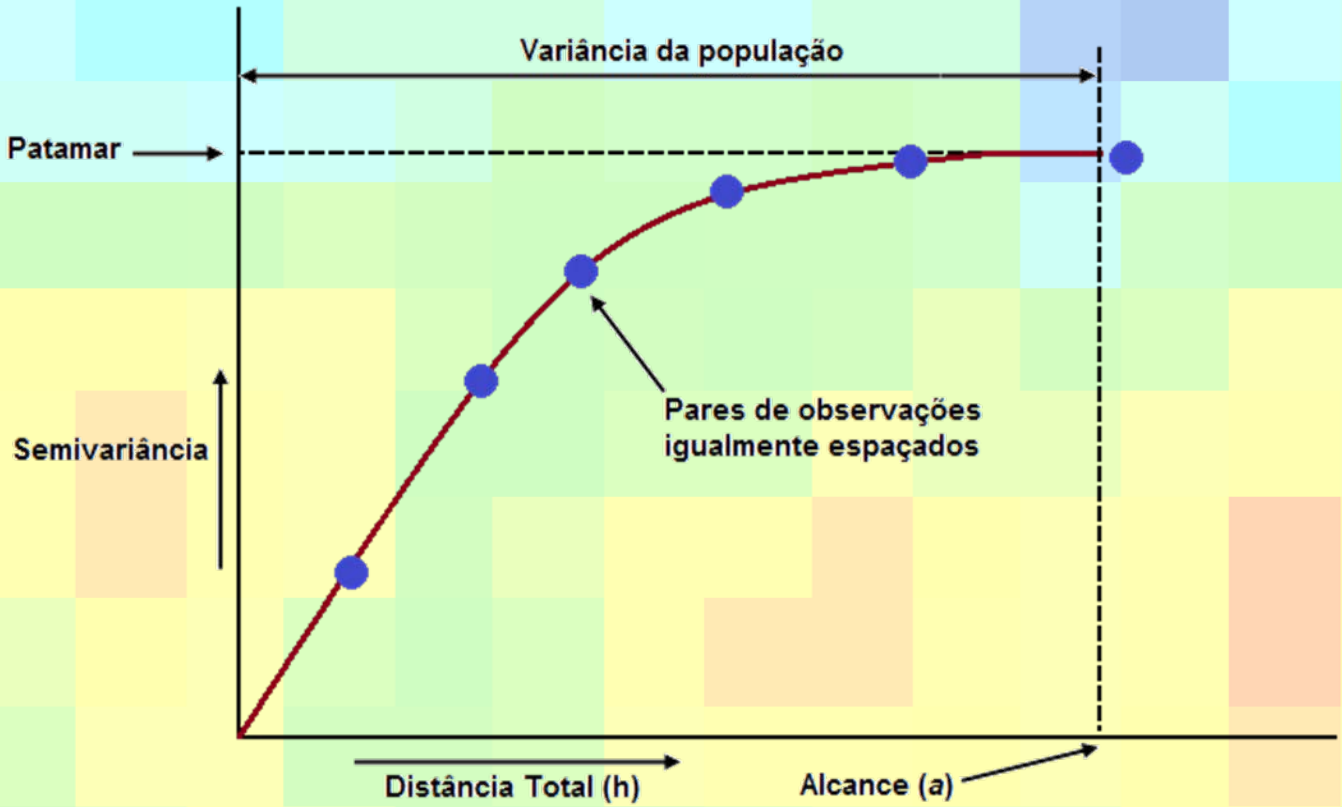
O variograma normalmente atinge o patamar, chamado de ***sill***, o qual é aproximadamente igual a variância total dos valores amostrais (a variância a priori dos dados).

A distância em que o variograma alcança o patamar é chamado alcance (***range***).

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Embora o variograma deva passar pela origem $\gamma(0)=0$, frequentemente ele aparece cortando o eixo vertical em um valor positivo, chamado efeito pepita (***nugget effect***).



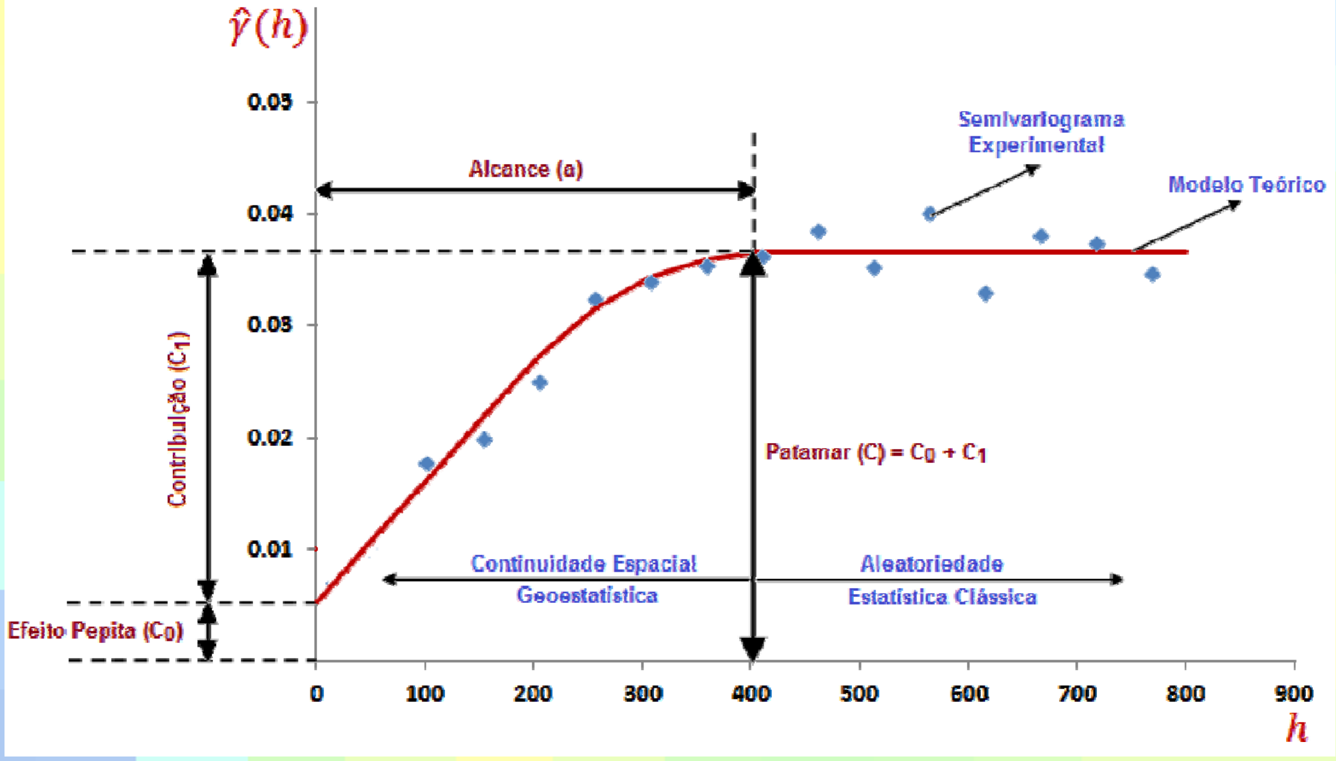
Variograma ideal.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Parâmetros do variograma

O seu padrão represente o que, intuitivamente, se espera de dados de campo, isto é, que as diferenças $\{Z(x_i)-Z(x_i+h)\}$ decresçam, a medida que h, a distância que os separa, decresce.



É esperado que observações mais próximas geograficamente tenha um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias.

$\gamma(h)$ aumenta com a distância h.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Alcance (a): distância, dentro da qual, as amostras apresentam-se correlacionadas espacialmente.

Patamar (C): é o valor do variograma correspondente a seu *alcance (a)*.

<u>Desse ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras torna-s invariante com a distância.</u>

Efeito Pepita (C_0): idealmente igual a zero.

Entretanto, na prática, à medida que **h** tende a 0 (zero), γ(**h**) se aproxima de um valor positivo, (C_0), que revela <u>a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras.</u>

Parte desta descontinuidade pode ser também devida a erros de medição (Isaaks; Srivastava, 1989).

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Contribuição (C_1): é a diferença entre o patamar (**C**) e o efeito pepita (C_0).

Cálculo do semivariograma para amostras regularmente espaçadas

Para determinar o semivariograma experimental, por exemplo, na direção de 90º, o cálculo de $\hat{\gamma}(h)$ é repetido para todos os intervalos de **h**.

Suponha a distância entre dois pontos consecutivos igual a 100m.

Então, qualquer par de observações, na direção de 90º, cuja distância é igual a 100m, será incluído no cálculo de $\hat{\gamma}(h)$(90º,100m).

Isto feito, os cálculos são repetidos para a próxima distância, por exemplo, 200m.

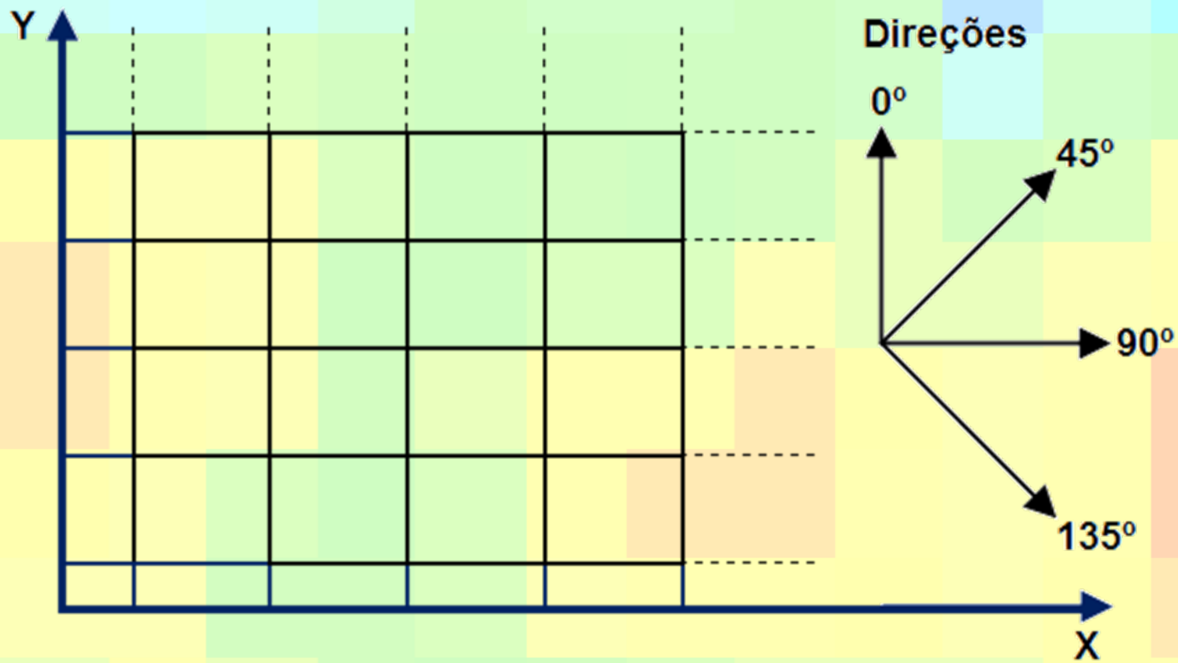
Isso inclui todos os pares de pontos, cuja distância é igual a 200m.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Cálculo do semivariograma para amostras regularmente espaçadas

O processo é repetido até que algum ponto de parada desejado seja alcançado.



Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Cálculo do semivariograma para amostras regularmente espaçadas

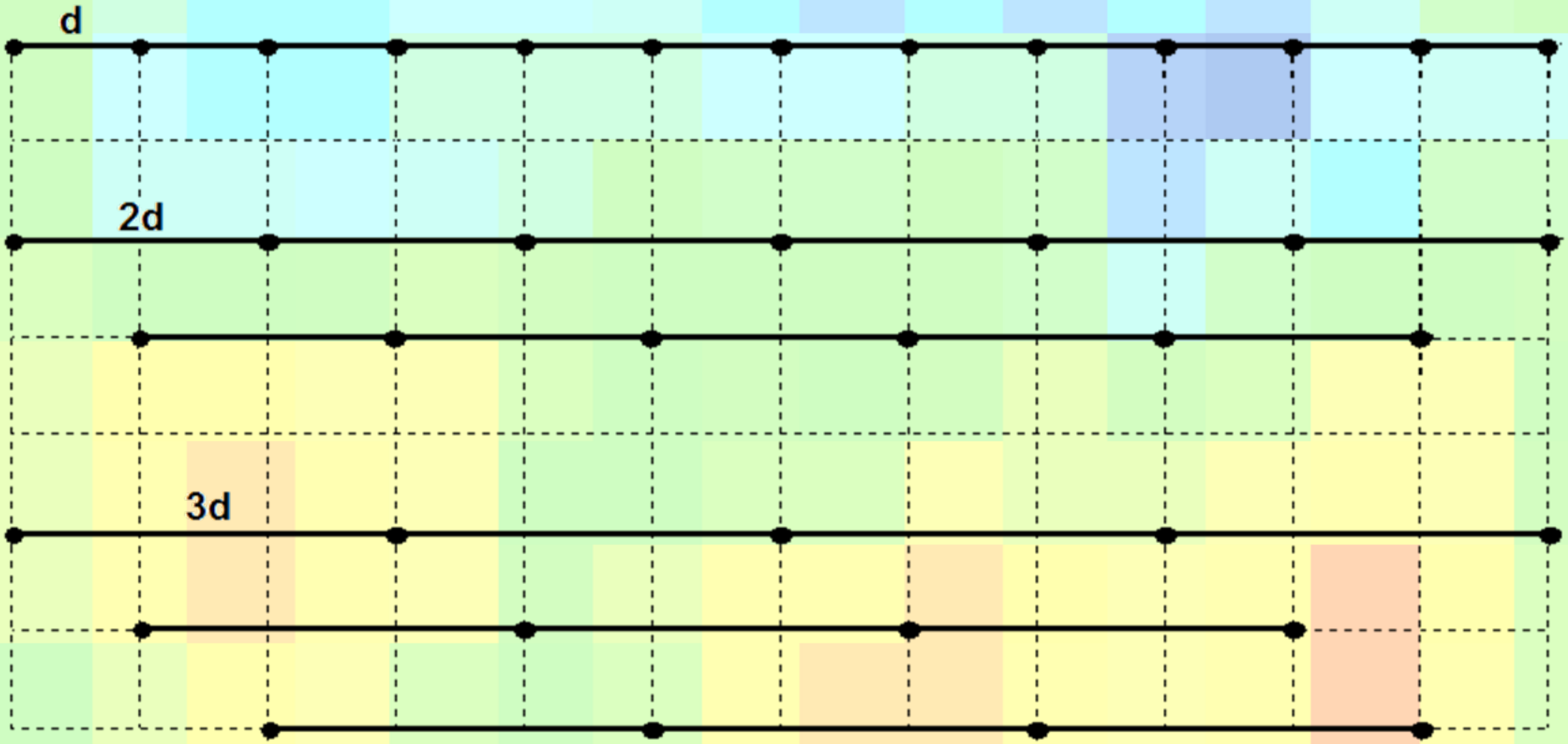


Ilustração para o cálculo do semivariograma a partir de uma malha de amostragem regularmente espaçada.

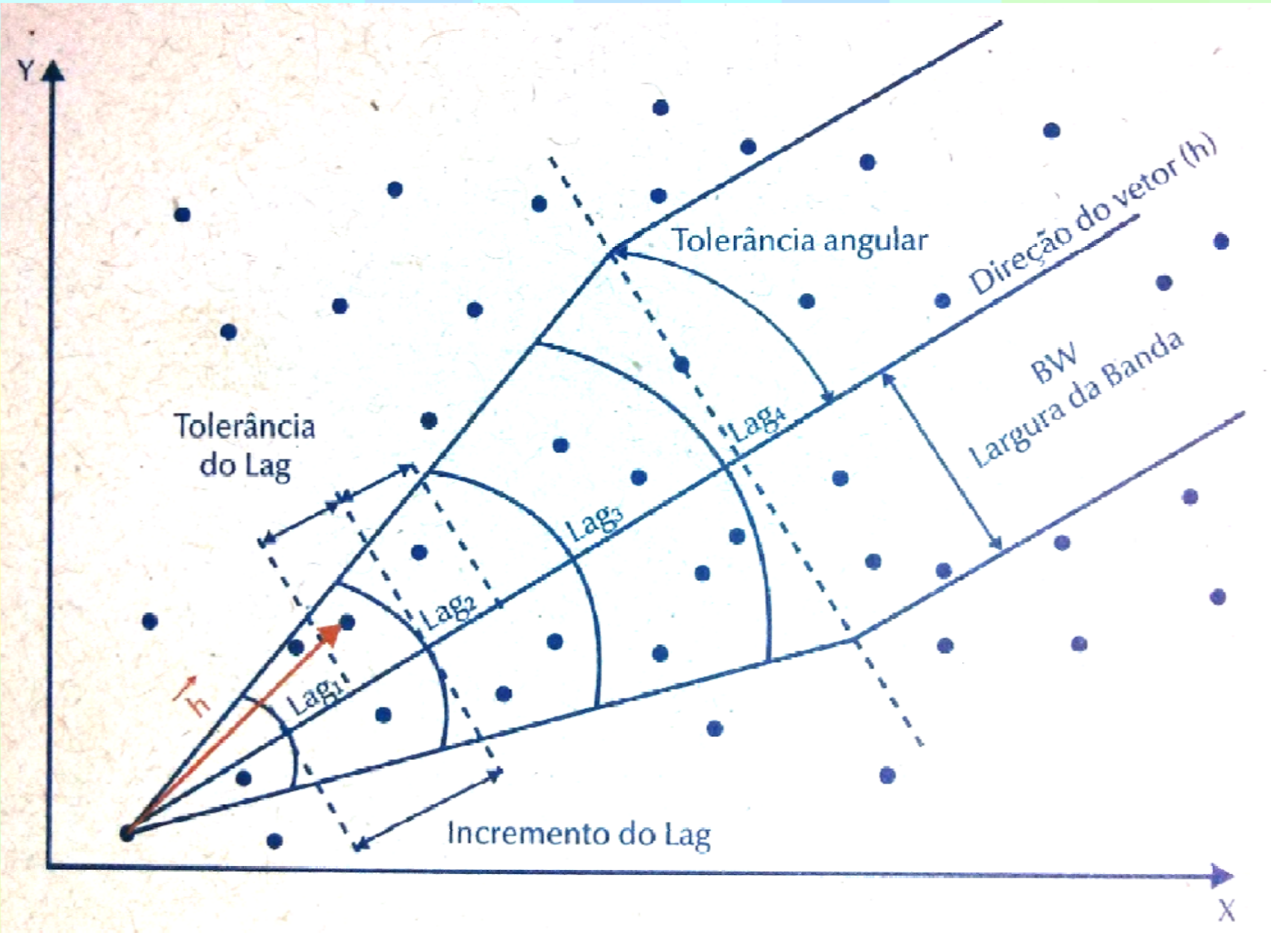
Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Cálculo do semivariograma a partir de uma malha de amostragem irregularmente espaçada.

Neste caso é necessário introduzir limites de tolerância para direção e distância.

Tome como referência o Lag2 da figura.



Geoestatística:

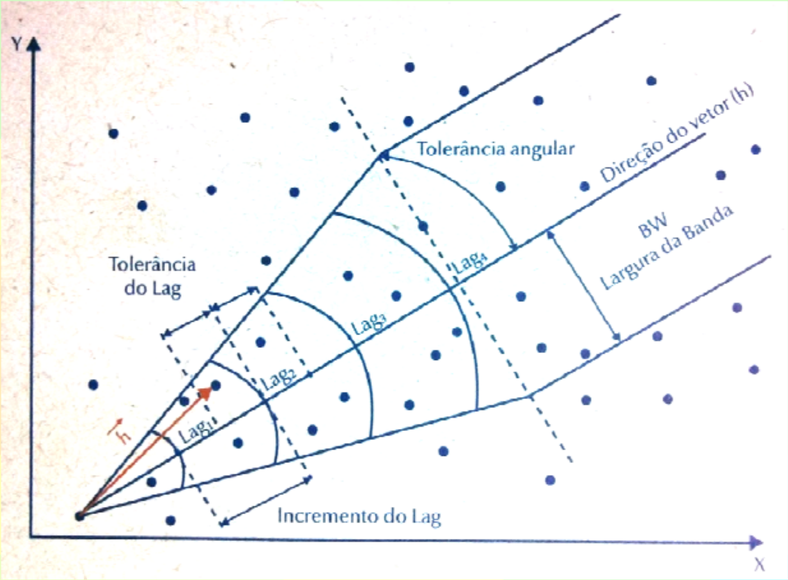
VARIOGRAMA (SEMIVARIOGRAMA)

Cálculo do semivariograma a partir de uma malha de amostragem irregularmente espaçada

Lag refere-se a uma distância pré-definida, a qual é utilizada no cálculo do semivariograma.

Suponha um incremento de Lag igual a 100m com tolerância de 50m.

Considere ainda a direção de medida de 45º com tolerância angular de 22,5º.



Então, qualquer par de observações cuja distância está compreendida entre 150m e 250m e 22,5º e 67,5º será incluído no cálculo do semivariograma de Lag2.

Este processo se repete par todos os Lags.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Cálculo do semivariograma a partir de uma malha de amostragem irregularmente espaçada

A largura da banda (BW) se refere a um valor de ajuste a partir do qual se restringe o número de pares de observações para o cálculo do semivariograma.

A próxima etapa constitui o ajuste de um modelo teórico ao semivariograma experimental.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

O ajuste de um modelo teórico ao semivariograma experimental não é direto e automático, pois, nesse processo, o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico.

Dependendo do ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Os modelos podem ser de dois tipos: **modelos com patamar** e **modelos sem patamar**.

Os **modelos com patamar** são conhecidos como **modelos transitivos**.

Alguns modelos transitivos atingem o patamar (C) assintoticamente, onde o alcance (a) é arbitrariamente definido como a distância correspondente a 95% do patamar.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Os **modelos sem patamar** continuam aumentando enquanto a distância aumenta.

Tais modelos são utilizados para modelar fenômenos que possuam capacidade infinita de dispersão (**modelo potência**).

Os **modelos transitivos** mais utilizados são:

- *Modelo esférico (Sph);*

- *Modelo Exponencial (Exp);* e

- *Modelo Gaussiano.*

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Modelo esférico (Sph)

O modelo esférico e um dos modelos mais utilizados e a equação normalizada é:

$$\begin{cases} \gamma(h) = C_0 + C\left[1{,}5\dfrac{h}{a} - 0{,}5\left(\dfrac{h}{a}\right)^3\right] & para\, h < a \\ \gamma(h) = C_0 + C & para\, h \geq a \end{cases}$$

Para dados ambientais é aplicado em 95% dos casos.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Modelo exponencial (Exp)

A equação normalizada é:

$$\gamma(h) = C_0 + C\left[1 - exp\left(-\frac{h}{a}\right)\right]$$

Esse modelo atinge o patamar assintoticamente, com alcance prático definido como a distância na qual o valor do modelo é 95% do patamar (Isaaks; Srivastava, 1989).

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Modelo gaussiano (Gau)

É um modelo transitivo, muitas vezes, usada para modelar fenômenos extremamente contínuos. Sua formulação é dada por:

$$\gamma(h) = C_0 + C\left[1 - exp\left(-\left(\frac{h}{a}\right)^2\right)\right]$$

Esse modelo também atinge o patamar assintoticamente, com alcance prático definido como a distância na qual o valor do modelo é 95% do patamar (Isaaks; Srivastava, 1989).

O que caracteriza esse modelo é seu comportamento parabólico próximo à origem.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Linear

Sua formulação é dada por:

$$\gamma(h) = \begin{cases} C_0 + \dfrac{C}{a}h & para\, h \leq a \\ C_0 + C & para\, h > a \end{cases}$$

Neste modelo o patamar é determinado por inspeção.

O coeficiente angular C/a é determinado pela inclinação da reta que passa pelos primeiros pontos de $\gamma(\mathbf{h})$ dando-se maior peso àqueles que correspondem a maior número de pares.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Modelo cúbico

$$\begin{cases} \gamma(h) = C_0 + C\left[7\left(\frac{h}{a}\right)^2 - \frac{35}{4}\left(\frac{h}{a}\right)^3 + \frac{7}{2}\left(\frac{h}{a}\right)^5 - \frac{3}{4}\left(\frac{h}{a}\right)^7\right] & para\, h < a \\ \gamma(h) = C_0 + C & para\, h \geq a \end{cases}$$

Modelo Pentaesférico

$$\begin{cases} \gamma(h) = C_0 + C\left[\frac{15}{8}\left(\frac{h}{a}\right) - \frac{5}{4}\left(\frac{h}{a}\right)^3 + \frac{3}{8}\left(\frac{h}{a}\right)^5\right] & para\, h < a \\ \gamma(h) = C_0 + C & para\, h \geq a \end{cases}$$

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Modelo efeito furo (*hole effect*)

$$\gamma(h) = C_0 + C\left[1 - \frac{sen\,\pi\,\frac{h}{a}}{\pi\,\frac{h}{a}}\right]$$

Modelo potência

O modelo potência não é um modelo transitivo, portanto não atinge o patamar.

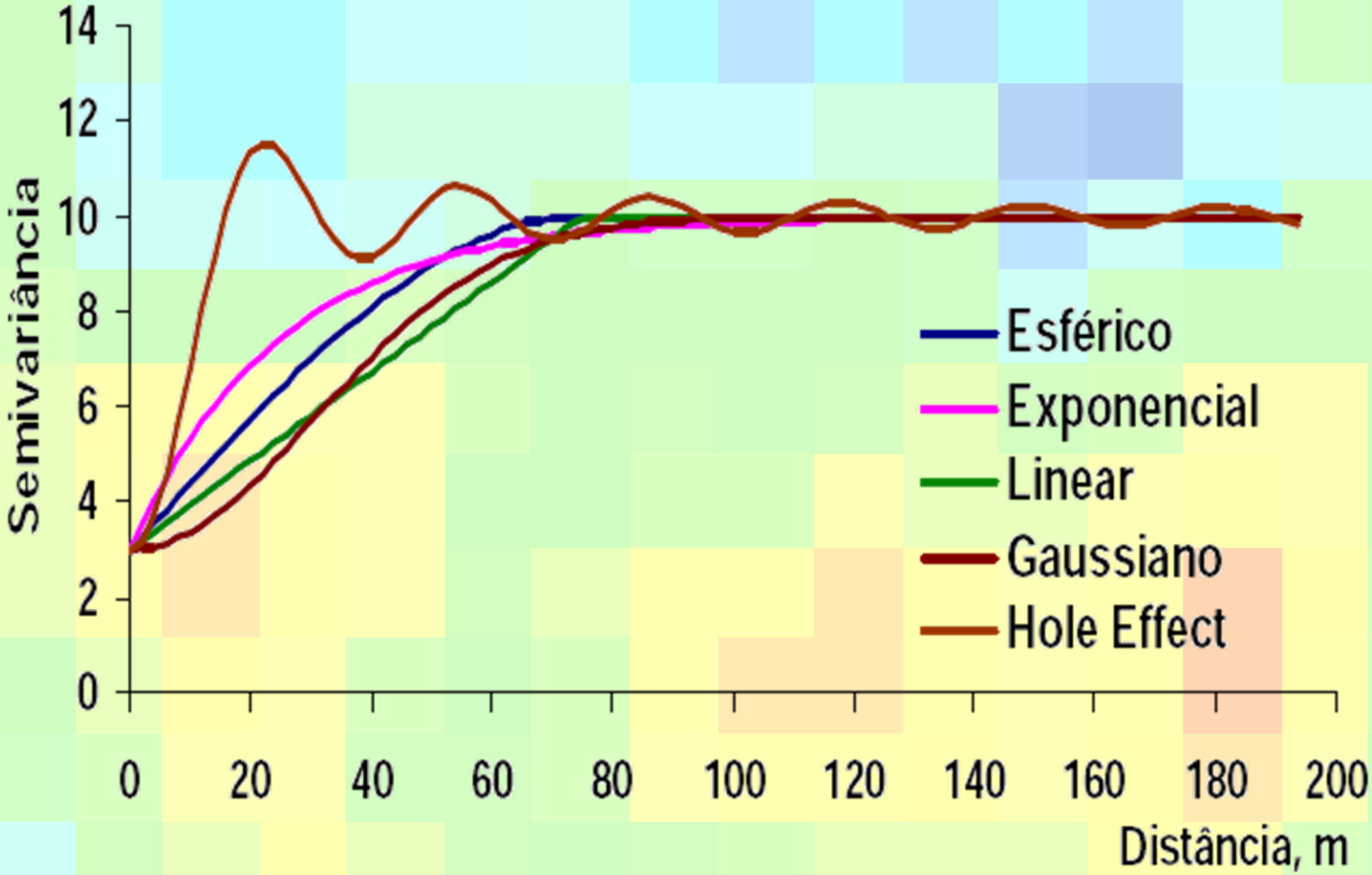
$$\gamma(h) = C_0 + \alpha[h]^{\beta}$$

α representa umaconstante positiva que multiplica a distância elevada a uma potência β.

Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)



Geoestatística:

VARIOGRAMA (SEMIVARIOGRAMA)

Modelos de Semivariogramas (Variogramas)

Modelos aninhados

Os modelos aninhados são combinações de modelos simples utilizados para modelar fenômenos mais complexos.

Por exemplo, um modelo aninhado útil em estudos de mineração e pesquisa de solo é o duplo esférico, definido como:

$$\begin{cases} \gamma(h) = C_0 + C_1\left[1{,}5\dfrac{h}{a_1} - 0{,}5\left(\dfrac{h}{a_1}\right)^3\right] & para\, 0 < h \leq a_1 \\ \gamma(h) = C_0 + C_2\left[1{,}5\dfrac{h}{a_2} - 0{,}5\left(\dfrac{h}{a_2}\right)^3\right] & para\, a_1 < h \leq a_2 \\ \gamma(h) = C_0 + C_1 & para\, h \geq a_2 \end{cases}$$

Geoestatística:

Grau de Dependência Espacial

Os variogramas expressam o comportamento espacial da variável regionalizada ou de seus resíduos e mostram o tamanho da zona de influência em torno de uma amostra, a variação nas diferentes direções do terreno, indicando também continuidade da característica estudada no terreno (Landim, 1998).

Trangmar et al (1985) sugeriram o uso da % da variância do efeito pepita para mensurar a dependência espacial (IDE), sendo que Cambardella et al (1994) propuseram os seguintes intervalos para avaliar a % da variância do efeito pepita:

IDE ≤ 25% = forte dependência espacial;

25% ≤ IDE ≤ 75% = moderada dependência espacial;

≥ 75% = fraca dependência espacial.

$$IDE = \frac{C_0}{(C_0 + C_1)} \times 100$$

Geoestatística:

Grau de Dependência Espacial

Zimback (2001) propôs a inversão dos fatores, como:

$$IDE = \frac{C_1}{(C_0 + C_1)} \times 100$$

e a classificação quanto ao grau de dependência espacial da variável em estudo é:

i) variável independente espacialmente – se a relação entre a componente estrutural e o patamar for igual a 0%, neste caso o variograma será com efeito pepita puro;

ii) variável com fraca dependência espacial – se a componente estrutural for menor ou igual a 25% do patamar;

iii) Variável com moderada dependência espacial – se a componente estrutural estiver entre 25% e 75% do patamar;

iv) Variável com forte dependência espacial – se a relação entre a componente estrutural e o patamar for maior ou igual 75%.