

População, Espaço e Ambiente

Abordagens Espaciais em Estudos de População: Métodos Analíticos e Técnicas de Representação

Parte IV – Integração de Dados

De pontos a superfícies 1: geoestatística

Silvana Amaral
Antonio Miguel V. Monteiro

{silvana@dpi.inpe.br, miguel@dpi.inpe.br}

Interpoladores *Geográficos*

Modelos de População em Superfícies

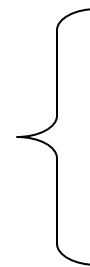
Baseados em Pontos – *Point-based*

Baseados em Áreas – *Areal-Based*

Informação Auxiliar

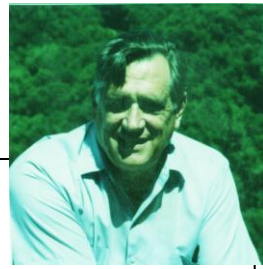
Conhecimento dos Processos

População



Preservar Volume
Restrições

Interpolação Geográfica

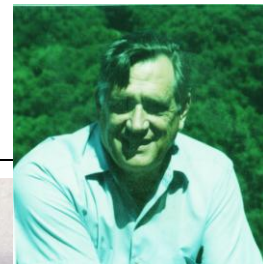


- Interpolação requer pressupostos sobre a natureza do fenômeno a ser interpolado...
...seguido por um “intelligent” guessing
- Por exemplo:

**Amostragem aleatória seguida por interpolação
seriam eficientes para representar esta superfície ??**



Outro desafio para interpolação



Amostras de elevação em localidades aleatórias em uma cidade e depois interpolação.

O que teríamos?



Interpolação de áreas é um caso especial de Interpolação Geográfica



Observações numéricas dadas em localizações pontuais

- Objetivo é produzir um mapa de contornos (isolinhas)
- Há vasta literatura para interpolação a partir de dados pontuais
- Clássicos: Kriging, inverso da distância, splining, etc.

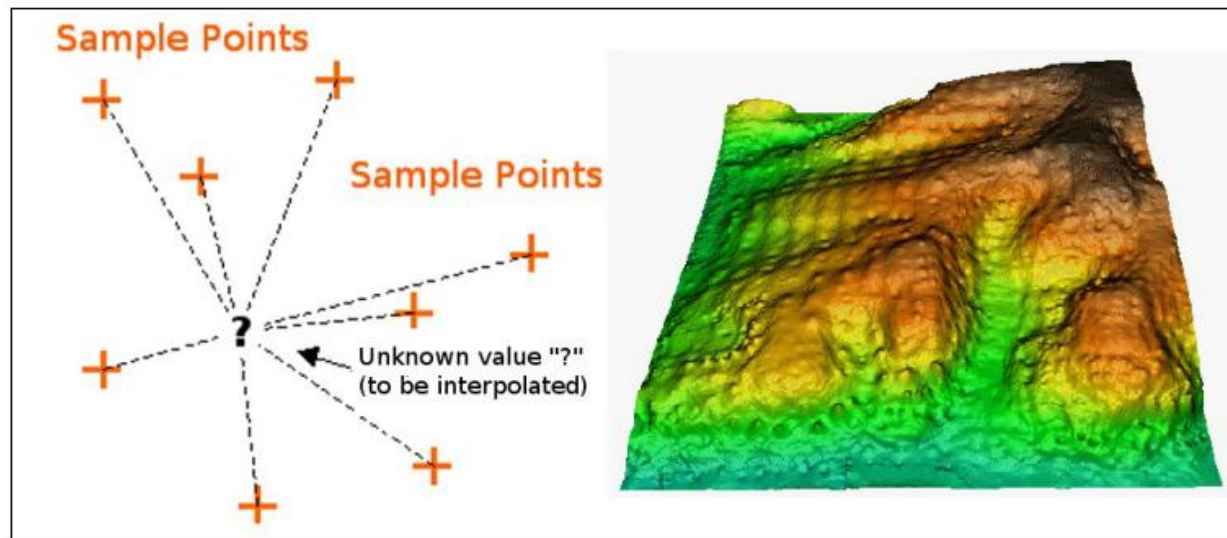


Illustration 2: Inverse Distance Weighted interpolation based on weighted sample point distance (left). Interpolated IDW surface from elevation vector points (right). Image Source: Mitas, L., Mitasova, H. (1999)).

Referência



Geoestatística para geoprocessamento

Eduardo C. G. Camargo

Geoestatística para geoprocessamento

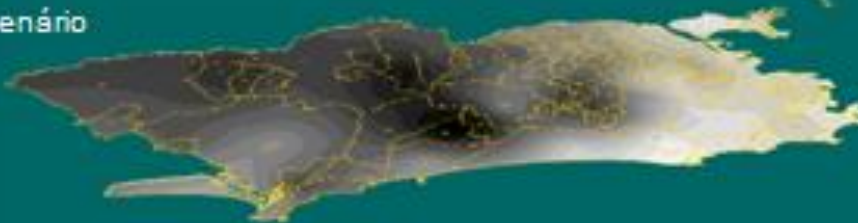
realidade



análise estrutural



cenário



Palavras-chave

- variáveis regionalizadas
- semivariograma empírico
- efeito pepita, alcance e patamar
- análise estrutural
- isotropia e anisotropia
- validação cruzada
- krigagem

O que é geoestatística?



É uma abordagem **PROBABILÍSTICA** de modelagem, que engloba um conjunto de métodos estatísticos, para a análise e mapeamento de dados distribuídos no espaço e/ou no tempo.

Requer o conhecimento de alguns conceitos básicos:

- Variável aleatória (V.A.)
- Momentos da V.A. Exs: $E[X]$, $C[X,Y]$;
- Função densidade de probabilidade (FDP);
- Função de Distribuição Acumulada (FDA): univariada e bivariada;
- Função aleatória (FA), etc.

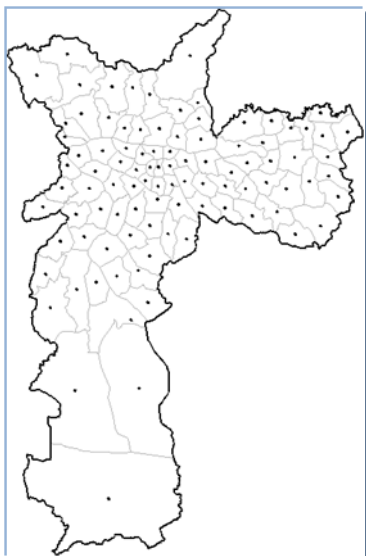
- 1) **Análise:** objetiva descrever a variabilidade espacial do fenômeno em estudo, denominada de análise estrutural ou modelagem do semivariograma.
- 2) **Inferência:** objetiva estimar valores de uma variável distribuída no espaço em locais não amostrados, denominada de krigagem.
- 3) **Simulação:** objetiva construir um conjunto de realizações equiprováveis ou igualmente representativa do fenômeno em estudo.

Etapas da modelagem geoestatística



Etapas da modelagem geoestatística

Região de estudo

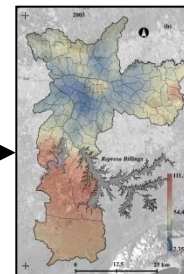


análise
exploratória

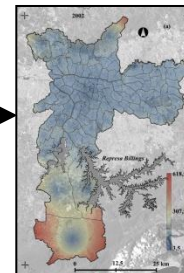
análise
estrutural

interpolação
krigeagem

simulação
condicionada

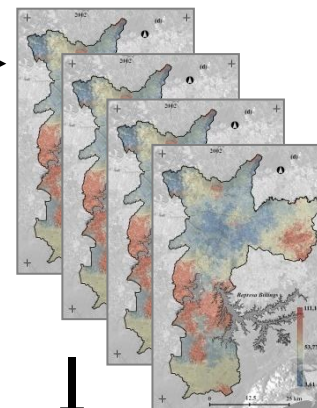


Superfície estimada
do fenômeno
investigado



Superfície da variância
da estimativa

Realizações equiprováveis



- Construção de cenários
- Mapas de incerteza

Porque usar geoestatística?

Interpoladores

Procedimentos determinísticos

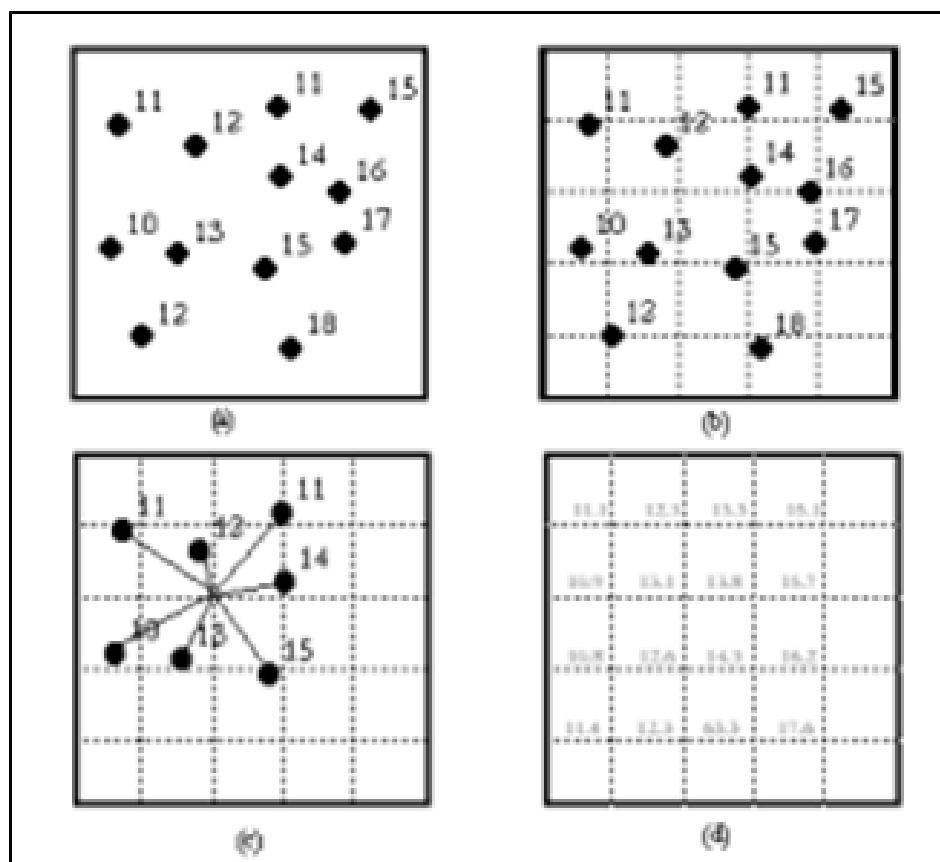


Figura 3-3 Ilustração do processo de interpolação por estimador local: (a) configuração original de amostras; (b) grade regular superposta às amostras; (c) interpolação de um valor a partir dos vizinhos; (d) grade regular resultante

Porque usar geoestatística?

Interpoladores

Procedimentos determinísticos

Média Móvel

Média Simples

Inverso do quadrado da distância

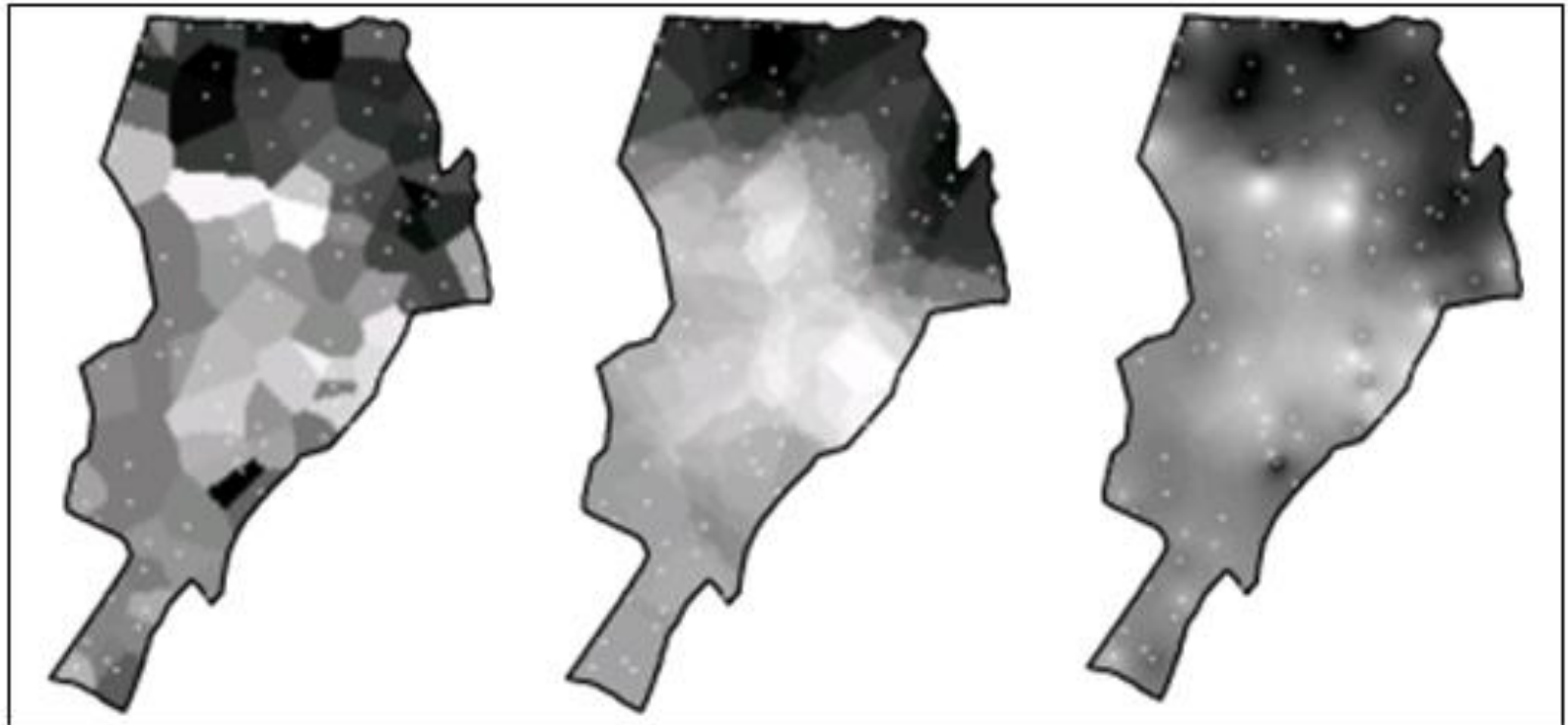
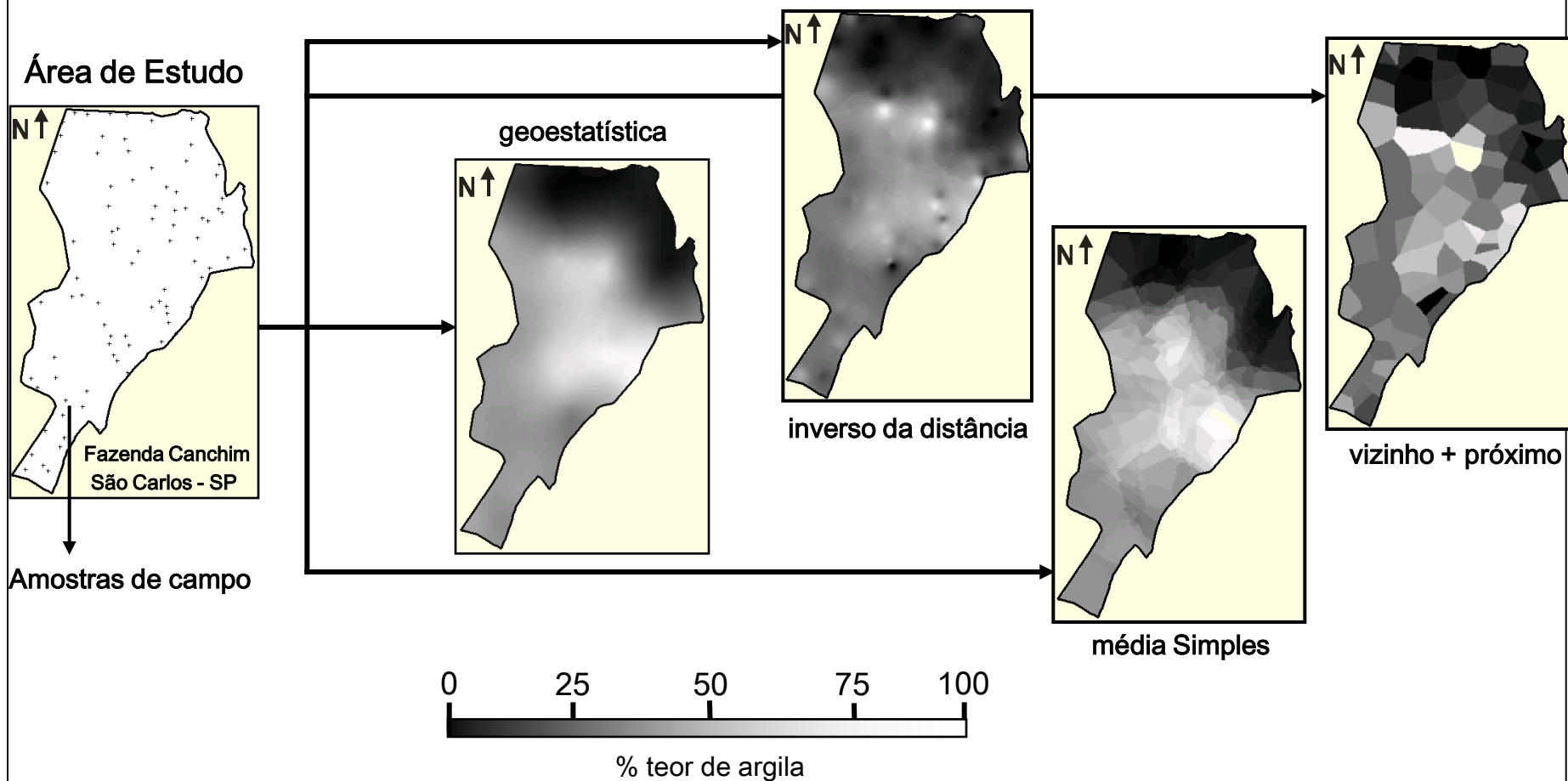


Figura 3-4 - Comparação entre interpoladores de média móvel, para o mesmo conjunto de amostras. À direita, inverso do quadrado da distância; no centro, média simples; à esquerda, vizinho mais próximo. Regiões mais claras representam alto valores e vice-versa.

Porque usar geoestatística?

Interpoladores

Procedimentos determinísticos



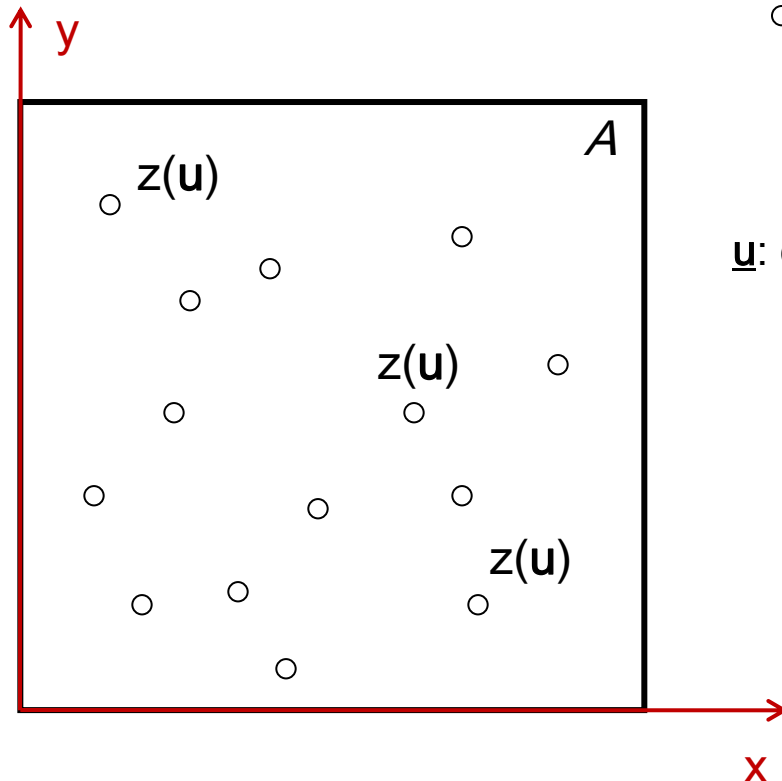
Principais conceitos teóricos

Variável aleatória (V.A.): Uma visão prática no contexto da geoestatística.

Z representa a variável em estudo.

- localizações geográficas onde a variável Z é medida ou observada, denotado por $z(u)$, $u \in A$.

$\underline{u}$: é um **vetor** de coordenadas geográficas $[u(x,y)]$;

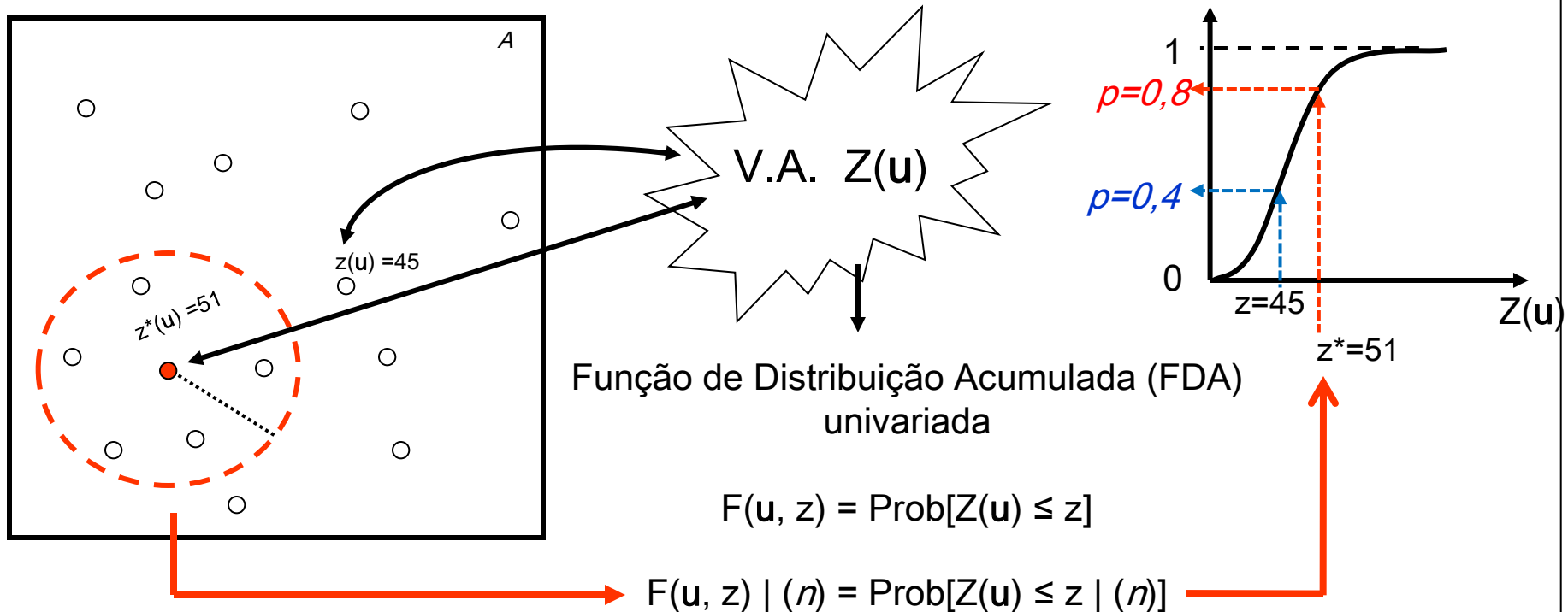


Principais conceitos teóricos



Variável aleatória (V.A.): Uma visão prática no contexto da geoestatística.

Localmente, um valor $z(u)$, $u \in A$, é interpretado como uma das possíveis realizações da variável aleatória $Z(u)$.

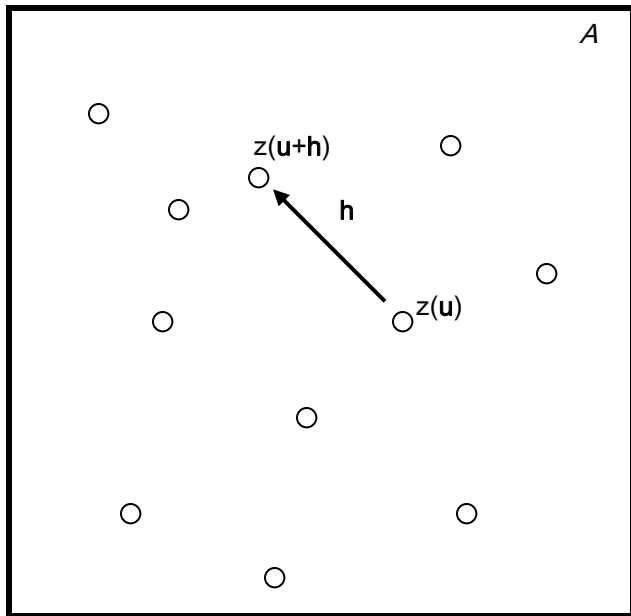


Principais conceitos teóricos



Função aleatória (F.A.): Uma visão prática no contexto da geoestatística.

O conjunto de V.A., $\{Z(u), u \in A\}$, é uma F.A. $Z(u)$.



Na geoestatística um caso de particular interesse de F.A. é a Função de Distribuição Aleatória (FDA) bivariada.

$$F(u, u+h; z_1, z_2) = \text{Prob}[Z(u) \leq z_1, Z(u+h) \leq z_2]$$



Momento da FDA bivariada: Covariância

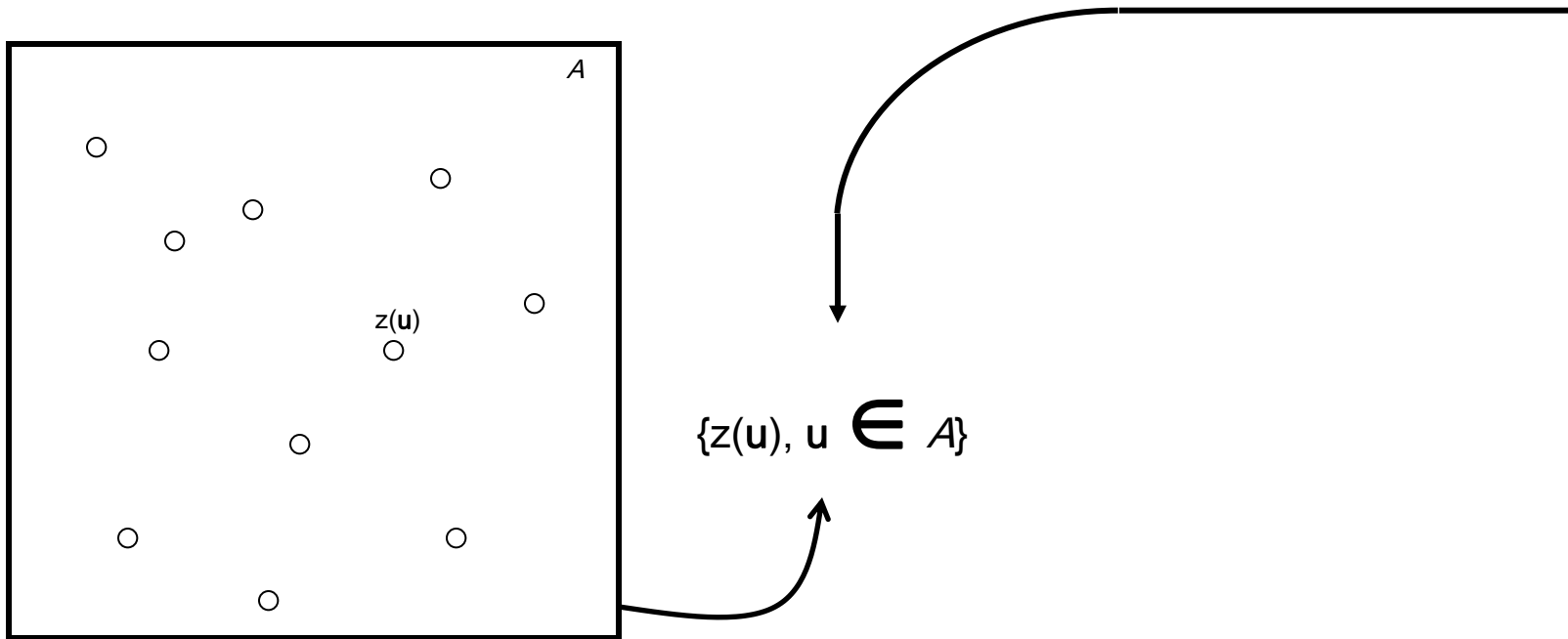
$$C[Z(u), Z(u+h)] = E[Z(u) \cdot Z(u+h)] - E[Z(u)] \cdot E[Z(u+h)]$$

h é um vetor distância entre dois pontos.

O que conhecemos de fato até agora?

Resposta: uma única amostragem do fenômeno de interesse.

Em outras palavras, tudo o que se sabe da F.A. $Z(u)$ é uma única realização.



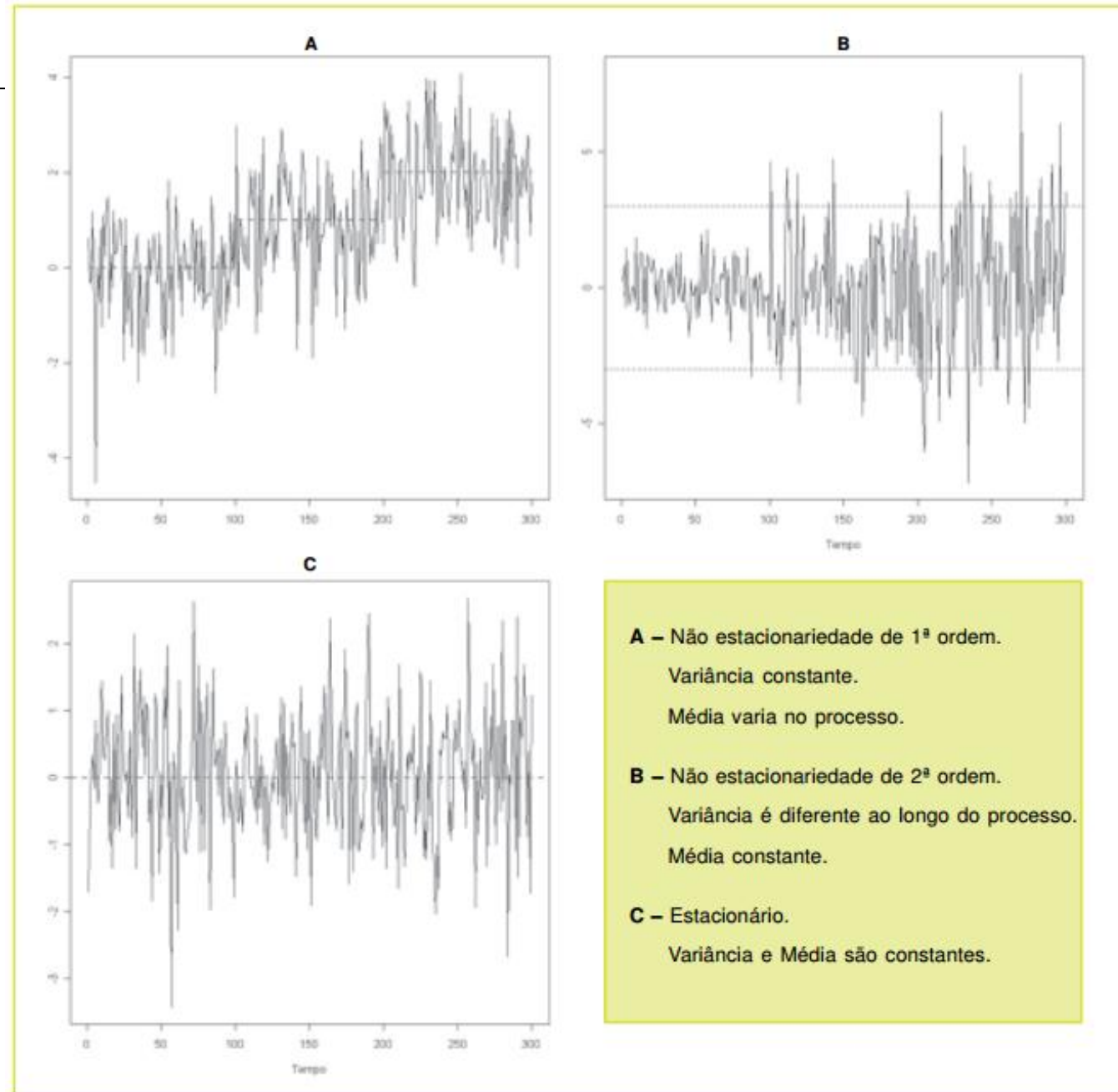
PROBLEMA: como deduzir a lei de probabilidade da F.A. $Z(u)$ a partir de uma única realização da mesma?

O paradigma que se estabelece, para inferir as FDA e interpolar valores em localizações não amostradas, é o de assumir a hipótese de ***estacionariedade***.

- a **estacionariedade** é uma propriedade do modelo probabilístico, uma hipótese necessária para realização de inferências;
- não é uma característica do fenômeno espacial em estudo;
- é uma decisão feita pelo analista, afim de verificar a adequação do modelo à realidade a ser investigada.

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. ”

FIGURA 1.6 – Exemplos de análise da estacionariedade numa Série Temporal.



Estacionariedade

Principais conceitos teóricos



Admite-se também que o fenômeno considerado seja *estacionário de segunda ordem*, isto é, a covariância entre dois pares quaisquer $Z(\mathbf{x})$ e $Z(\mathbf{x} + \mathbf{h})$, separados por um vetor distância $\mathbf{h}$, existe e depende somente de $\mathbf{h}$. Então:

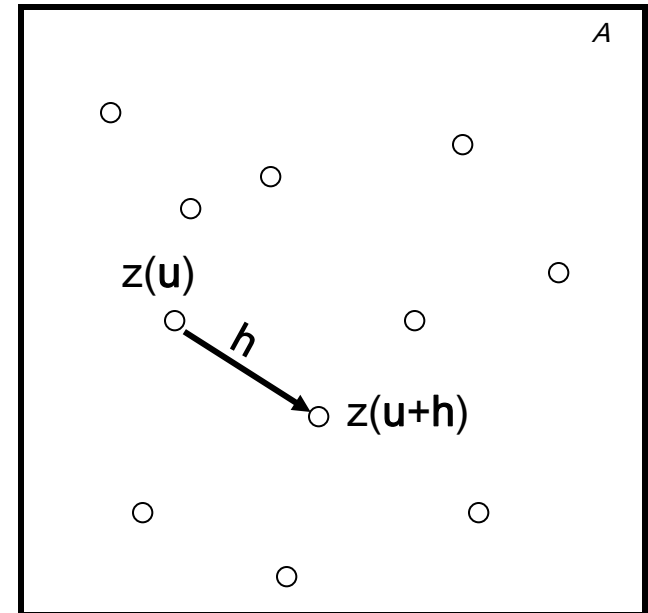
$$C(\mathbf{h}) = \text{COV} [Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})] = E[Z(\mathbf{x}).Z(\mathbf{x} + \mathbf{h})] - m^2$$

Hipótese de estacionariedade de 2ª ordem

Considera somente o primeiro e o segundo momentos invariantes da F.A.

$$1) E[Z(\mathbf{u})] = m, \quad \forall \mathbf{u} \in A$$

$$E[Z(\mathbf{u})] = E[Z(\mathbf{u} + \mathbf{h})] = m \quad \text{ou} \quad E[Z(\mathbf{u})] - E[Z(\mathbf{u} + \mathbf{h})] = E[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})] = 0$$

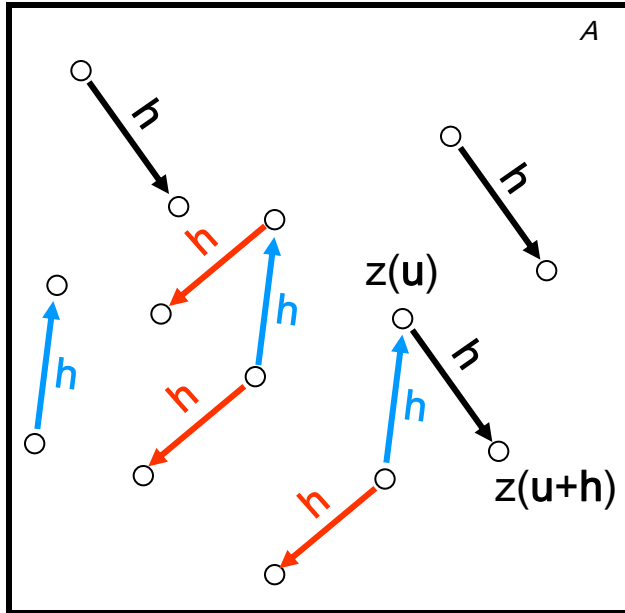


Principais conceitos teóricos



Hipótese de estacionariedade de 2ª ordem

2) a covariância entre os pares $Z(u)$ e $Z(u + h)$, separados por um vetor distância h , é estacionária.



Estacionariedade da covariância significa que a covariância entre dois pares quaisquer,

$Z(x)$ e $Z(x+h)$, é invariante para um mesmo vetor distância h .

ou

que a covariância entre dois pares quaisquer $Z(x)$ e $Z(x + h)$, separados por um vetor distância h , existe e depende somente de h

$$C[Z(u), Z(u + h)] = E[(Z(u) \cdot Z(u + h)) - E[Z(u)] \cdot E[Z(u + h)]] \quad \forall u \in A$$

Principais conceitos teóricos



Admite-se também que o fenômeno considerado seja *estacionário de segunda ordem*, isto é, a covariância entre dois pares quaisquer $Z(x)$ e $Z(x+h)$, separados por um vetor distância h , existe e depende somente de h . Então:

A estacionariedade da covariância implica na estacionariedade da variância:

$$\text{Var}(Z(x)) = E[Z^2(x)] - 2m.m + m^2 = E[Z^2(x)] - m^2 = C(0)$$

Hipótese de estacionariedade de 2ª ordem

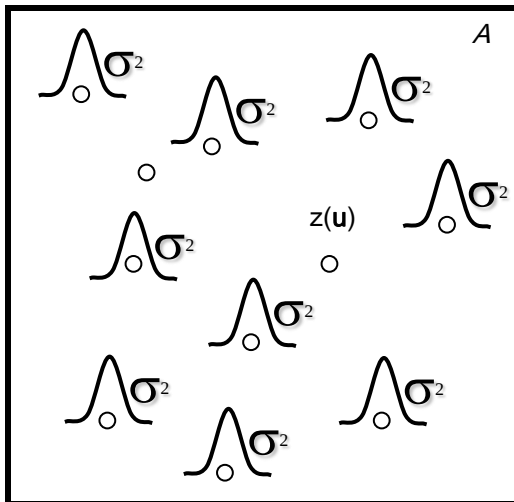
3) A estacionariedade da covariância implica na estacionariedade da variância:

$$\text{Var}[Z(u)] = E[Z(u) - m]^2 = E[Z^2(u)] - 2.E[Z(u)].m + m^2 =$$

$$= E[Z(u).Z(u+0)] - 2m^2 + m^2 =$$

$$= E[Z(u).Z(u+0)] - m^2 = C(0), \quad \forall u \in A$$

Covariância



Principais conceitos teóricos



Hipótese de estacionariedade intrínseca

estabelece que os incrementos $[Z(u) - Z(u + h)]$ tem esperança zero e variância somente em função de h , assim:

$$1) E[Z(u) - Z(u + h)] = 0, \quad \forall u \in A.$$

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

em que: $2\gamma(h)$ é denominado de função variograma
e $\gamma(h)$ de semivariograma

$$\gamma(h) = C(0) - C(h)$$

Principais conceitos teóricos



$$1) E[Z(u) - Z(u + h)] = 0, \quad \forall u \in A.$$

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

em que: $2\gamma(h)$ - função variograma
 $\gamma(h)$ - semivariograma

$$\gamma(h) = C(0) - C(h)$$

A covariância $C(h)$ e o semivariograma $\gamma(h)$ são ferramentas equivalentes para caracterizar a dependência espacial.

$\gamma(h)$ representa o *semivariograma*, que é metade do variograma

Sob a hipótese de estacionariedade de 2ª ordem, a covariância e o semivariograma são formas alternativas de caracterizar a autocorrelação dos pares $Z(x)$ e $Z(x+h)$ separados pelo vetor h .

Principais conceitos teóricos



$$2\gamma(\mathbf{h}) = 2C(0) - 2C(\mathbf{h}) \quad \text{ou} \quad \gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$$

O **semivariograma** é metade do **variograma**.

A relação indica que sob a hipótese de estacionariedade de 2ª ordem, que a covariância e o semivariograma *são formas alternativas de caracterizar a autocorrelação dos pares $Z(x)$ e $Z(x+h)$ separados pelo vetor h .*

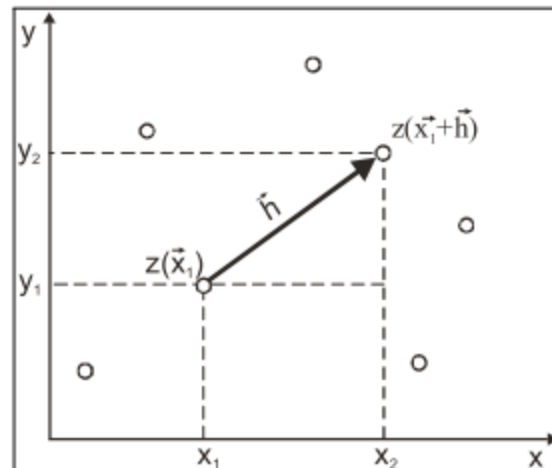


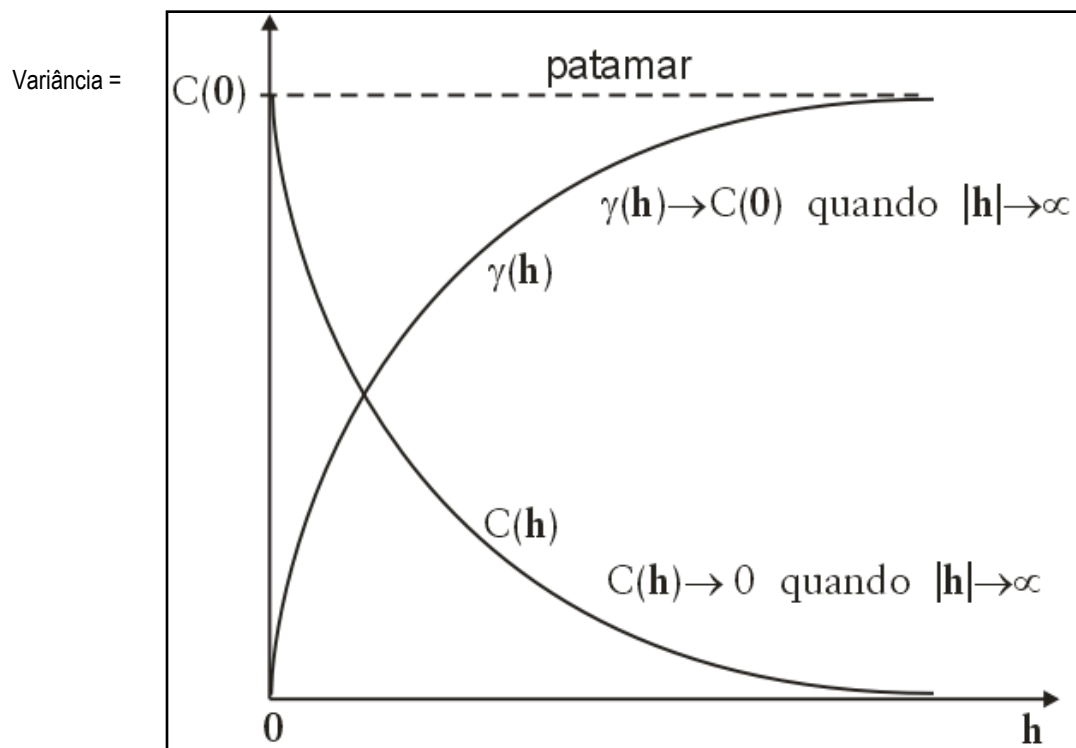
Figura 3-8 – Amostragem em duas dimensões.

Principais conceitos teóricos



relação entre as funções semivariograma e covariância

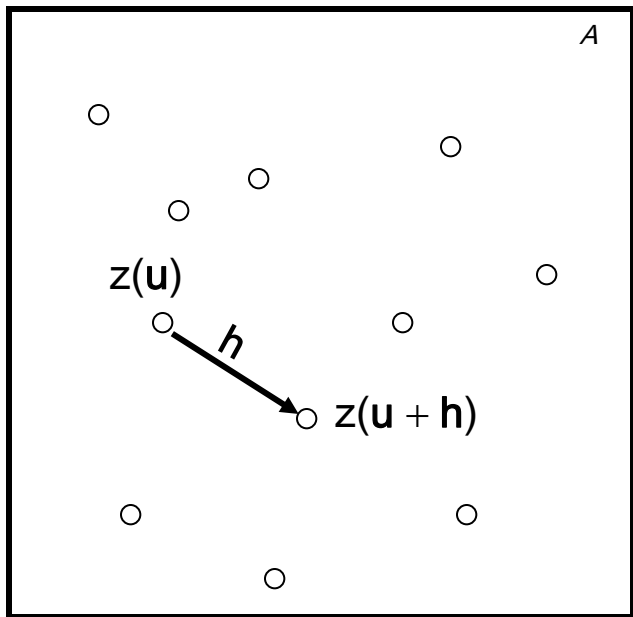
$$\gamma(h) = C(0) - C(h)$$



Variograma $2\gamma(h)$



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



- $2\gamma(h)$ mede o grau de dissimilaridade entre pares
- de observação separados pelo vetor distância h ;
- é função do vetor distância h ;
- depende da geometria de amostragem.

Variograma $2\gamma(\mathbf{h})$



Definição: esperança matemática (E) do quadrado da diferença entre os valores de pontos no espaço separados pelo vetor distância $\mathbf{h}$.

$$2\gamma(\mathbf{h}) = E\{[z(\mathbf{u}) - z(\mathbf{u} + \mathbf{h})]^2\}$$

Através de um conjunto amostral, $\{z(\mathbf{u}_1), z(\mathbf{u}_2), \dots, z(\mathbf{u}_N)\}$, o variograma pode ser estimado por:

$$\hat{2\gamma}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2$$

em que:

$\hat{2\gamma}(\mathbf{h})$: é o estimador de variograma;

$\mathbf{h}$: é o vetor distância (modulo e direção) entre pares de observação;

$N(\mathbf{h})$: é o número de pares, $z(\mathbf{u}_i)$ e $z(\mathbf{u}_i + \mathbf{h})$, separados por $\mathbf{h}$;

$z(\mathbf{u}_i)$ e $z(\mathbf{u}_i + \mathbf{h})$: são valores observados nas localizações $\mathbf{u}_i$ e $\mathbf{u}_i + \mathbf{h}$.

Semivariograma $\gamma(\mathbf{h})$



Definição: metade da esperança matemática (E) do quadrado da diferença entre os valores de pontos no espaço separados pelo vetor distância $\mathbf{h}$.

$$\gamma(\mathbf{h}) = \frac{1}{2} E\{[z(\mathbf{u}) - z(\mathbf{u} + \mathbf{h})]^2\}$$

Através de um conjunto amostral, $\{z(\mathbf{u}_1), z(\mathbf{u}_2), \dots, z(\mathbf{u}_N)\}$, o semivariograma pode ser estimado por:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2$$

em que:

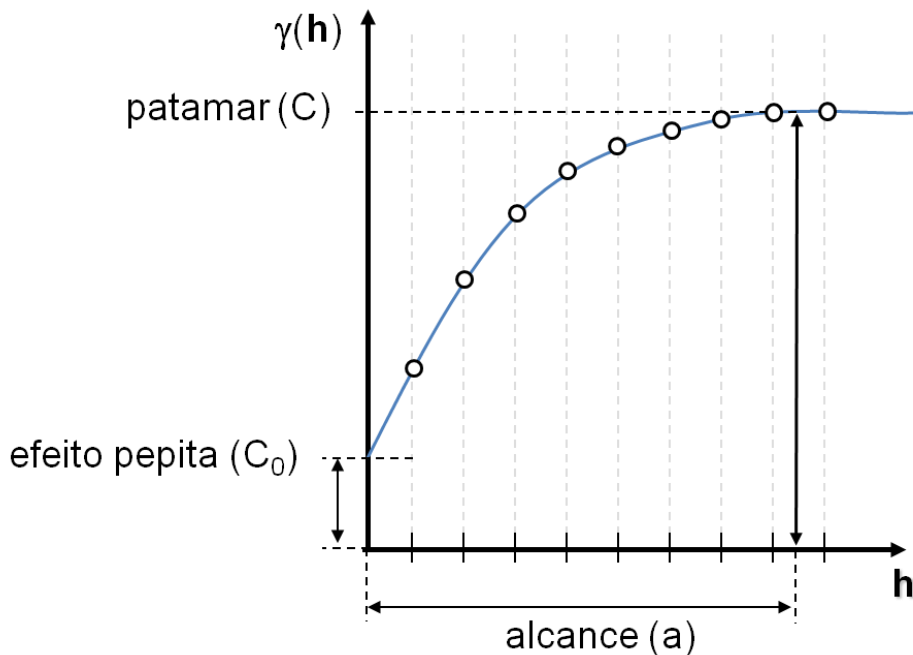
$\hat{\gamma}(\mathbf{h})$: é o estimador de semivariograma;

$\mathbf{h}$, $N(\mathbf{h})$, $z(\mathbf{u}_i)$ e $z(\mathbf{u}_i + \mathbf{h})$: conforme definidos anteriormente.

Semivariograma $\gamma(h)$



A figura ilustra um semivariograma empírico (ou experimental) com características muito próximas do ideal.



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

- Patamar (C): é o valor do semivariograma correspondente a seu alcance (a).

Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ($\text{Var}[Z(x) - Z(x+h)]$) torna-se aproximadamente constante.

- Efeito Pepita (C_0): idealmente, $\gamma(0)=0$.

Entretanto, na prática, à medida que h tende para zero, $\gamma(h)$ se aproxima de um valor positivo chamado Efeito Pepita (C_0), que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras.

O efeito pepita é o valor da semivariância para a distância zero e representa a componente da variabilidade espacial que não pode ser relacionado com uma causa específica (variabilidade ao acaso)

Semivariograma $\gamma(h)$

A figura ilustra um semivariograma empírico (ou experimental) com características muito próximas do ideal.

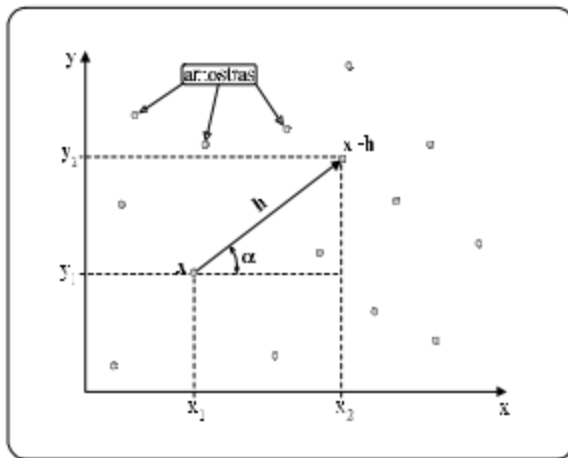


Fig. 2.2 - Amostragem em duas dimensões.

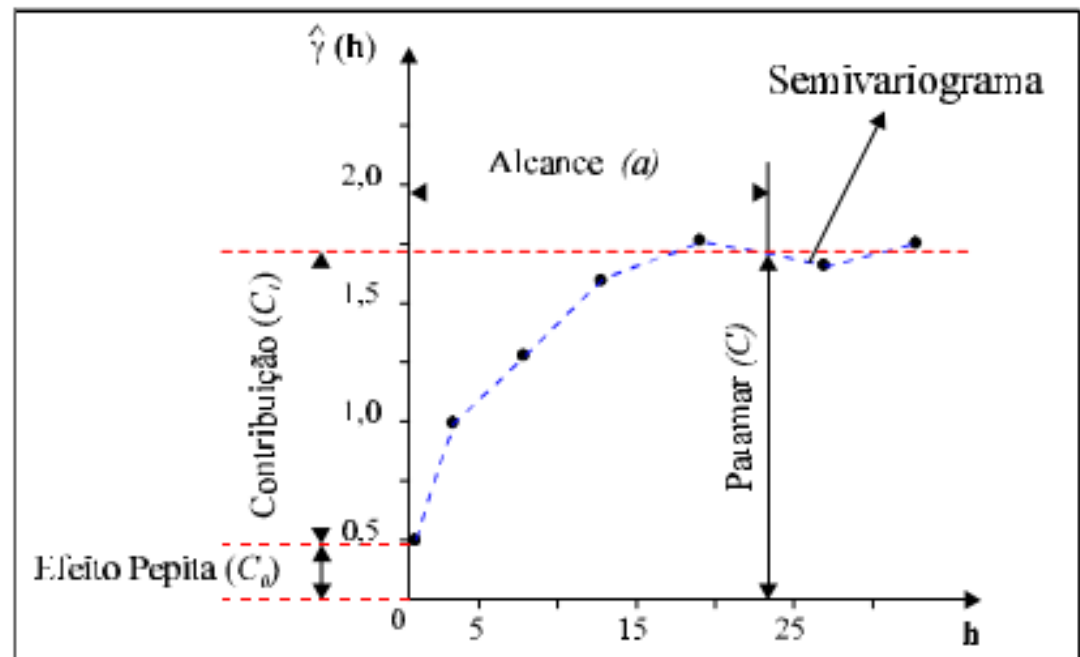
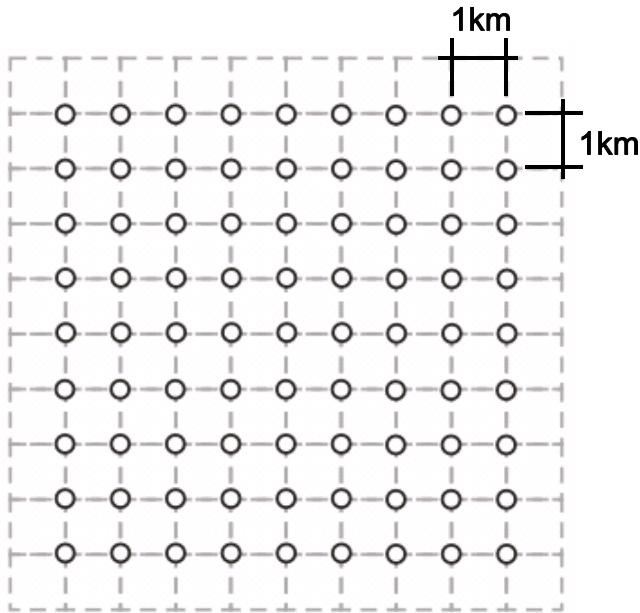


Fig. 2.3 - Exemplo de semivariograma.

Semivariograma $\gamma(h)$

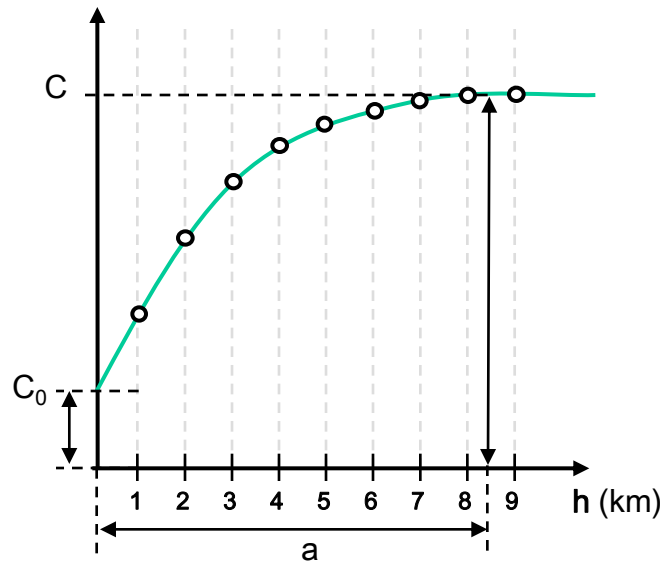
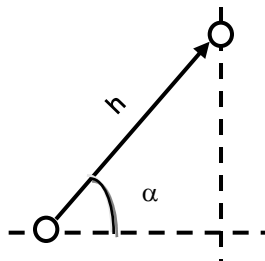


Cálculo do semivariograma a partir de amostras regularmente espaçadas.

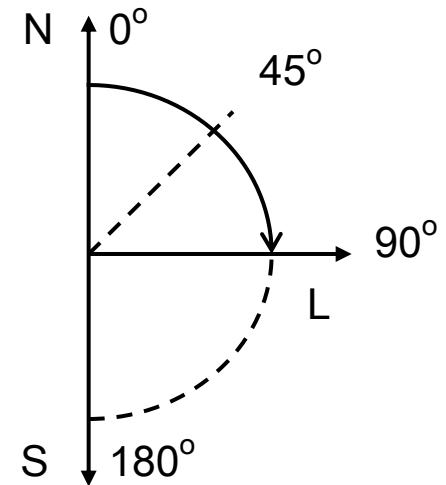


$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u_i) - z(u_i + h)]^2$$

vetor distância h



direções de análise



função simétrica

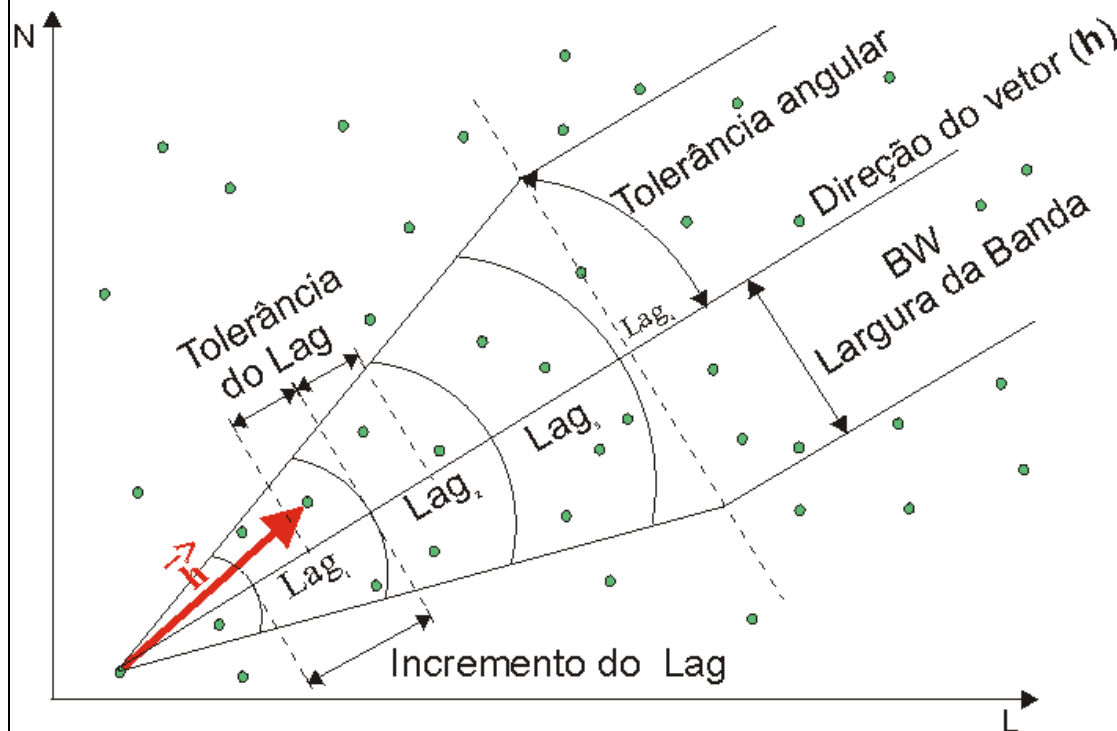
$$\gamma(h) = -\gamma(h)$$

Semivariograma $\gamma(h)$



Cálculo do semivariograma a partir de amostras irregularmente espaçadas.

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



parâmetros adicionais

- tolerância do incremento (lag)
- tolerância angular
- largura de banda

Semivariograma $\gamma(h)$

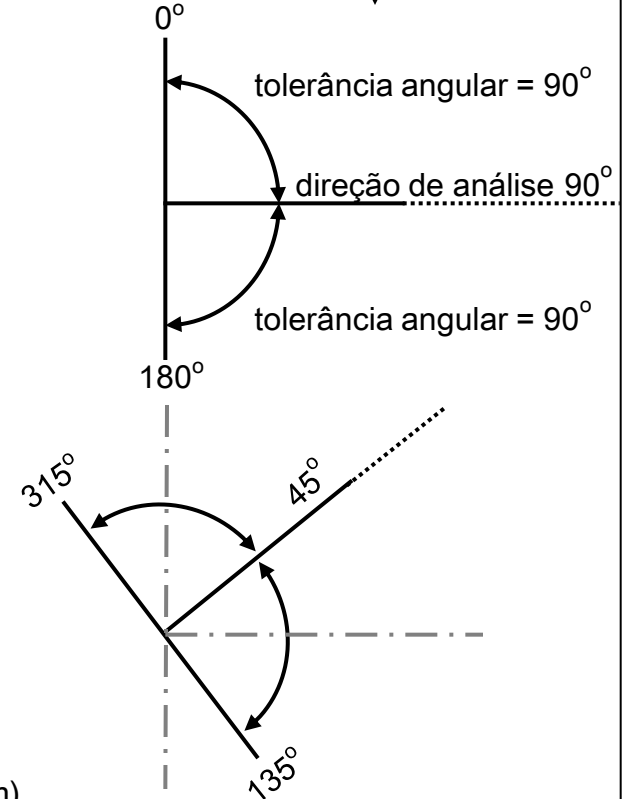
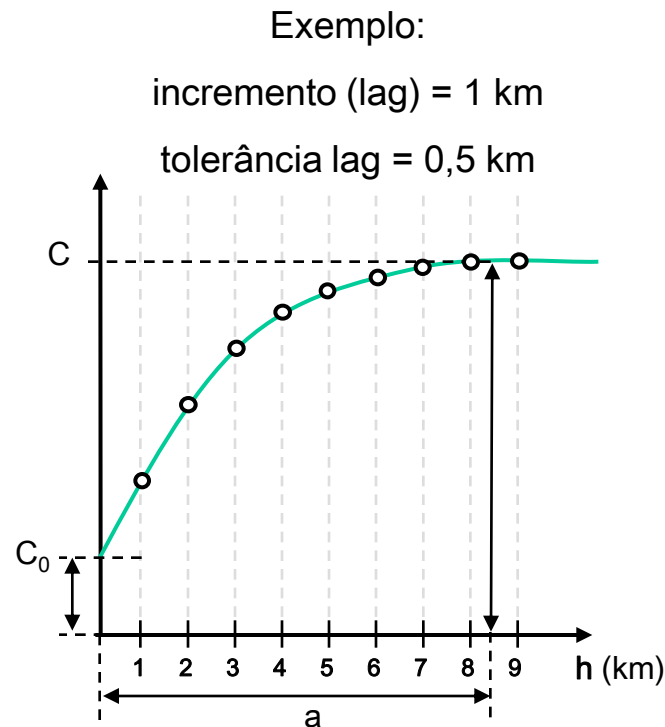
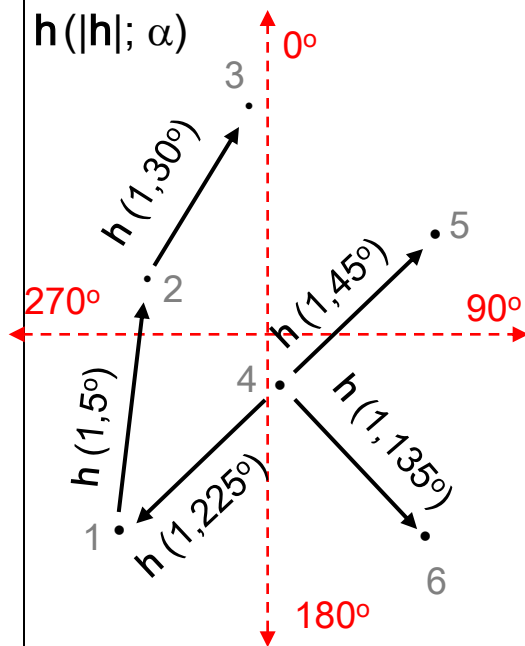


Cálculo do semivariograma a partir de amostras irregularmente espaçadas.

Semivariograma omnidirecional => tolerância angular = 90°

direção de análise (do vetor h) não importa.

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



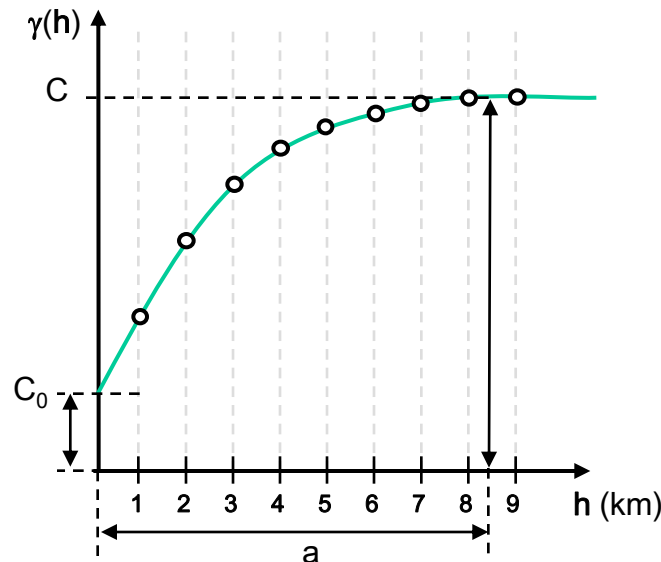
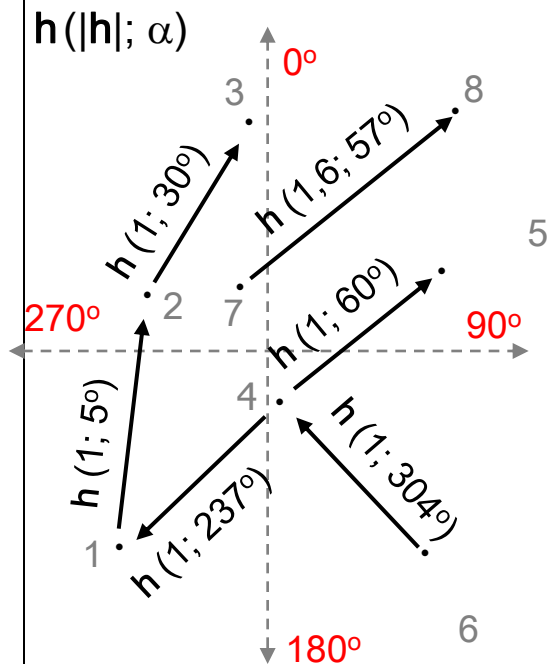
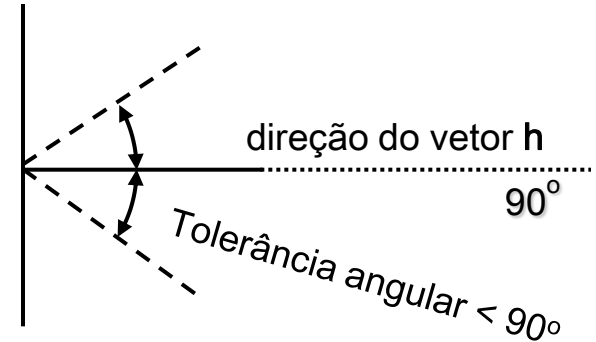
Semivariograma $\gamma(h)$



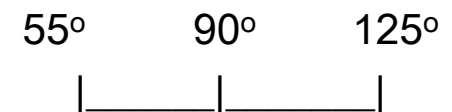
Cálculo do semivariograma a partir de amostras irregularmente espaçadas.

Semivariograma direcional => tolerância angular $< 90^\circ$

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

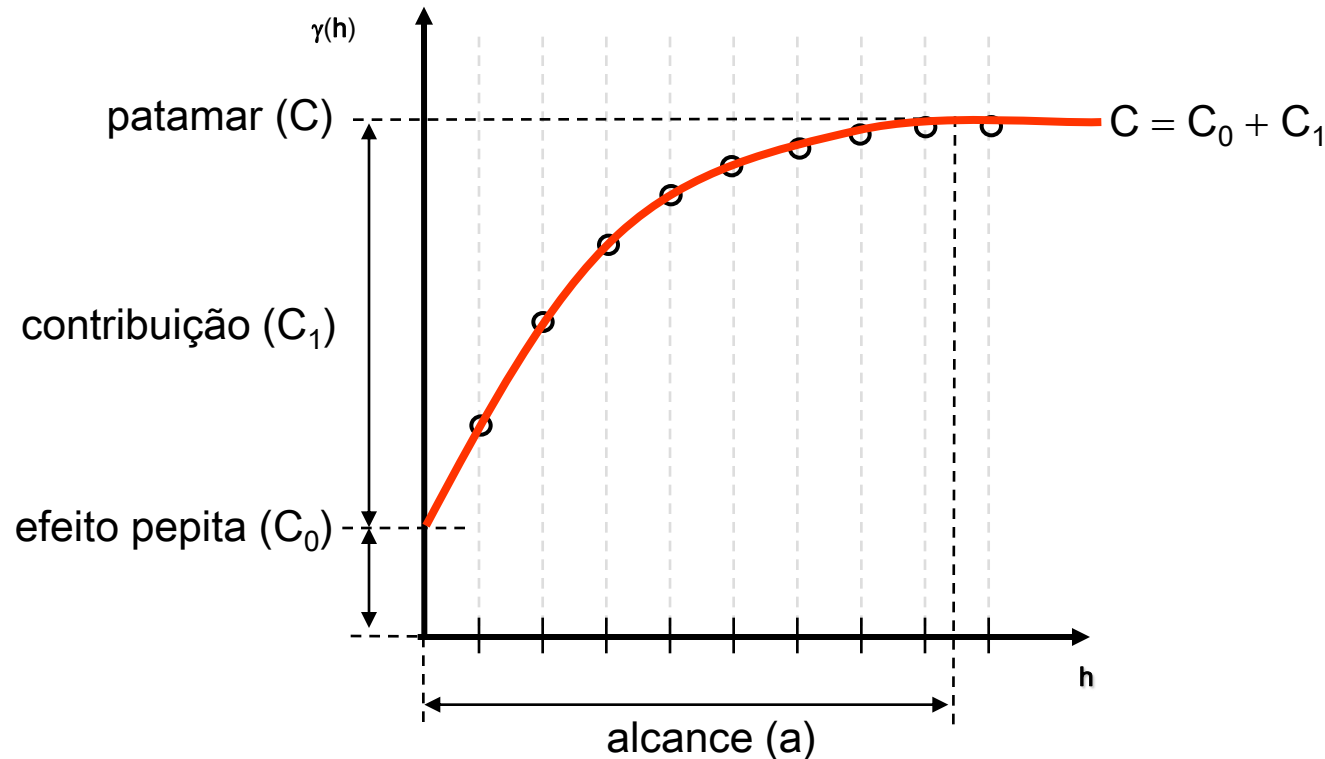


Exemplo:
 incremento (lag) = 1 km
 tolerância lag = 0,5 km
 direção de análise = 90°
 tolerância angular = 35°



Modelos teóricos de semivariograma

O gráfico do semivariograma empírico estimado por $\hat{\gamma}(h)$ é formado por uma série de valores, sobre os quais se objetiva ajustar uma função.



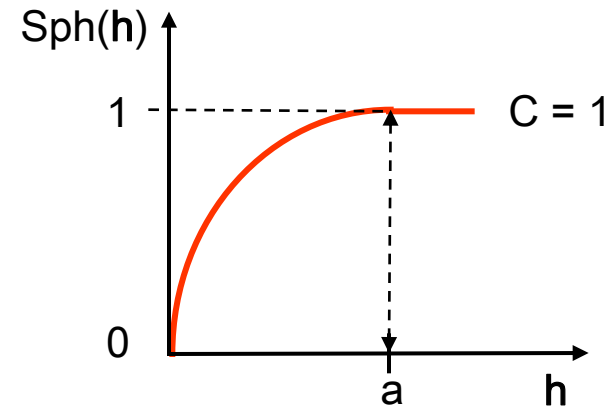
O modelo de ajuste deve representar o melhor possível o comportamento de $\gamma(h)$.

Modelos teóricos de semivariograma

Modelo de ajuste esférico

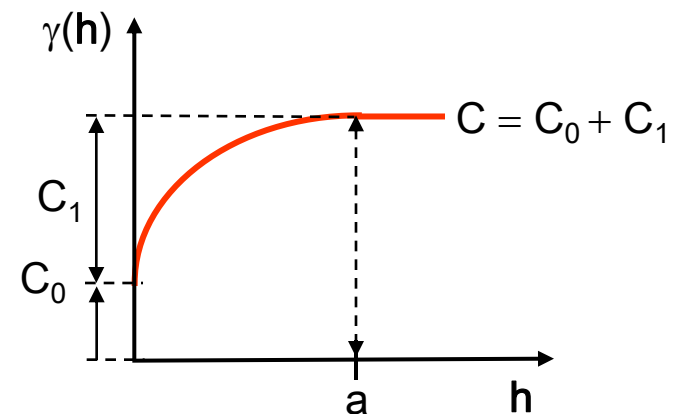
- Normalizado

$$\text{Sph}(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}|=0 \\ 1,5\left(\frac{|\mathbf{h}|}{a}\right) - 0,5\left(\frac{|\mathbf{h}|}{a}\right)^3 & , 0 < |\mathbf{h}| \leq a \\ 1 & , |\mathbf{h}| > a \end{cases}$$



- Na prática: $C_0 > 0$ e $C_1 > 1$

$$\gamma(\mathbf{h}) = \begin{cases} 0, & C_0 \\ C_0 + C_1 \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{a} \right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a} \right)^3 \right] = C_0 + C_1 [\text{Sph}(\mathbf{h})], & 0 < |\mathbf{h}| \leq a \\ C_0 + C_1, & |\mathbf{h}| > a \end{cases}$$

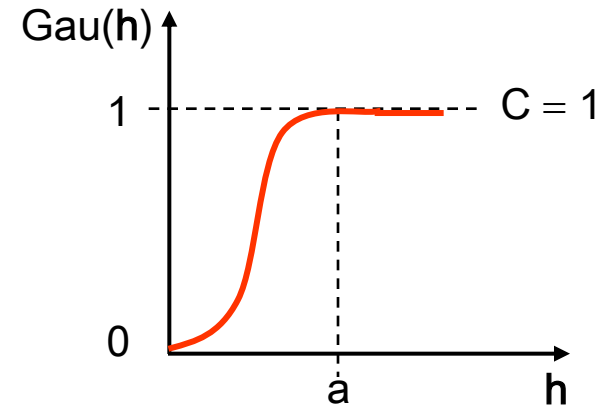


Modelos teóricos de semivariograma

Modelo de ajuste gaussiano

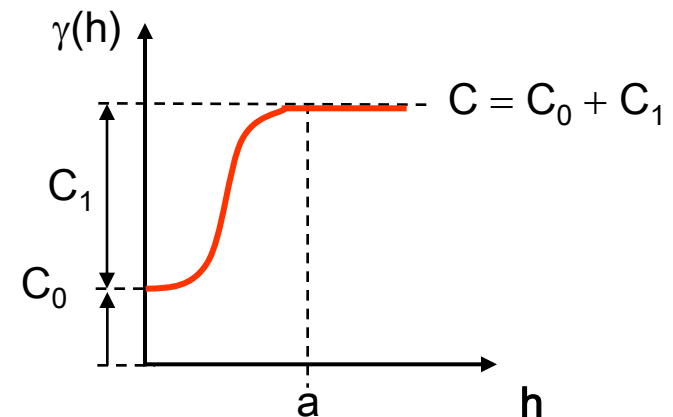
- Normalizado

$$\text{Gau}(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ 1 - \exp\left(-\frac{|\mathbf{h}|^2}{a}\right) & , |\mathbf{h}| \neq 0 \end{cases}$$



- Na prática: $C_0 > 0$ e $C_1 > 1$

$$\gamma(\mathbf{h}) = \begin{cases} 0, & C_0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{|\mathbf{h}|^2}{a}\right) \right] = C_0 + C_1 [\text{Gau}(\mathbf{h})], & 0 < |\mathbf{h}| \leq a \\ C_0 + C_1, & |\mathbf{h}| > a \end{cases}$$

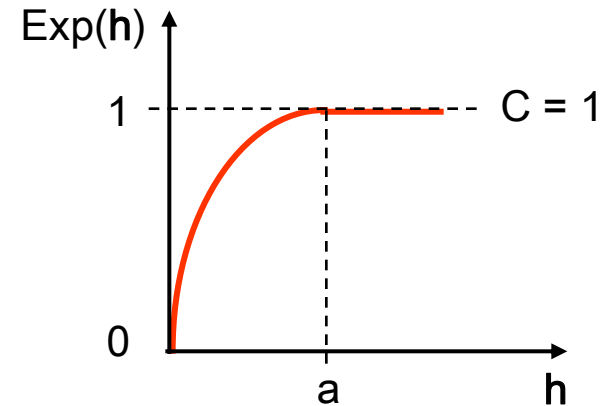


Modelos teóricos de semivariograma

Modelo de ajuste exponencial

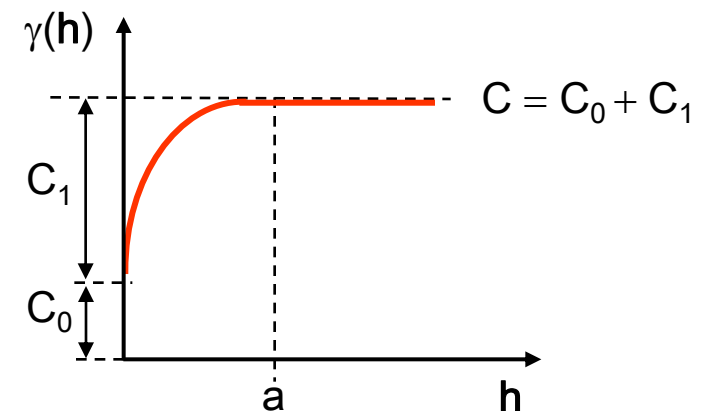
- Normalizado

$$\text{Exp}(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ 1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right) & , |\mathbf{h}| \neq 0 \end{cases}$$



- Na prática: $C_0 > 0$ e $C_1 > 1$

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \\ C_0 + C_1 \left[1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right) \right] = C_0 + C_1 [\text{Exp}(\mathbf{h})], & 0 < |\mathbf{h}| \leq a \\ C_0 + C_1, & |\mathbf{h}| > a \end{cases}$$

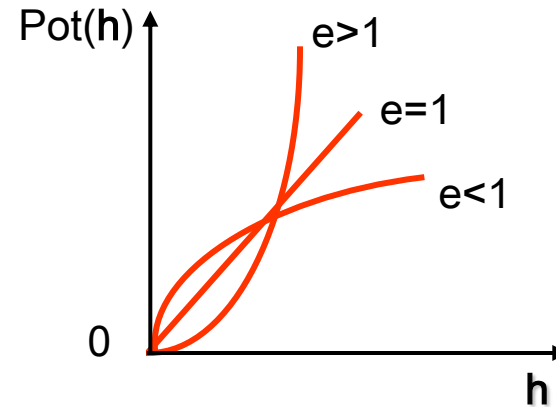


Modelos teóricos de semivariograma

Modelo de ajuste potência

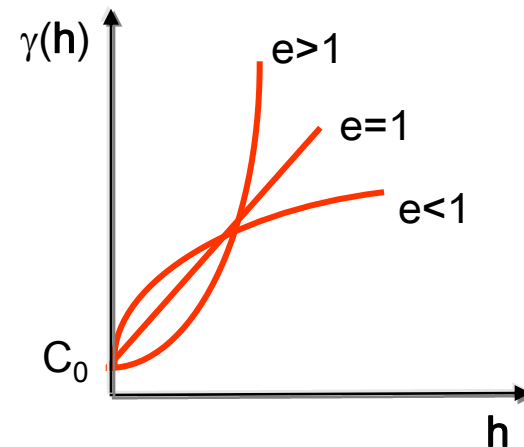
- Normalizado

$$\text{Pot}(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ c \cdot |\mathbf{h}|^e & , |\mathbf{h}| \neq 0 \end{cases}$$



- Na prática: $C_0 > 0$ e $C_1 > 1$

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , C_0 \\ C_0 + c \cdot |\mathbf{h}|^e = C_0 + \text{Pot}(|\mathbf{h}|) & , |\mathbf{h}| > 0 \end{cases}$$



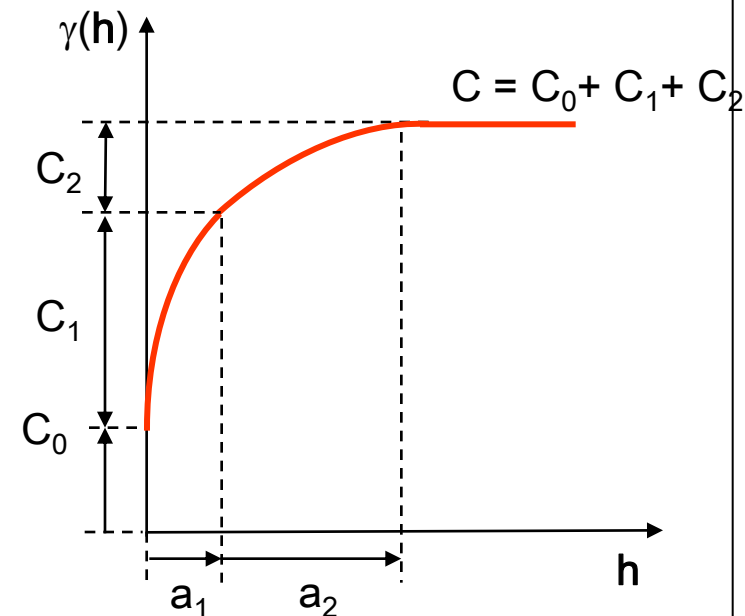
Modelos teóricos de semivariograma

Modelo de ajuste aninhados

Existem determinados fenômenos em que são necessários modelos mais complexos de semivariograma para explicar suas variações espaciais. Estes modelos são combinações de modelos simples, denominados *aninhados*.

Ex: Modelo aninhado duplo esférico

$$\gamma(\mathbf{h}) = \begin{cases} 0, & C_0 \\ C_0 + C_1 \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{a_1} \right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a_1} \right)^3 \right] = \gamma_1(\mathbf{h}), & 0 < |\mathbf{h}| \leq a_1 \\ C_0 + C_2 \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{a_2} \right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a_2} \right)^3 \right] = \gamma_2(\mathbf{h}), & a_1 < |\mathbf{h}| \leq a_2 \\ C_0 + C_1 + C_2, & |\mathbf{h}| > a_2 \end{cases}$$



Isotropia



Quando a variabilidade espacial de um fenômeno em estudo é a mesma em todas as direções, diz-se que o fenômeno é *isotrópico*.

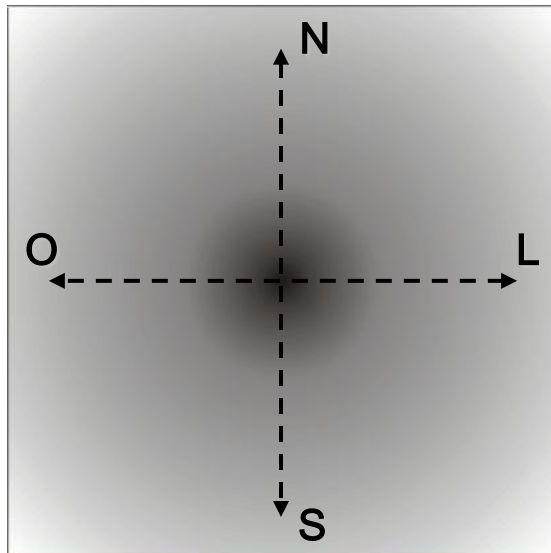
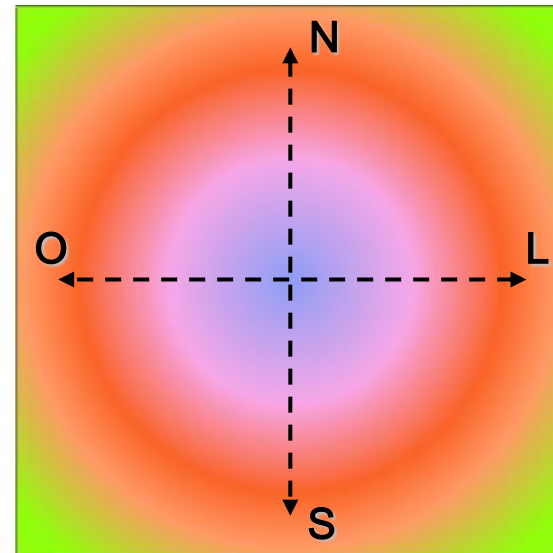


Imagem nível de cinza

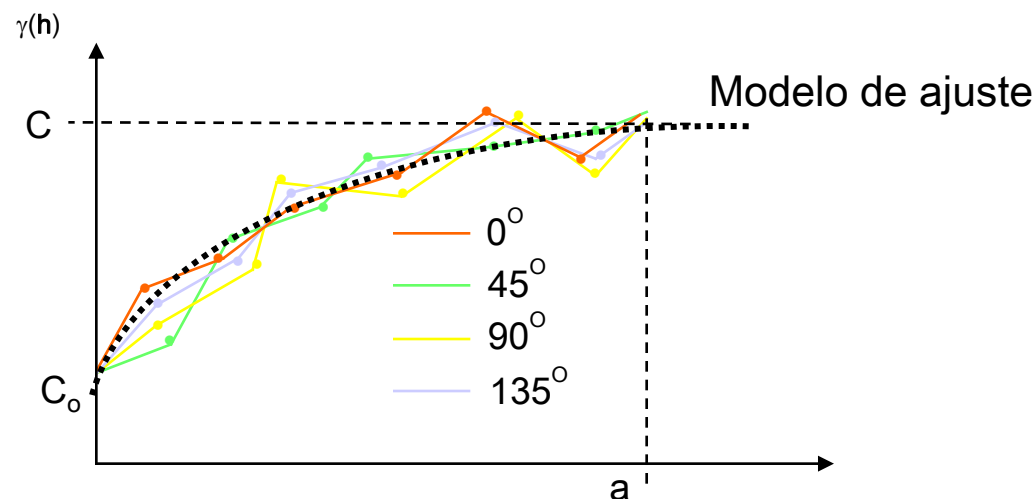


Composição Colorida

Isotropia



Considere os semivariogramas ilustrados na figura abaixo



Esta é a representação de um caso simples e menos freqüente, em que a distribuição espacial do fenômeno é denominada *isotrópica*.

Neste caso, um único modelo é suficiente para descrever a variabilidade espacial do fenômeno em estudo.

Anisotropia



Quando a variabilidade espacial de um fenômeno em estudo não é a mesma em todas as direções, diz-se que o fenômeno é *anisotrópico*.

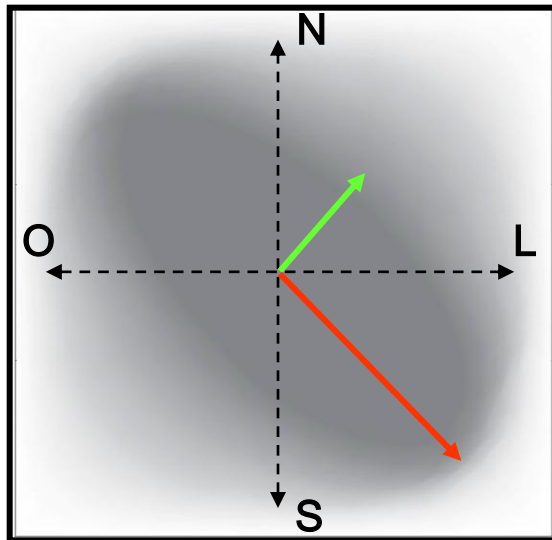
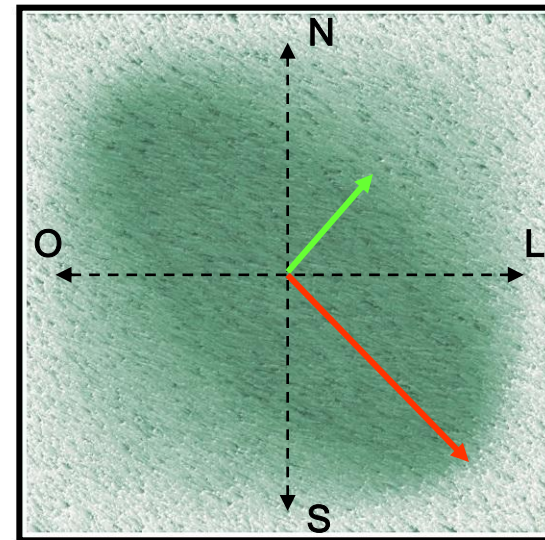


Imagem nível de cinza



Composição Colorida

direções de continuidade espacial

— maior

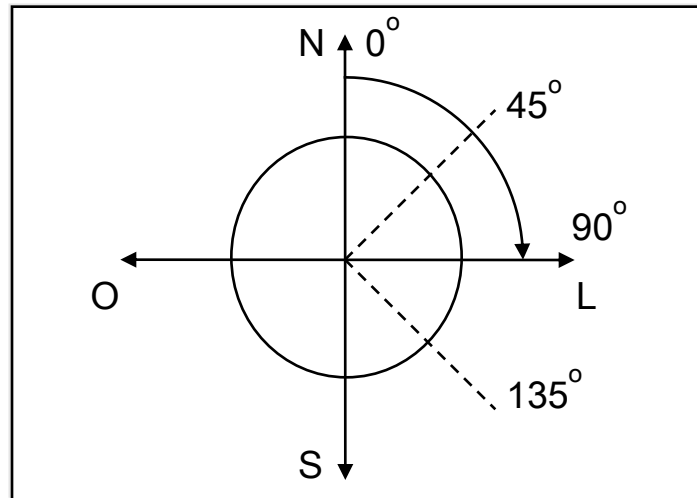
— menor

Anisotropia



A análise da ***anisotropia*** objetiva detectar as direções de maior e menor continuidade espacial do fenômeno investigado.

Convenções direcionais usadas na geoestatística

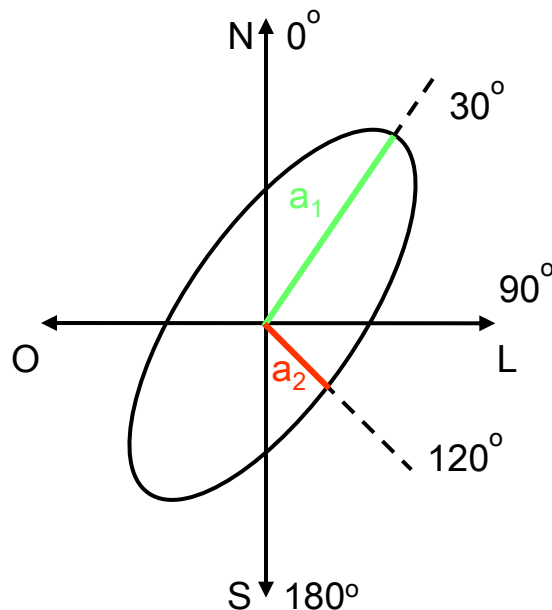


Uma forma de detectar a ***anisotropia*** é através da observação dos semivariogramas obtidos para diferentes direções.

Anisotropia



Um modo direto de visualizar e calcular os parâmetros (fator e ângulo) da anisotropia é através do esboço gráfico de uma elipse (ou *diagrama de rosa*).



Parâmetros da anisotropia

Fator de anisotropia (Fa)

$$Fa = a_2 / a_1$$

Ângulo de anisotropia (Aa)

Aa = tomado da direção Norte para o eixo de maior continuidade. No exemplo = 30°.

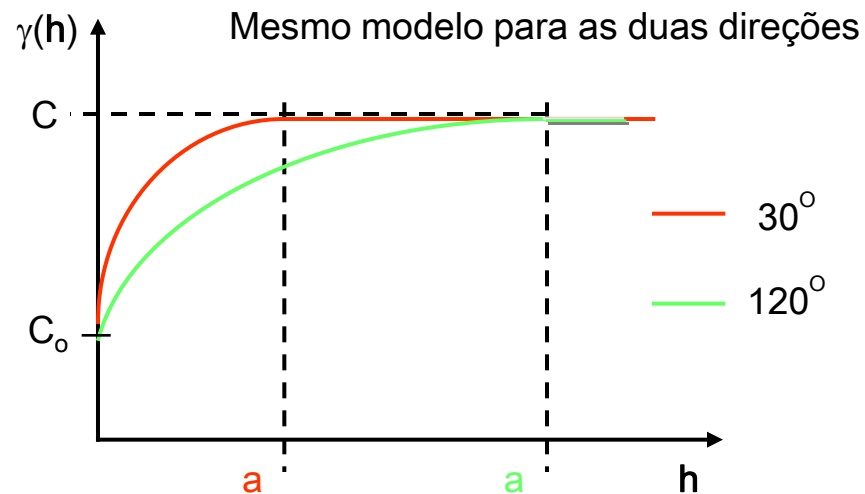
Tipos de anisotropia: geométrica, zonal e combinada.

Anisotropia



Anisotropia geométrica

Neste caso, os semivariogramas apresentam o mesmo patamar (C) com diferentes alcances (a) para o mesmo modelo.



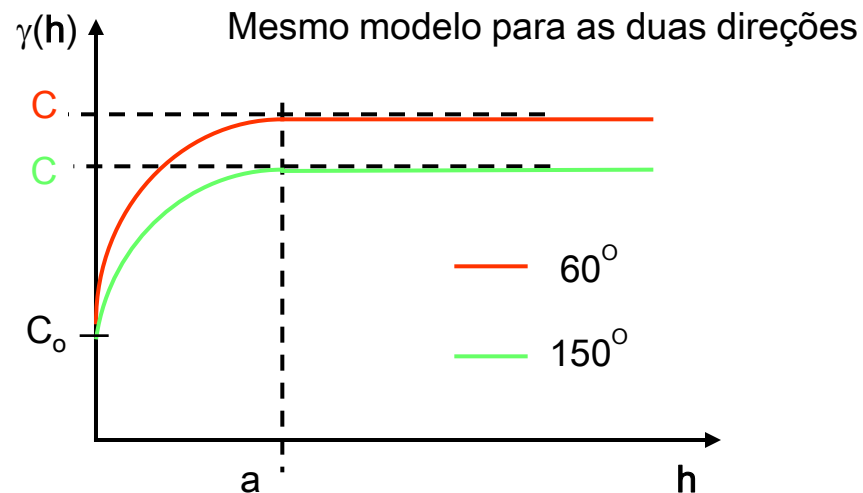
Anisotropia



Anisotropia zonal

Neste caso, os semivariogramas apresentam diferentes patamares (C) com mesmo alcance (a) para o mesmo modelo.

Como a isotropia, a **anisotropia zonal** é um caso menos freqüente presente nos fenômenos naturais.

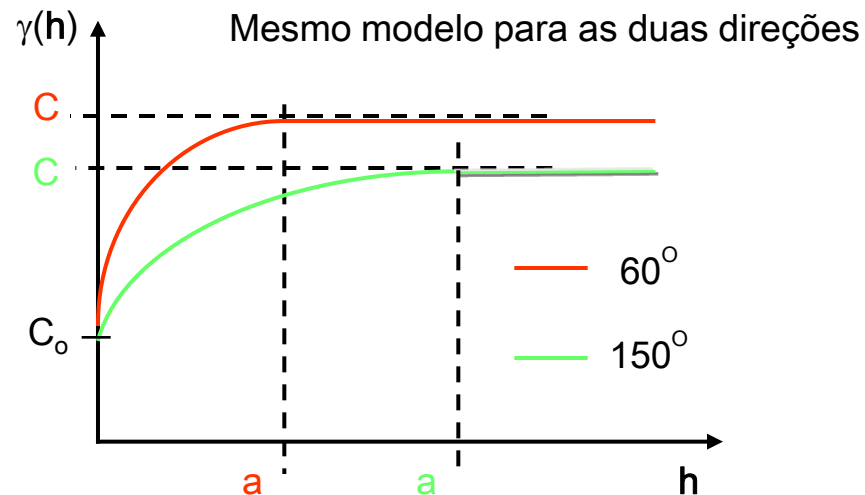


Anisotropia



Anisotropia combinada (geométrica + zonal)

Neste caso, os semivariogramas apresentam diferentes patamares (C) e diferentes alcances (a) para o mesmo modelo. Pode apresentar também diferentes efeitos pepita.

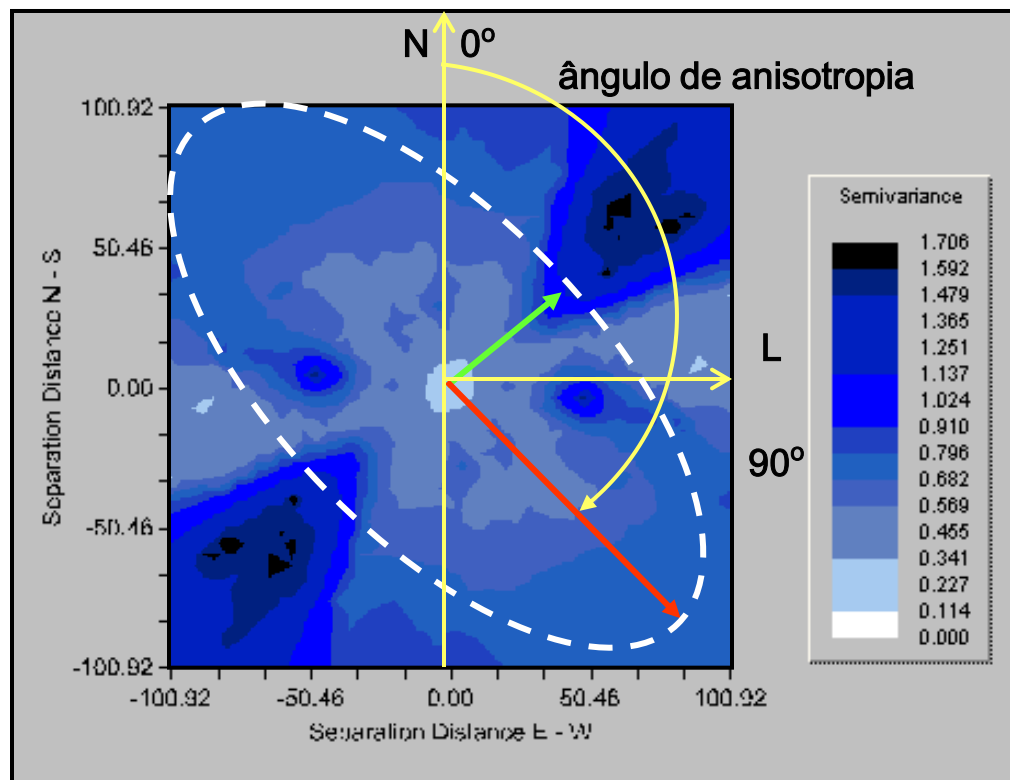


Semivariograma de superfície



É um gráfico 2D que fornece uma visão geral da variabilidade espacial do fenômeno em estudo. Também conhecido como *Mapa de Semivariograma*.

Utilizado para detectar os eixos de *Anisotropia* (direções de maior e menor continuidade espacial).

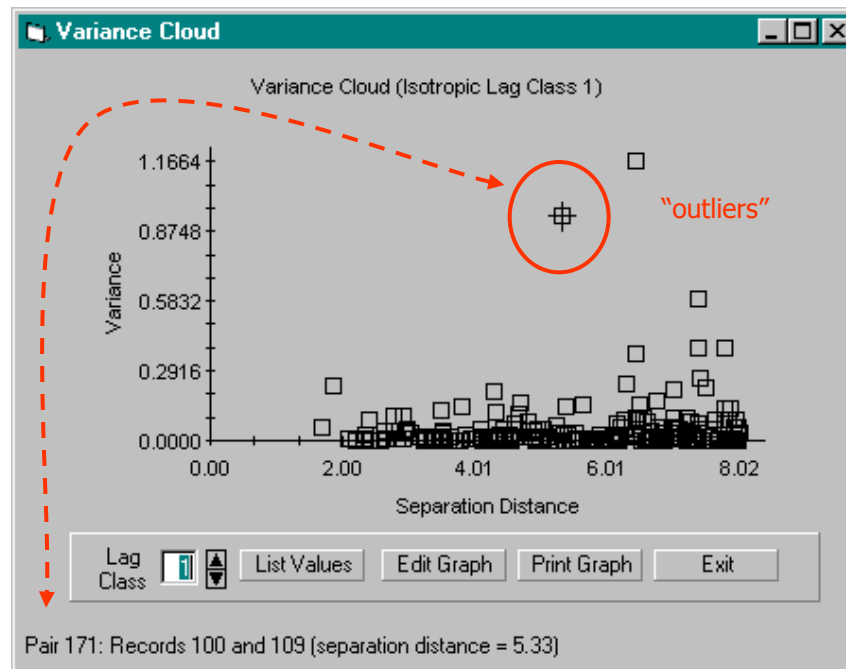


Semivariograma de nuvem



É um gráfico das semivariâncias de todos os pares de pontos tomados para um determinado lag (distância).

O variograma de nuvem é útil para detectar a presença de “outliers”.

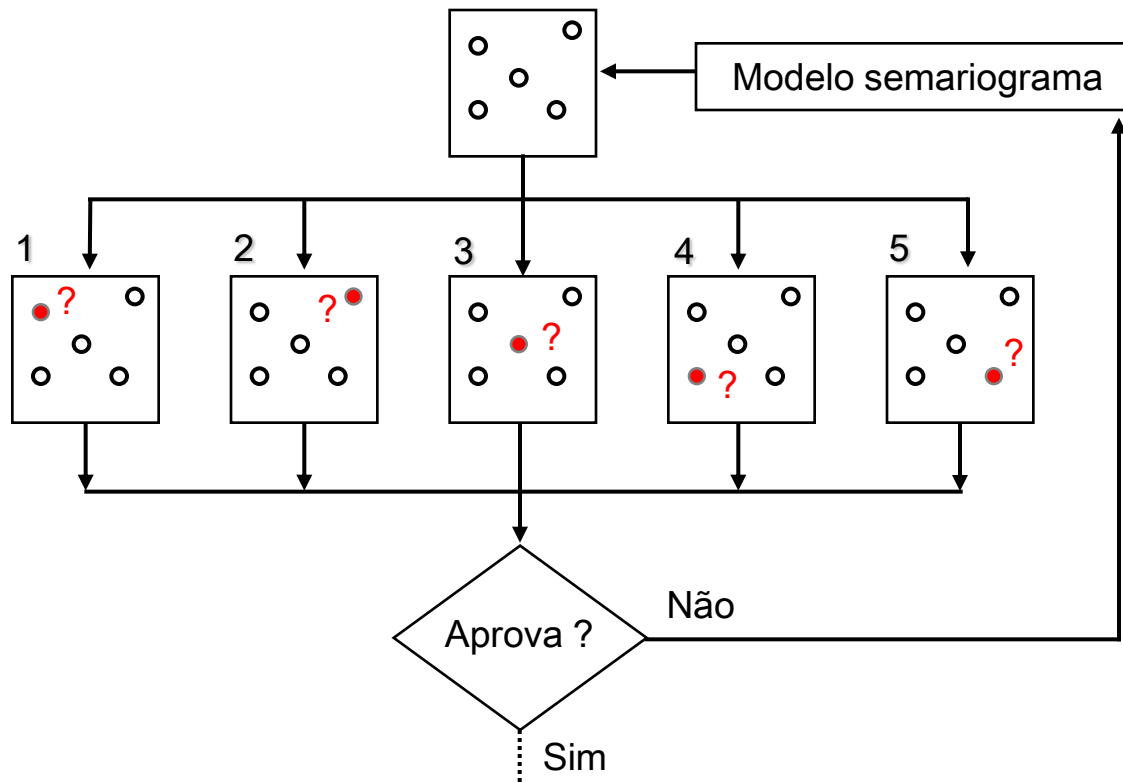


Pair	Variance	Distance	Point 1 Record #	Point 2 Record #
1	0.0013	9.02	1	7
2	0.0002	15.43	1	10
3	0.0088	9.26	1	16
4	0.0030	10.18	1	18
5	0.0030	15.44	1	19
6	0.0230	12.74	1	25
7	0.0522	14.99	1	26
8	0.0343	12.52	1	27
9	0.0126	15.16	2	4
10	0.0084	13.70	2	9
11	0.0219	13.63	2	11
12	0.0042	8.78	2	19
13	0.0314	9.49	2	20
14	0.0010	12.30	2	21
15	0.0160	14.34	2	28
16	0.0126	12.16	3	4
17	0.0219	10.90	3	11
18	0.0042	11.32	3	19
19	0.0314	10.00	3	20

Validação cruzada



É um procedimento para verificar a adequação do modelo de ajuste ao semivariograma



Análises

- estatísticas do erro
- histograma do erro
- diagrama espacial do erro
- diagrama de valores observados versus estimados

Validação cruzada

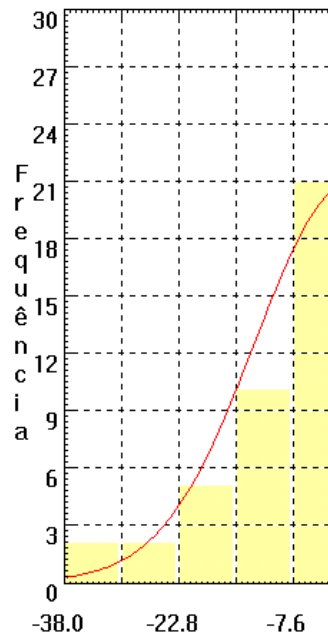


Análise de resultados

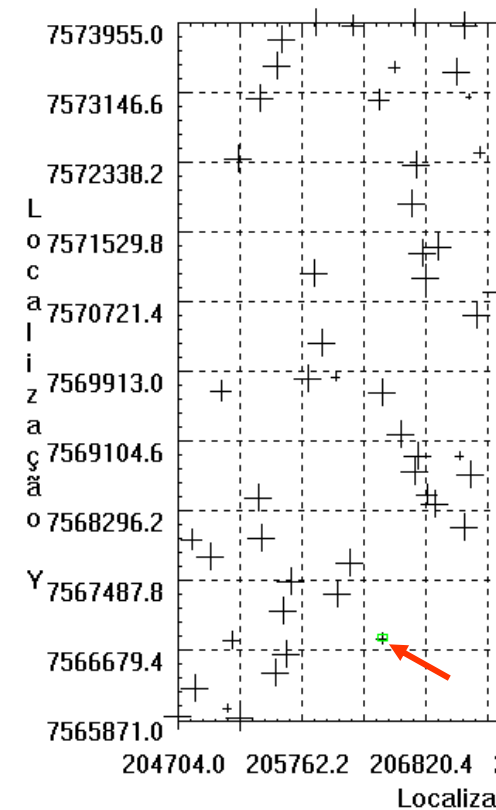
Histograma do Erro

Exibir Executar

Número de Classes: 10



Distribuição espacial do erro



N.pontos=85,Reestimados=85,Local..X=206455.094 Y=

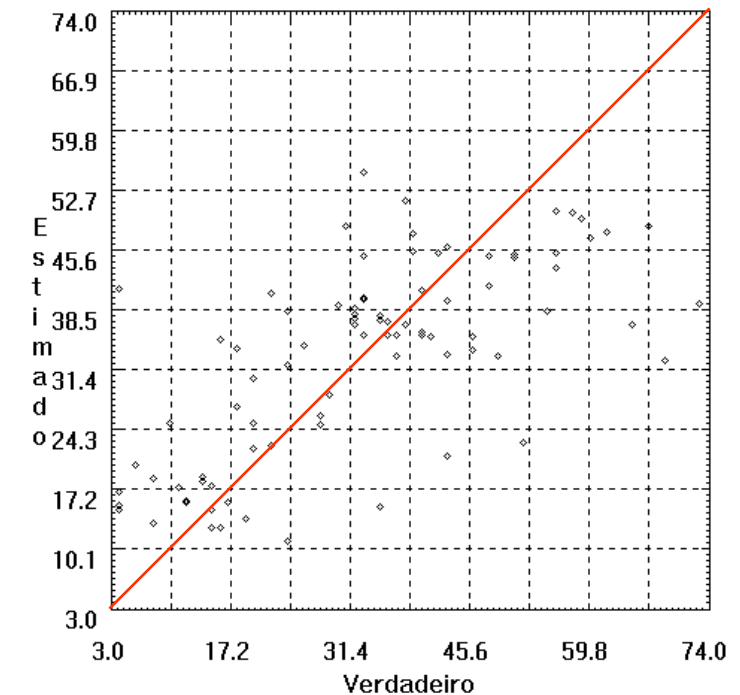
Relatório de Dados

STATISTICS OF THE ERROR

Amplitude

85
-0.252
153.789
12.401

Diagrama Verdadeiros X Estimados



N.pontos=85,Reestimados=85,Verd.=35.837,Estim.=35.483

Krigeagem



O termo **krigeagem** é derivado do nome *Daniel G. Krige*

A **krigeagem** é um estimador estocástico que depende da análise de correlação espacial baseada em semivariograma.

Áreas de Aplicações:

- mapeamento geológico (*Verly et al., 1984*)
- mapeamento solo (*Burgess e Webster, 1980*)
- mapeamento hidrológico (*Kitanidis et. al., 1983*)
- mapeamento atmosférico (*Lajaunie, 1984*)

A krigeagem engloba um conjunto de métodos de estimação que podem ser estacionários* ou não-estacionários:

- krigeagem Simples (*)
- krigeagem Ordinária (*)
- krigeagem Universal
- co-krigeagem
- krigeagem por indicação , outros

Krigeagem



A **krigeagem** é um estimador estocástico que depende da análise de correlação espacial baseada em semivariograma.

Na Krigeagem, o procedimento é semelhante ao de interpolação **por média móvel ponderada**, exceto que aqui os pesos são determinados a partir de uma **análise espacial**, baseada no **semivariograma experimental**.

Além disso, a krigagem fornece, em média, estimativas não tendenciosas e com variância mínima

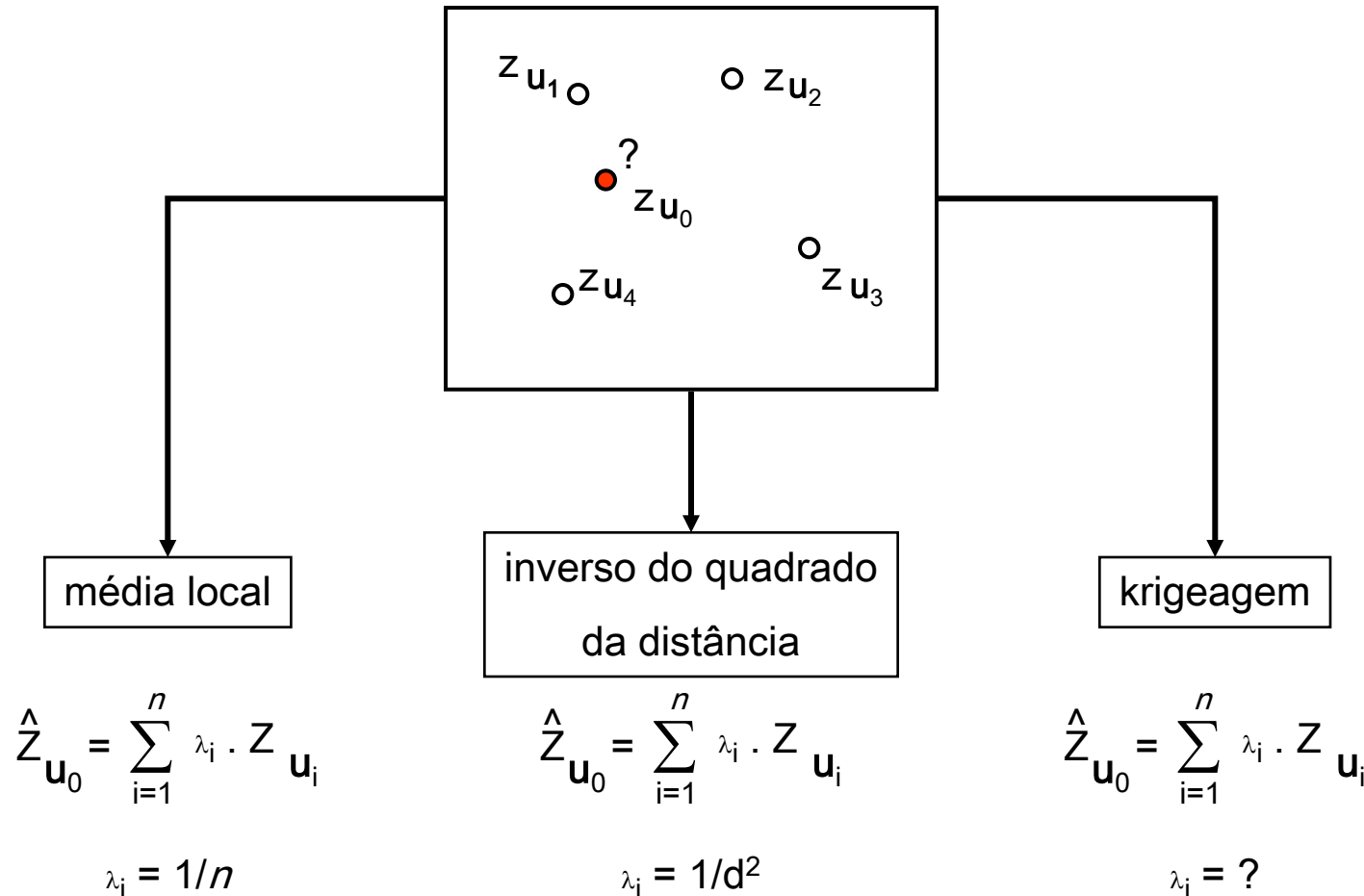
Estimativas não tendenciosas significam que, em média, a diferença entre valores estimados e observados para o mesmo ponto deve ser nula;

Variância mínima significa que estes estimadores possuem a menor variância dentre todos os estimadores não tendenciosos.

Krigeagem



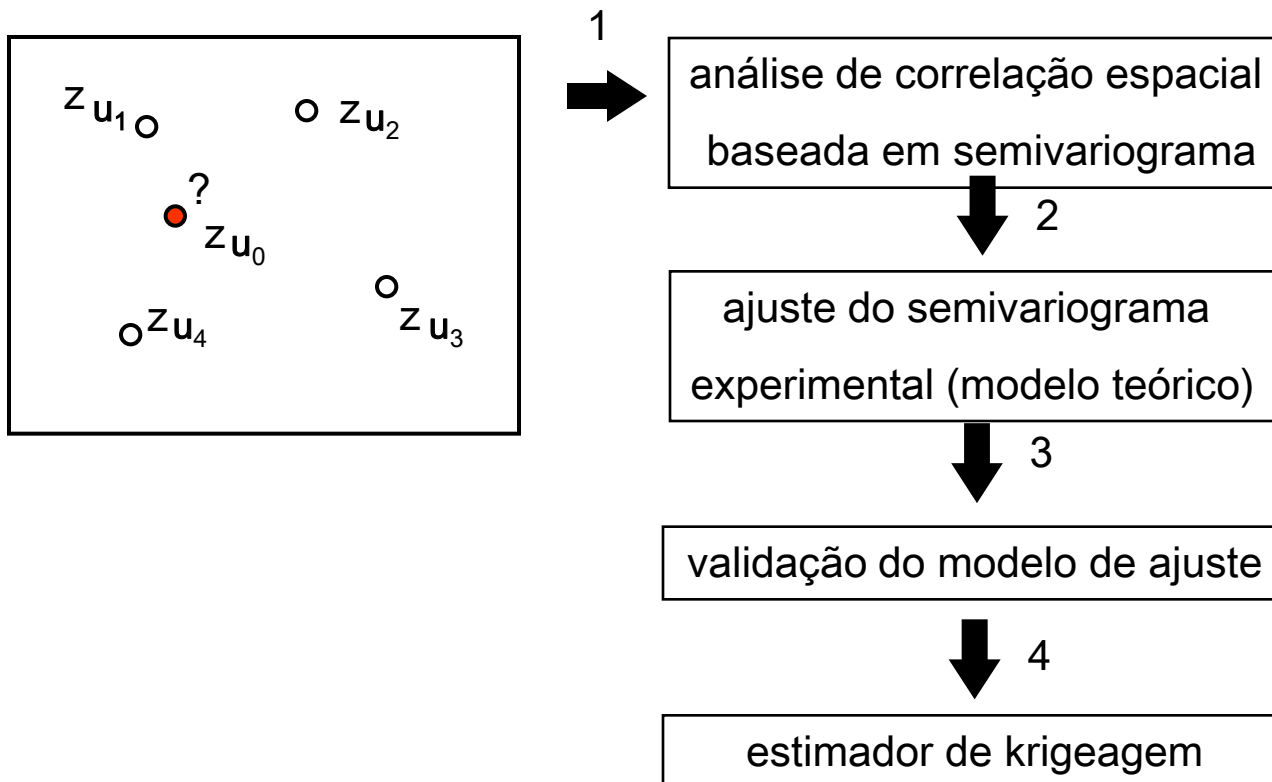
Envolve uma combinação linear de n valores em pontos vizinhos.



Krigeagem



Os pesos são calculados considerando a estrutura de correlação espacial imposta pelo semivariograma



Krigeagem



Segundo *Journal (1988)*: $\mathbf{K} \cdot \boldsymbol{\lambda} = \mathbf{k} \Rightarrow \boldsymbol{\lambda} = \mathbf{K}^{-1} \mathbf{k}$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \alpha \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}$$

Os elementos das matrizes de covariâncias são calculados da seguinte forma (*Journal, 1988*):

$$C_{ij} = C(\mathbf{0}) - \gamma(\mathbf{h}) = C_0 + C_1 - \gamma(\mathbf{h})$$

Substituindo os valores de C_{ij} nas matrizes encontram-se os pesos $\lambda_1, \lambda_2, \dots$, e λ_n .

Estimador de Krigeagem (*Journal, 1988*): $Z_{\mathbf{x}_0}^* = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)$

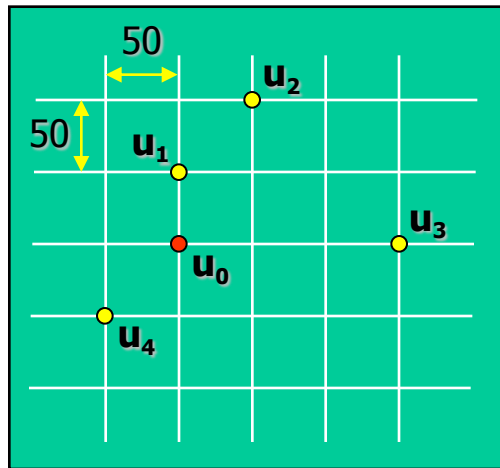
Variância de Krigeagem (*Journal, 1988*): $\sigma_{ko}^2 = (C_0 + C_1) - \boldsymbol{\lambda}^T \mathbf{k}$

Krigeagem



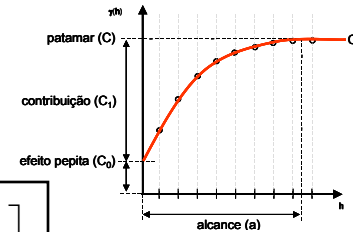
EXEMPLO

Considere o espaço amostral na figura abaixo. Deseja-se estimar o valor da variável Z no ponto u_0 , a partir de $z(u_1)$, $z(u_2)$, $z(u_3)$ e $z(u_4)$. Considere ainda, que o semivariograma empírico foi ajustado através de um modelo esférico, com $a = 200$, $C_1 = 20$, e $C_0 = 2$.



krigeagem ordinária

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \alpha \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & 1 \\ C_{21} & C_{22} & C_{23} & C_{24} & 1 \\ C_{31} & C_{32} & C_{33} & C_{34} & 1 \\ C_{41} & C_{42} & C_{43} & C_{44} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{01} \\ C_{02} \\ C_{03} \\ C_{04} \\ 1 \end{bmatrix}$$

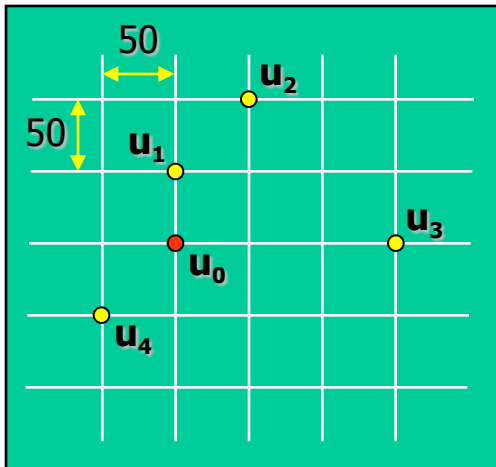


Os elementos das matrizes são calculados: $C_{ij} = C_0 + C_1 - \gamma(h)$ Modelo Teórico

$$C_{12} = C_{21} = C_{04} = C_0 + C_1 - \gamma(50\sqrt{2})$$

$$= (2+20) - \left[2 + 20 \left(1,5 \frac{50\sqrt{2}}{200} - 0,5 \frac{(50\sqrt{2})^3}{(200)^3} \right) \right] = 9,84$$

EXEMPLO



$$C_{13} = C_{31} = (C_0 + C_1) - \gamma [\sqrt{(150)^2 + (50)^2}] = 1,23$$

$$C_{14} = C_{41} = C_{02} = (C_0 + C_1) - \gamma [\sqrt{(100)^2 + (50)^2}] = 4,98$$

$$C_{23} = C_{32} = (C_0 + C_1) - \gamma [\sqrt{(100)^2 + (100)^2}] = 2,33$$

$$C_{24} = C_{42} = (C_0 + C_1) - \gamma [\sqrt{(100)^2 + (150)^2}] = 0,29$$

$$C_{34} = C_{43} = (C_0 + C_1) - \gamma [\sqrt{(200)^2 + (50)^2}] = 0$$

$$C_{01} = (C_0 + C_1) - \gamma (50) = 12,66$$

$$C_{03} = (C_0 + C_1) - \gamma (150) = 1,72$$

$$C_{11} = C_{22} = C_{33} = C_{44} = (C_0 + C_1) - \gamma (0) = 22$$

Krigeagem



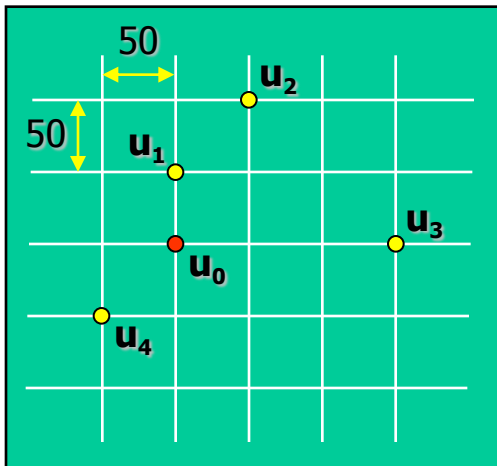
EXEMPLO

Substituindo os valores de C_{ij} nas matrizes, encontra-se os seguintes pesos:

$$\lambda_1 = 0,518 \quad \lambda_2 = 0,022 \quad \lambda_3 = 0,089 \quad \lambda_4 = 0,371$$

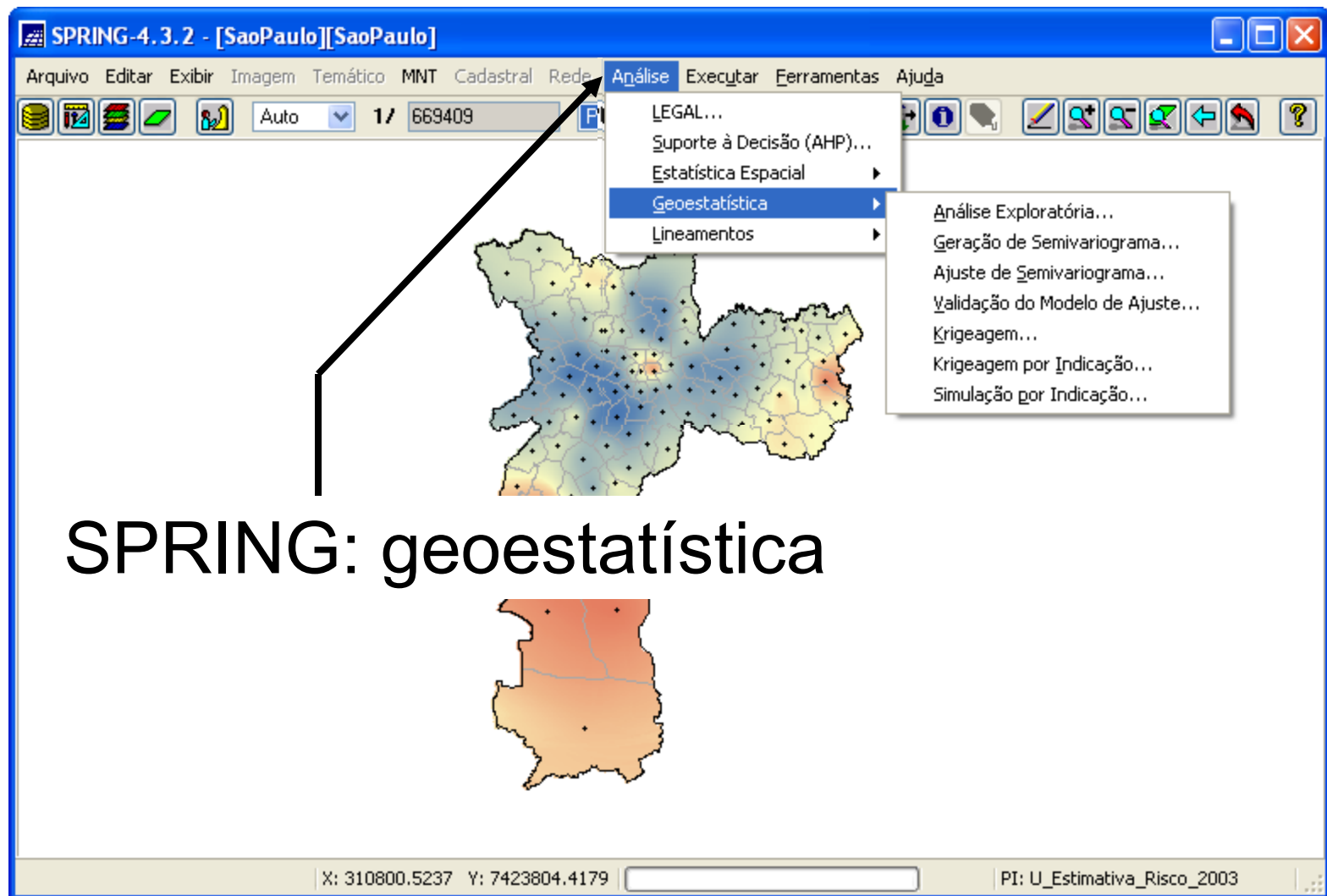
Finalmente o valor estimado é dado por:

$$\hat{Z}(u_0) = 0,518 z(u_1) + 0,022 z(u_2) + 0,089 z(u_3) + 0,371 z(u_4)$$



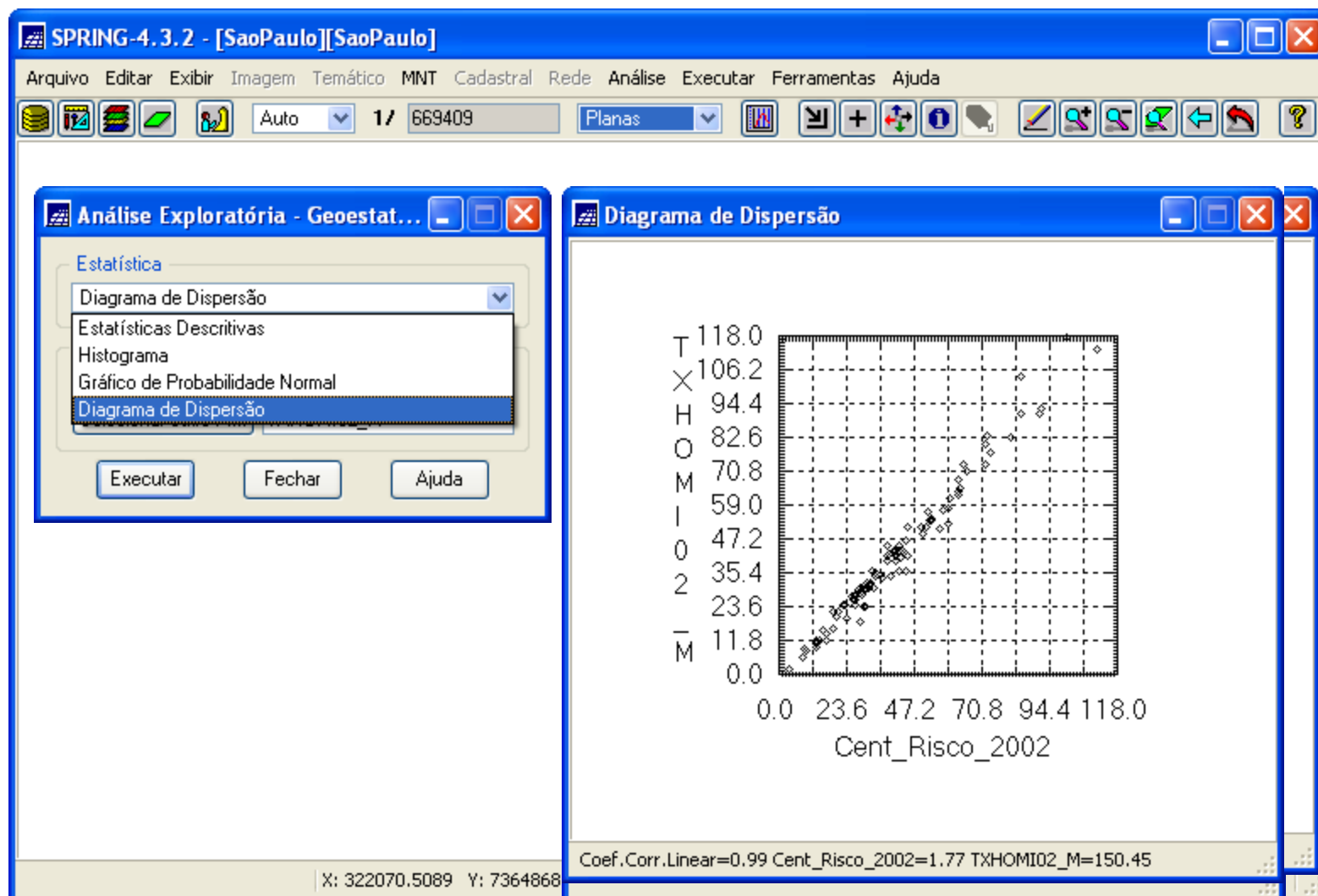
COMENTÁRIO: embora as amostras Z_2 e Z_3 tenham pouca influência na estimativa final de Z_0 , suas influências não são lineares em relação às suas distâncias a partir de Z_0 . A amostra Z_3 está mais distante que Z_2 ; no entanto, tem mais influência, 8,9%, que Z_2 , 2,2%. Isto ocorre porque Z_0 está diretamente sobre a influência de Z_3 , enquanto Z_2 está muito próximo de Z_1 . Ao se introduzir as covariâncias no cálculo dos pesos, evita-se associar pesos indevidos a “clusters” (agrupamentos) de amostras, o que não ocorre com outros métodos baseados somente na distância.

SPRING e geoestatística

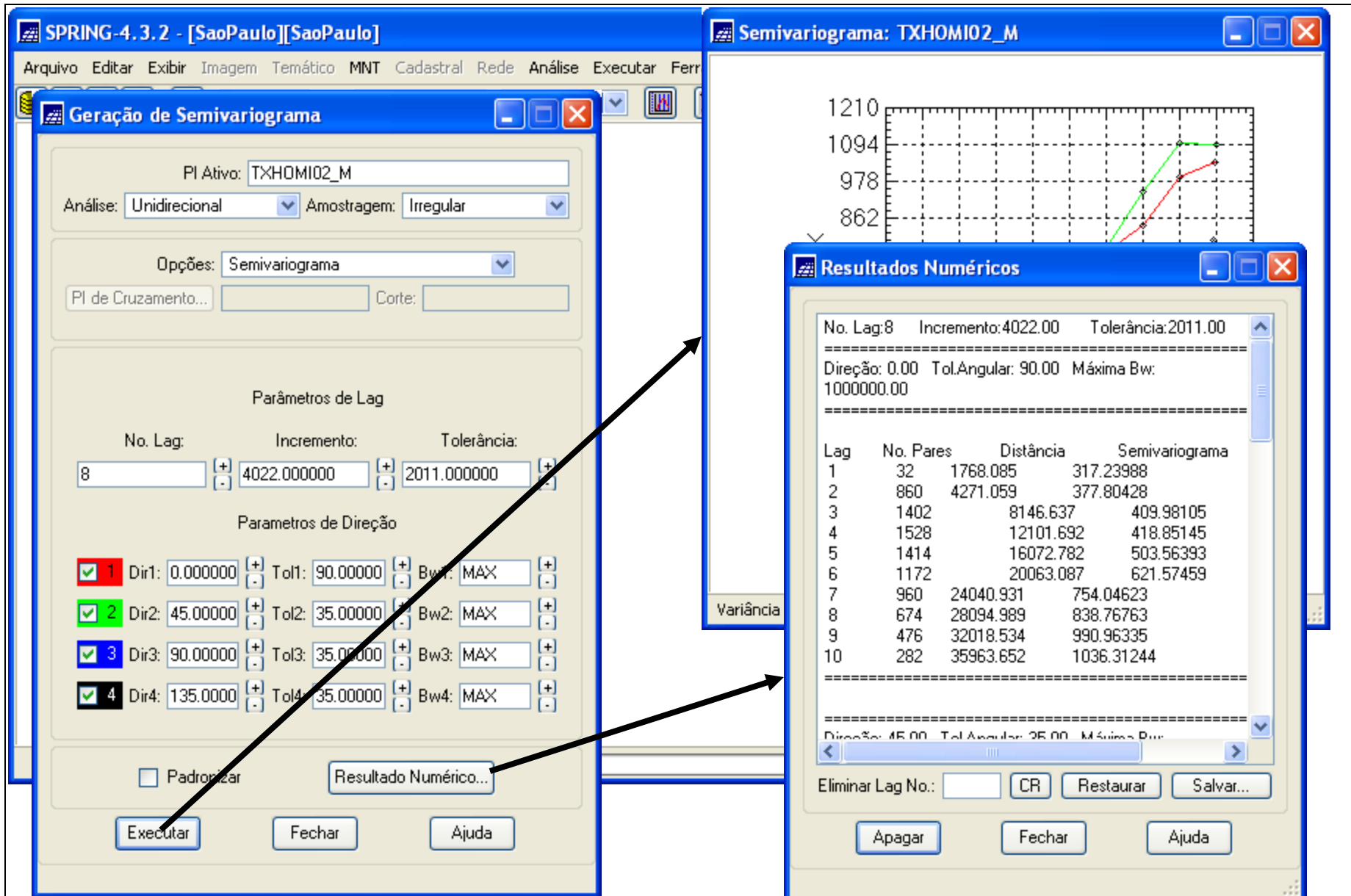


SPRING: geoestatística

SPRING e geoestatística



SPRING e geoestatística



SPRING e geoestatística



SPRING-4.3.2 - [SaoPaulo][SaoPaulo]

Arquivo Editar Exibir Imagem Temático

Auto 1/

Ajuste de Semivario...

Ajuste

☒ Automático ☐ Visual

Número de Estruturas

☒ 1 ☐ 2 ☐ 3

Modelos

Modelo 1:

Modelo 2:

Modelo 3:

Verificar Ajustes

TXHOMI02_M_0.var

TXHOMI02_M_45.var

TXHOMI02_M_90.var

TXHOMI02_M_135.var

Parâmetros Estruturais

Definir...

Executar Fechar Ajuda

Parâmetros Estruturais

Parâmetros

Número de Estruturas: ☒ 1 ☐ 2 ☐ 3

Efeito Pepita: 169.345

Primeira Estrutura

Tipo:

Contribuição: 509.714 Ângulo Anis.: 0

Alcance Máx.: 35077.21 Alcance Mín.: 35077.21

Segunda Estrutura

Tipo:

Contribuição:

Ângulo Anis.:

Alcance Máx.:

Alcance Mín.:

Terceira Estrutura

Tipo:

Contribuição:

Ângulo Anis.:

Alcance Máx.:

Alcance Mín.:

Executar Fechar Ajuda

10

42.730

vo: Esférico

620.748

.291

rama Esférico

	Efeito Pepita	Contribuição	Alcance
2	-43.002	173.707	480.713
3	-43.258	169.345	509.714

Salvar...

Apagar Fechar Ajuda

SPRING e geoestatística



SPRING-4.3.2 - [SaoPaulo][SaoPaulo]

Arquivo Editar Exibir Imagem Temático MNT Cadastral Rede Análise Executar Ferramentas Ajuda

Auto 1/ 669409 Planas

Validação do Modelo

PI Ativo

Nome: TXHOMI02_M Verificar Modelo...

Parâmetros de Interpolação

Número de Pontos no Elipsóide de Busca

Mínimo: 4 Máximo: 16

Elipsóide de Busca (Raio e Orientação)

R.Mín.: 93742.7 R.Máx.: 93742.7 Ângulo: 0

Resultados

Numérico

Diagrama Espacial do Erro

Histograma do Erro

Estatísticas do Erro

Diagrama Observado x Estimado

Numérico

Relatório de Dados

Infolayer: TXHOMI02_M

7 = Number of Variables

X Localization

Y Localization

Z Localization (not used)

Observed

Estimated

Variance Estimation

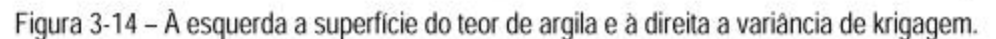
Error = Estimated - Observed

339408.500	7393048.000	0.000	23.710	26.575	241.579	2.865
325152.375	7394657.500	0.000	11.490	23.936	242.123	12.446
317380.656	7407413.000	0.000	11.160	25.010	262.904	13.850
345096.563	7392355.000	0.000	44.370	36.740	245.088	-7.630
348376.875	7396045.000	0.000	46.230	40.977	246.226	-5.253
329211.344	7397718.500	0.000	23.910	25.263	234.832	1.353
331758.750	7393425.500	0.000	11.320	25.043	224.516	13.723
337172.875	7396055.500	0.000	23.520	39.349	235.815	15.829
332882.969	7397314.500	0.000	39.360	38.044	227.058	-1.316
335082.625	7395369.000	0.000	103.760	33.671	225.486	-70.089

Salvar...

Apagar Fechar Ajuda

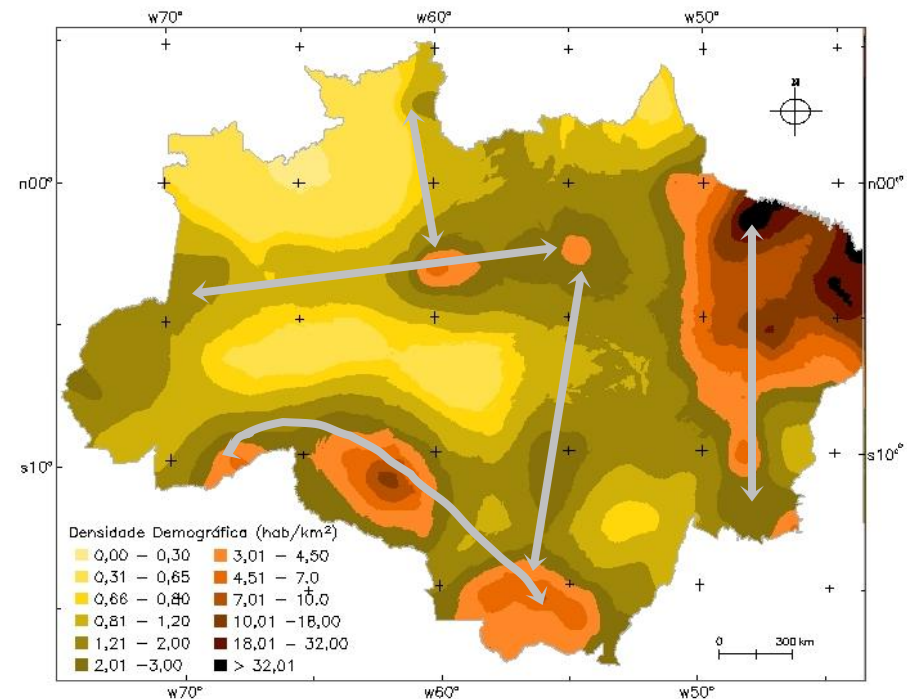
X: 322070.5089 Y: 7364868.0635



População na AmzL

Superfícies de Densidade - Krigeagem

- Densidade de População Total – sedes de município
- Das superfícies-> Mapas temáticos para análise
- Mapas de tendências, imprecisos e generalizados
- Evolução das superfícies - > tendências demográficas



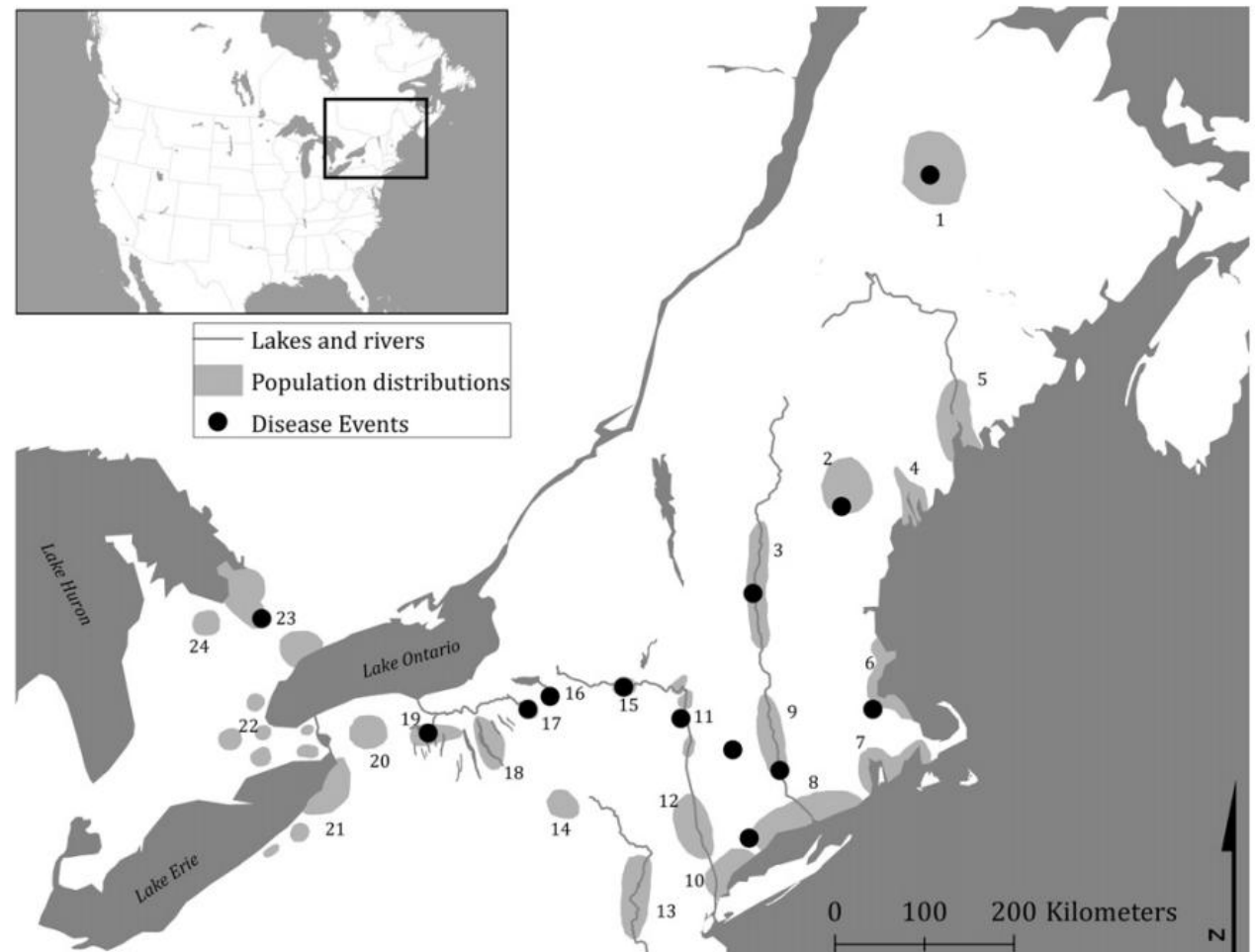
Densidade de População - 2000

Journal of Anthropological Archaeology 31 (2012) 83–92

Using spatial analysis to estimate depopulation for Native American populations in northeastern North America, AD 1616–1645

Eric E. Jones^{a,*}, Sharon N. DeWitte^b

E.E. Jones, S.N. DeWitte / Journal of Anthropological Archaeology 31 (2012) 83–92



Spatial analysis to estimate depopulation for Native American populations



Objetivos:

- mapear o conhecimento sobre a distribuição da população e do despovoamento (% pop perdida) das sociedades nativas americanas do sec 17 (NE US).
 - Padrões espaciais de despovoamento (*depopulation*);
- Usar interpolação espacial para determinar a *depopulation* dos grupos nativos para os quais não se tem info de população
 - Preencher lacunas de informação para entender o processo de *depopulation*
- *Final goal*: discutir fatores culturais, biológicos e históricos da *depopulation*

Spatial analysis to estimate depopulation for Native American populations

Métodos:

- Dados de distribuição espacial de publicações existentes e surveys arqueológicos
- Mapeamento prévio dos autores



Table 1

The 17th-century Native American cultures of the Northeast used in this study and the source of their geographic information. Those in *italics* previously had no population information but have depopulation percentage estimates from our results below.

Culture	Sources on geographic distributions
Maliseet-Passamaquoddy	Snow (1980, p. 341)
Mohegan-Pequot	Dincauze (1990)
<i>Kennebec</i>	Piotrowski (2002)
Mahican	Bradley (2007)
Eastern Abenaki	Snow (1980, p. 62)
Western Abenaki	Snow (1980, p. 342)
Massachusetts	Dincauze (1990) and Bragdon (1996)
Pocumtuc	Dincauze (1990) and Piotrowski (2002)
Quiripi-Unquachog	Dincauze (1990)
<i>Narragansett-Pokanoket</i>	Dincauze (1990)
Mohawk	Snow (1995a)
Oneida	Jones (2010a, 2010c)
Onondaga	Jones (2010a, 2010c)
Cayuga	Jones (2010c)
Seneca	Wray et al. (1987) and Jones (2010c)
Wendat	Engelbrecht (2003) and Warrick (2008)
<i>Neutral</i>	Snow (1994) and Engelbrecht (2003)
<i>Wenro</i>	Snow (1994) and Engelbrecht (2003)
<i>Erie</i>	Snow (1994) and Engelbrecht (2003)
<i>Munsee</i>	Dincauze (1990)
<i>Susquehannock</i>	Snow (1994) and Engelbrecht (2003)
Minisink	Kraft (1991)

- Digitalizaram informações dos mapas para um GIS
- Usando localizações dos sites ocupados no período, criaram *shapefiles* das extensões geográficas que os continhamfigura

Spatial analysis to estimate (populations

Métodos:

Estimativas de população
pré e pós contato
também de trabalhos
anteriores;
(diferentes acurácias)

Tabela 2 - mais precisos

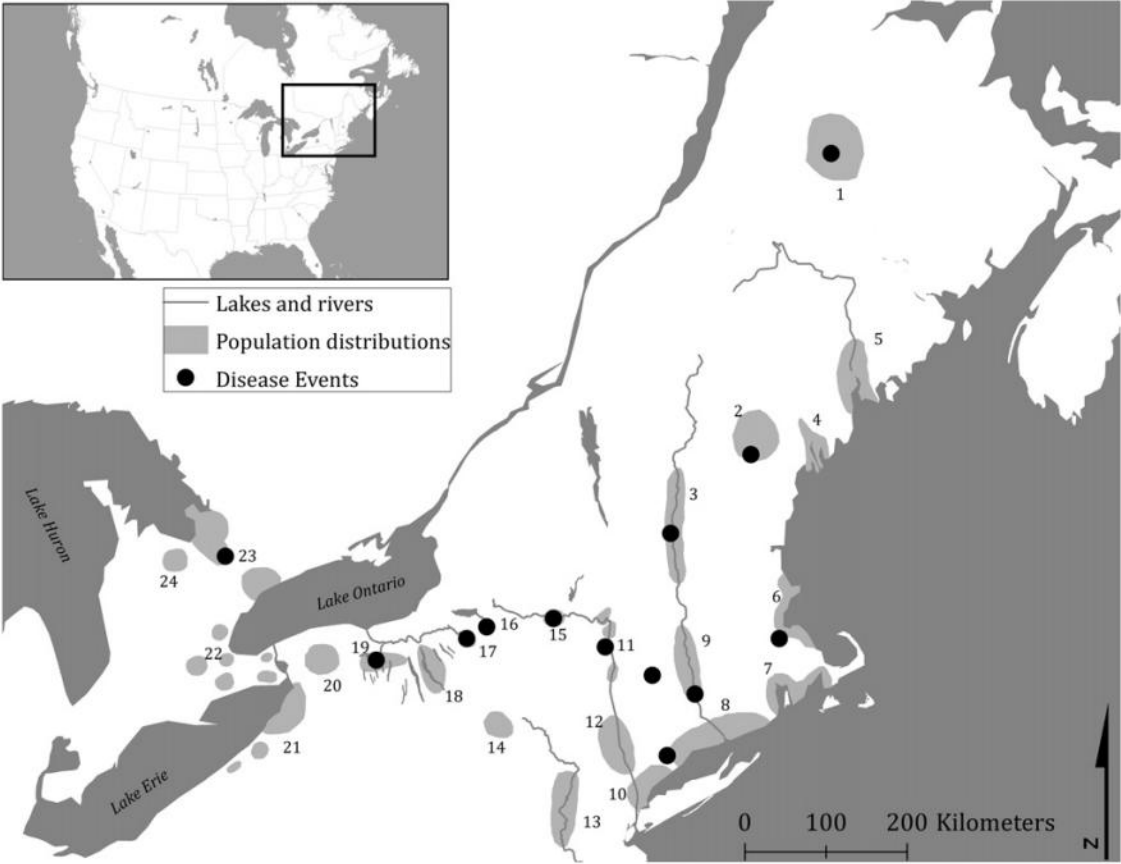


Table 2
Population data for the 17th-century Native American populations used in this study, including the year of the depopulation event being studied, beginning and ending populations for the depopulation event during the indicated year, percentage of total population that died, and sources.

Population	Date	Initial population	Resulting population	Percent lost	References
Maliseet-Passamaquoddy	1633	7600	2500	67	Snow and Lanphear (1988, p. 24)
Mohegan-Pequot	1633	16,000	3000	81	Bradford (1952, p. 260)
Mahican	1633	6400	500	92	Snow and Lanphear (1988, p. 24)
Eastern Abenaki	1633	13,800	3000	78	Snow and Lanphear (1988, p. 24)
Western Abenaki	1633	12,000	250	98	Snow and Lanphear (1988, p. 24)
Massachusetts	1633	44,000	6400	86	Snow and Lanphear (1988, p. 24)
Pocumtuc	1633	18,400	920	95	Snow and Lanphear (1988, p. 24)
Quiripi-Unquachog	1633	29,900	1500	95	Snow and Lanphear (1988, p. 24)
Mohawk	1633	8000	1750	78	Snow (1995a, 1995b), Jones (2010a), and Snow and Lanphear (1988, p. 24)
Seneca	1634	4850	3150	35	Jones (2010b)
Oneida	1635	1500	400	73	Jones (2010a)
Wendat	1639	20,000	10,000	50	Heidenreich (1976), Trigger (1976, pp. 588–589), and Warrick (2008)
Onondaga	1645	2700	1150	57	Jones (2010a)

Spatial analysis to estimate depopulation for Native American populations

Métodos:

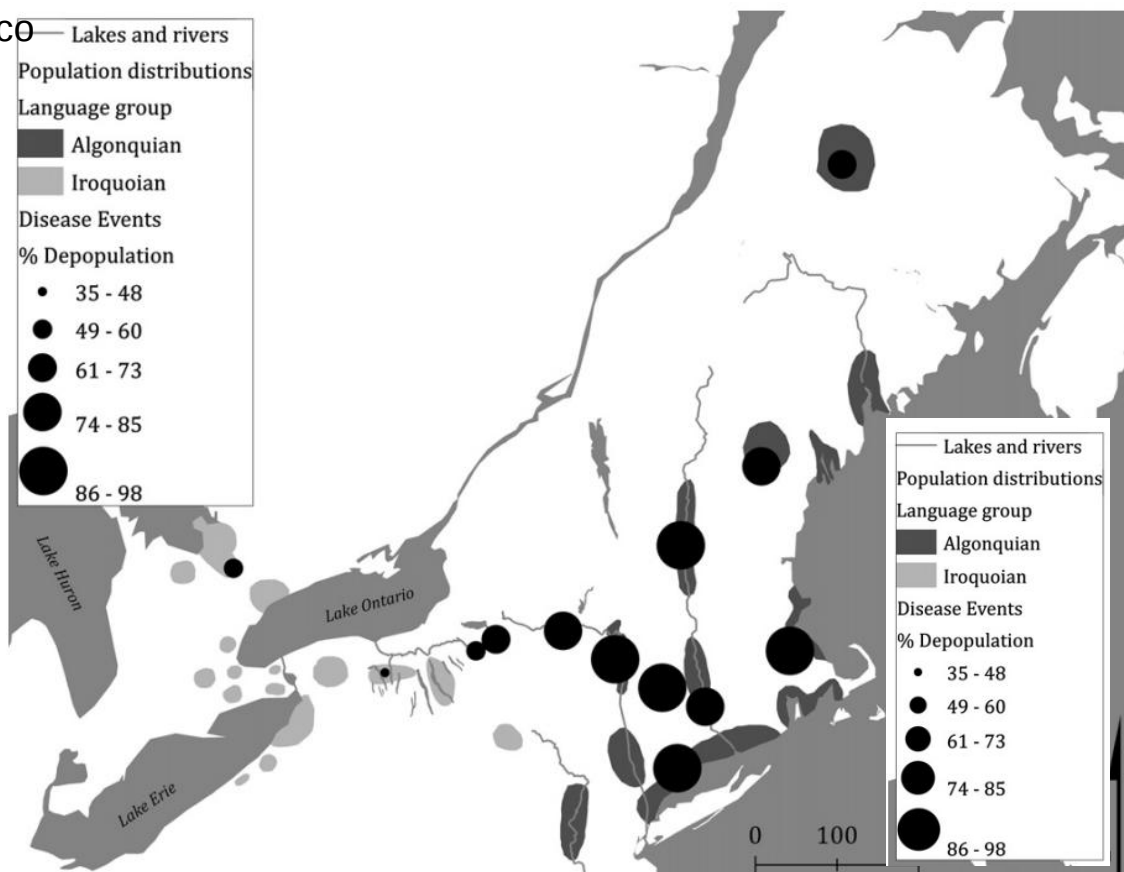
Despovoamento (*depopulation*): % perda entre estimativas pré e pós contato

- Georeferenciou *depopulation* associando ao ponto central da distribuição inicial mapeada
- krigagem da *depopulation* para estimar as taxas de *depopulation* para os grupos sem dados

"It is important to note that kriging can only produce a surface that is an approximation of the spatial variation in the dataset"

- krigagem ordinária, modelos esférico

- Sobreposição do mapa de distribuição geográfica das culturas nativas do sec 17 com a superfície de % de depopulation (buscando entender os padrões para os grupos para os quais não se tinha info)



Spatial analysis to estimate depopulation for Native American populations

Métodos:

Despovoamento (*depopulation*): % perda entre estimativas pré e pós contato

- Georeferenciou *depopulation* associando ao ponto central da distribuição inicial mapeada
- krigagem da *depopulation* para estimar as taxas de *depopulation* para os grupos sem dados

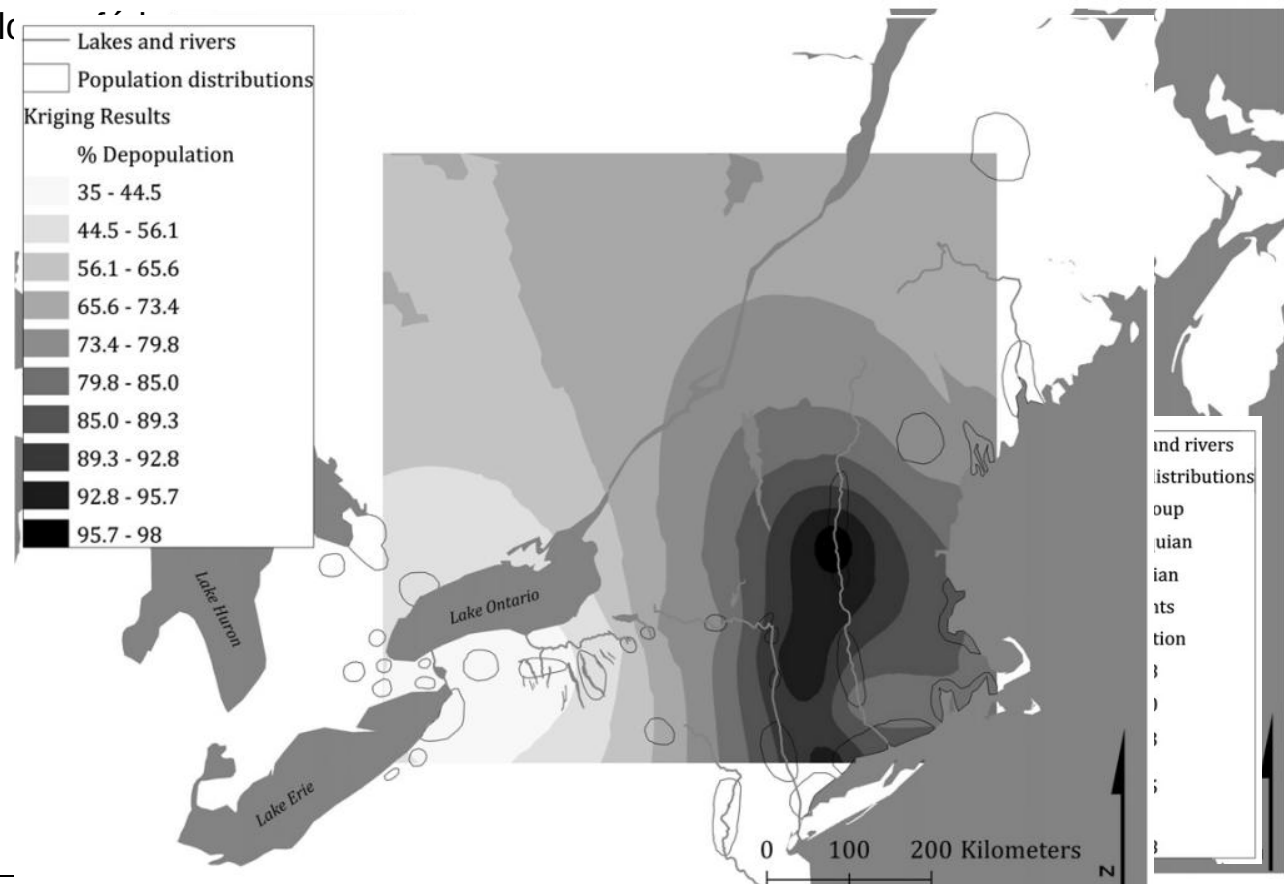
"It is important to note that kriging can only produce a surface that is an approximation of the spatial variation in the dataset"

- krigagem ordinária, modelo

- Sobreposição do mapa de distribuição geográfica das culturas nativas do sec 17 com a superfície de % de *depopulation*

- buscando entender os padrões para os grupos para os quais não se tinha info)

Discussão das causas:
exposição prévia a doenças,
variantes genéticas, práticas
culturais (adoção),
especulações...



- Exemplos de uso de krigagem na literatura sobre representação da distribuição de população

E. Yoo, P.C. Kyriakidis, W. Tobler

Reconstructing Population Density Surfaces from
Areal Data: A Comparison of Tobler's
Pycnophylactic Interpolation Method and Area-to-
Point Kriging

Geographical Analysis 42 (2010) 78–98 r 2010

Krig x Pycnophylatic



Estabelecer relações entre a variografia logarítmica utilizada em geoestatística e o critério de suavização do método de Tobler.

Construir uma superfície:

- A regularidade da superfície desconhecida é governada por uma função determinística;
- Equação que soluciona a superfícies é governada pela diferencial parcial de Laplace ($\sim$ Tobler)
- Apresenta uma solução para aproximar a superfície contínua a partir dos dados de área baseado na premissa que a regularidade (smoothness) da superfície adjacente é conhecida, e não necessariamente é a descrita pela equação diferencial de Laplace.

Geographical Analysis 42 (2010) 78–98 r 2010

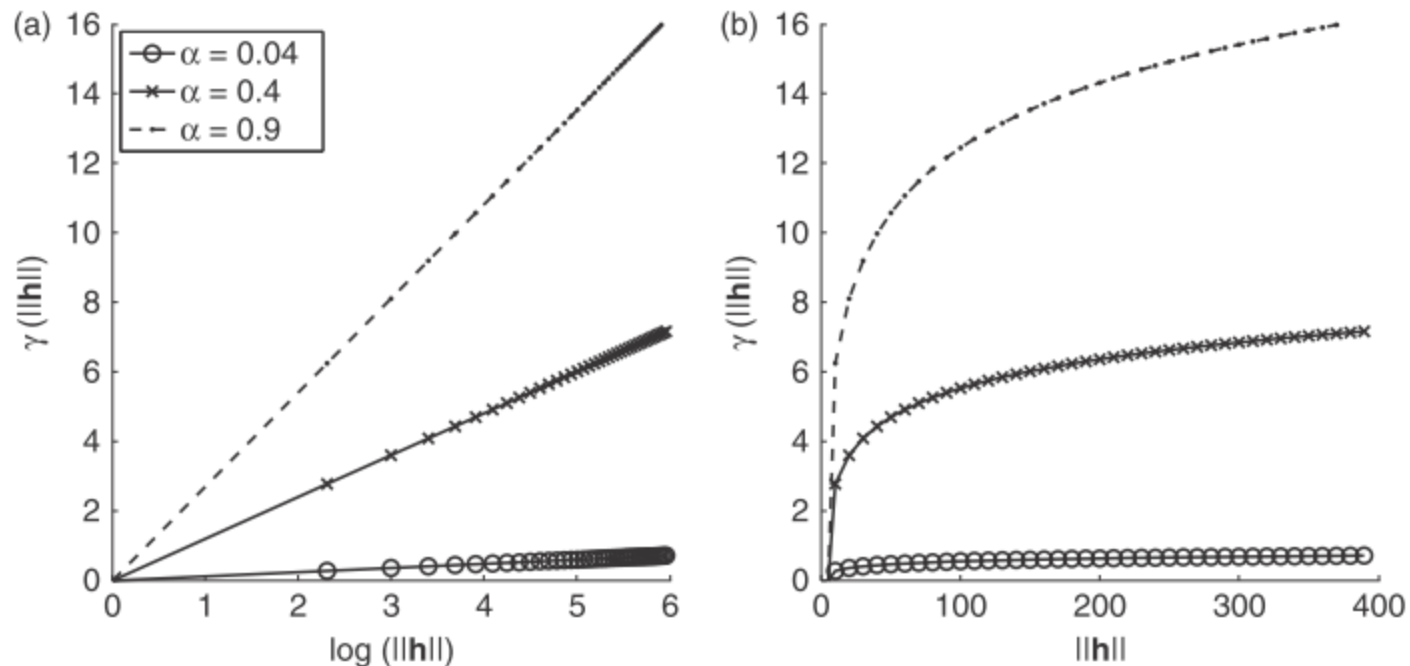


Figure 1. Illustration of logarithmic or de Wijsian variograms with different *absolute dispersion* parameters. The solid lines with circles and x-marks denote, respectively, the de Wijsian model with absolute dispersion parameters 0.04 and 0.4; the dashed-dotted line denotes the variogram model with $\alpha = 0.9$. (a) Plots of de Wijsian variograms versus distance on log scale. (b) Plots of de Wijsian variograms versus distance.

Krig x Pycnophylatic

Superfície de densidade de população a partir de dados coletados em áreas

18 census tracts, Ann Arbor 1970

Eun-Hye Yoo et al.

Reconstructing Population Density Surfaces from Areal Data

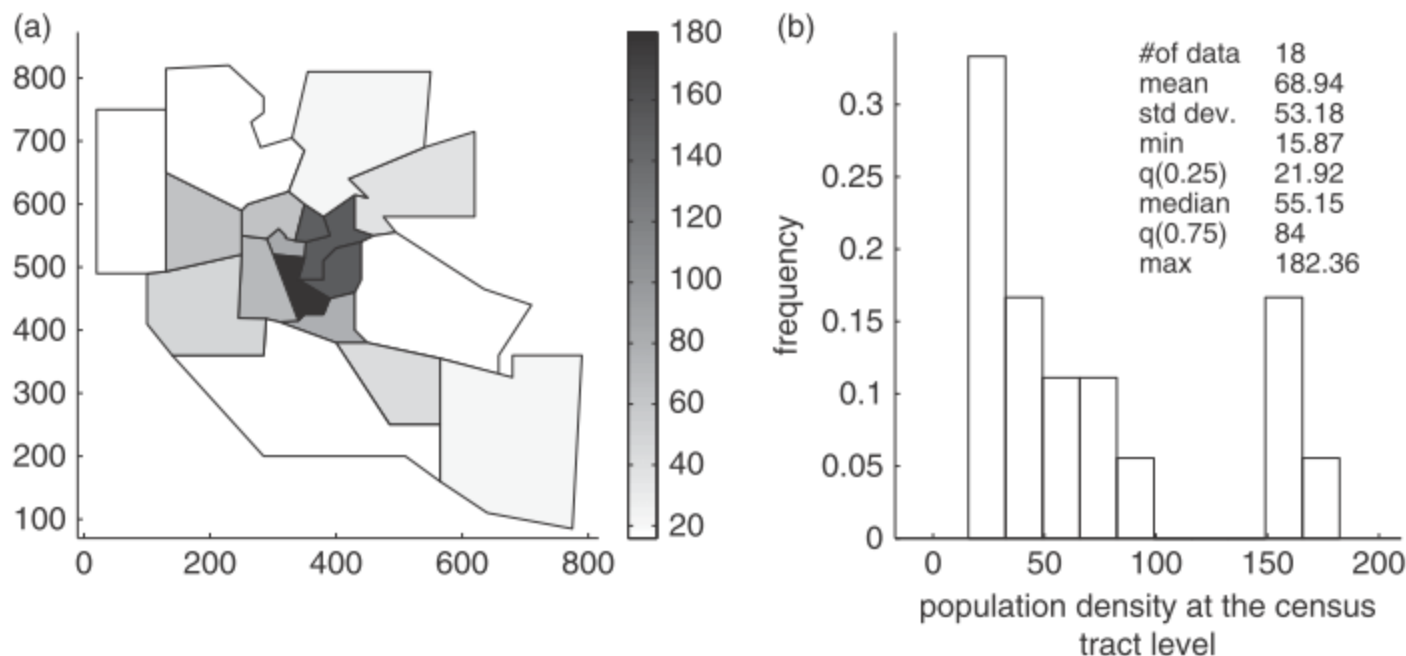


Figure 2. (a) Choropleth map representation of population density at the census tracts in Ann Arbor (1970). (b) Histogram and summary statistics of the mean population density values at census tracts.

Krig x Pycnophylatic

Comparação Tobler x ATP Kriging

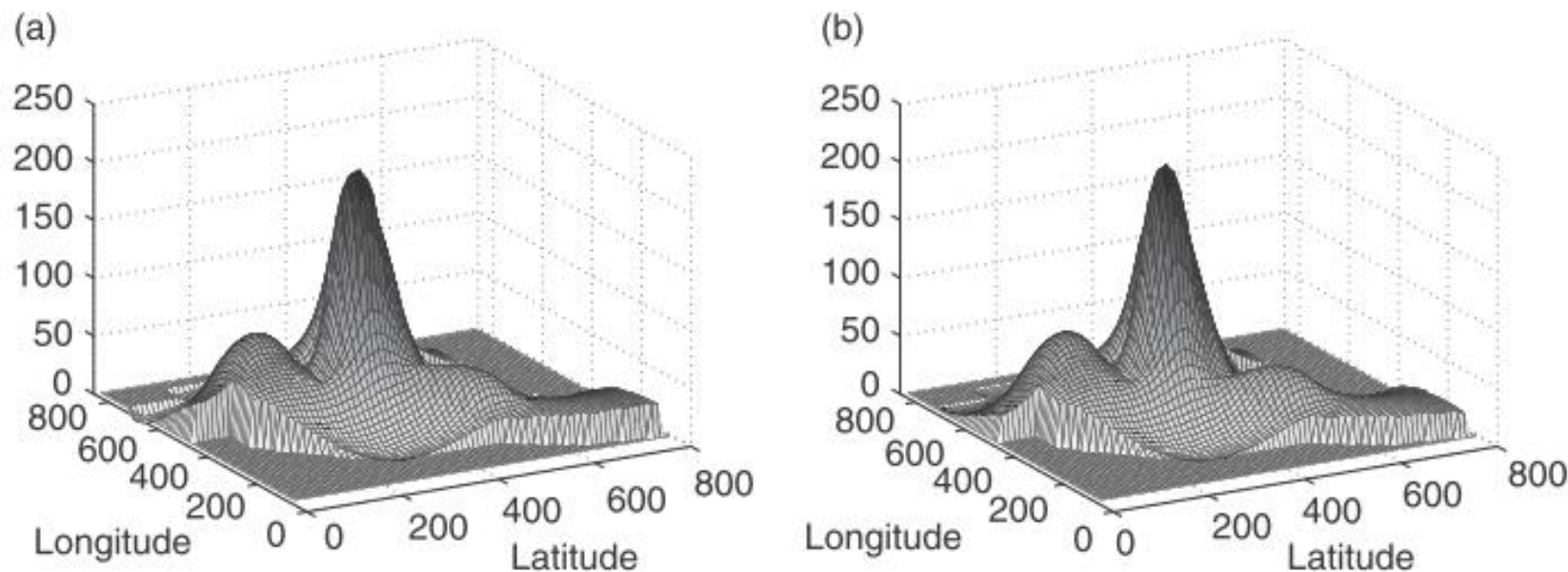
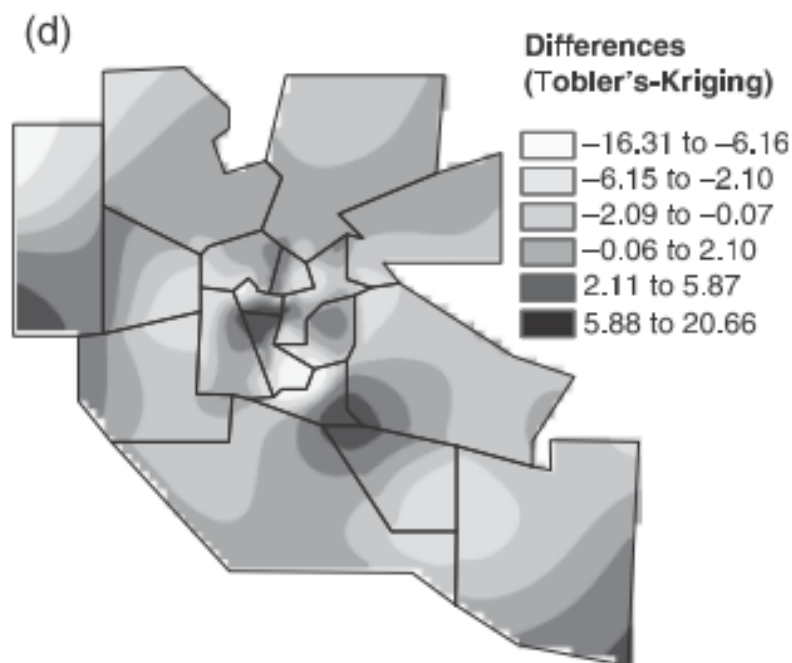
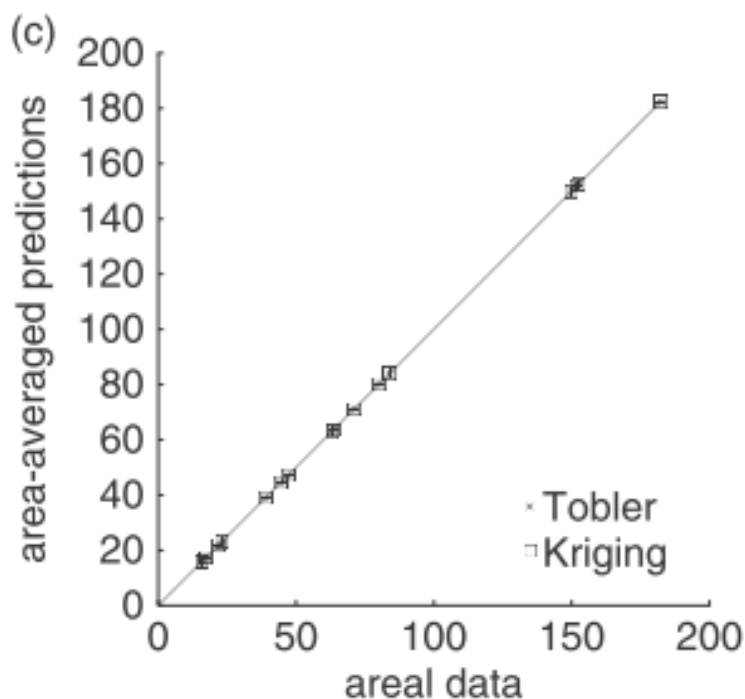


Figure 3. (a) Tobler's solution surface under a Neumann-type boundary condition.
(b) Area-to-point (ATP) kriging prediction surface with boundary conditions at infinity.

Krig x Pycnophylatic

Comparação Tobler x ATP Kriging



(c) Scatter-plot of original areal data (population density at the census tract level of Fig. 2a) versus area-averaged point predictions derived by Tobler's method (asterisks) and by kriging (squares).

(d) Map of differences between point predictions obtained by Tobler's interpolation method (a) and ATP kriging (b).

Comparação Tobler x ATP Kriging

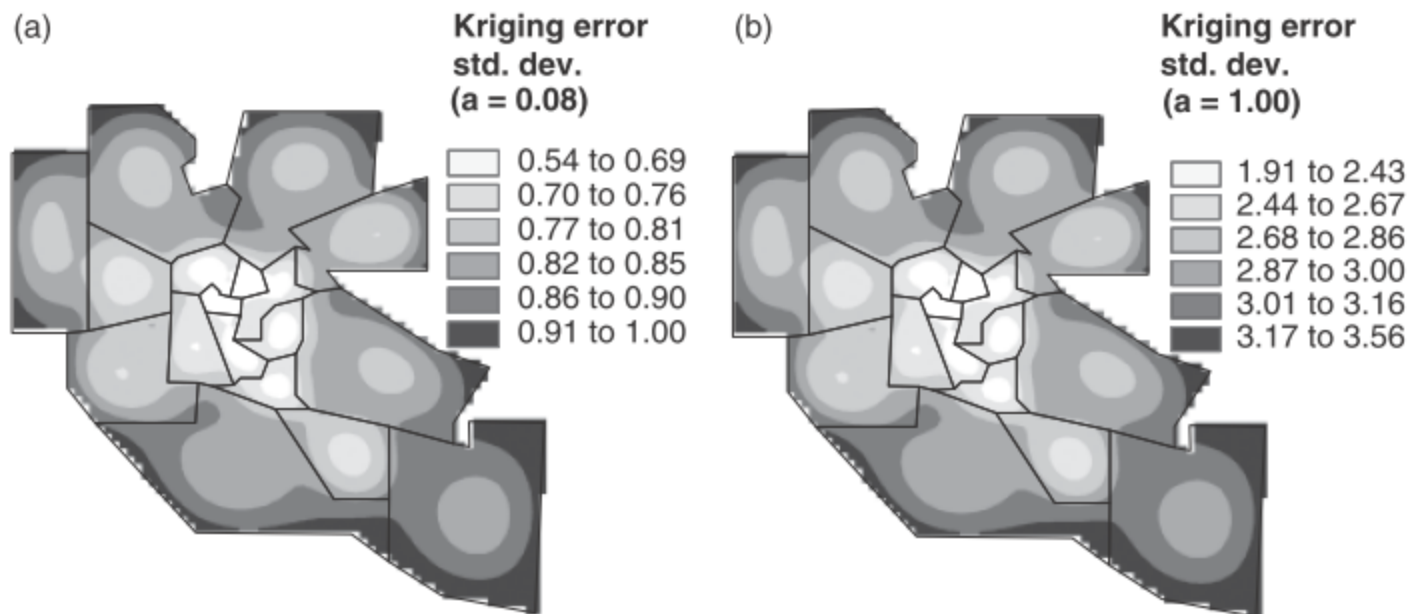


Figure 4. Error standard deviations associated with the area-to-point kriging predictions of Fig. 3b with absolute dispersion $\alpha \approx 0.08$ (a) and $\alpha \approx 1$ (b).

Resultados

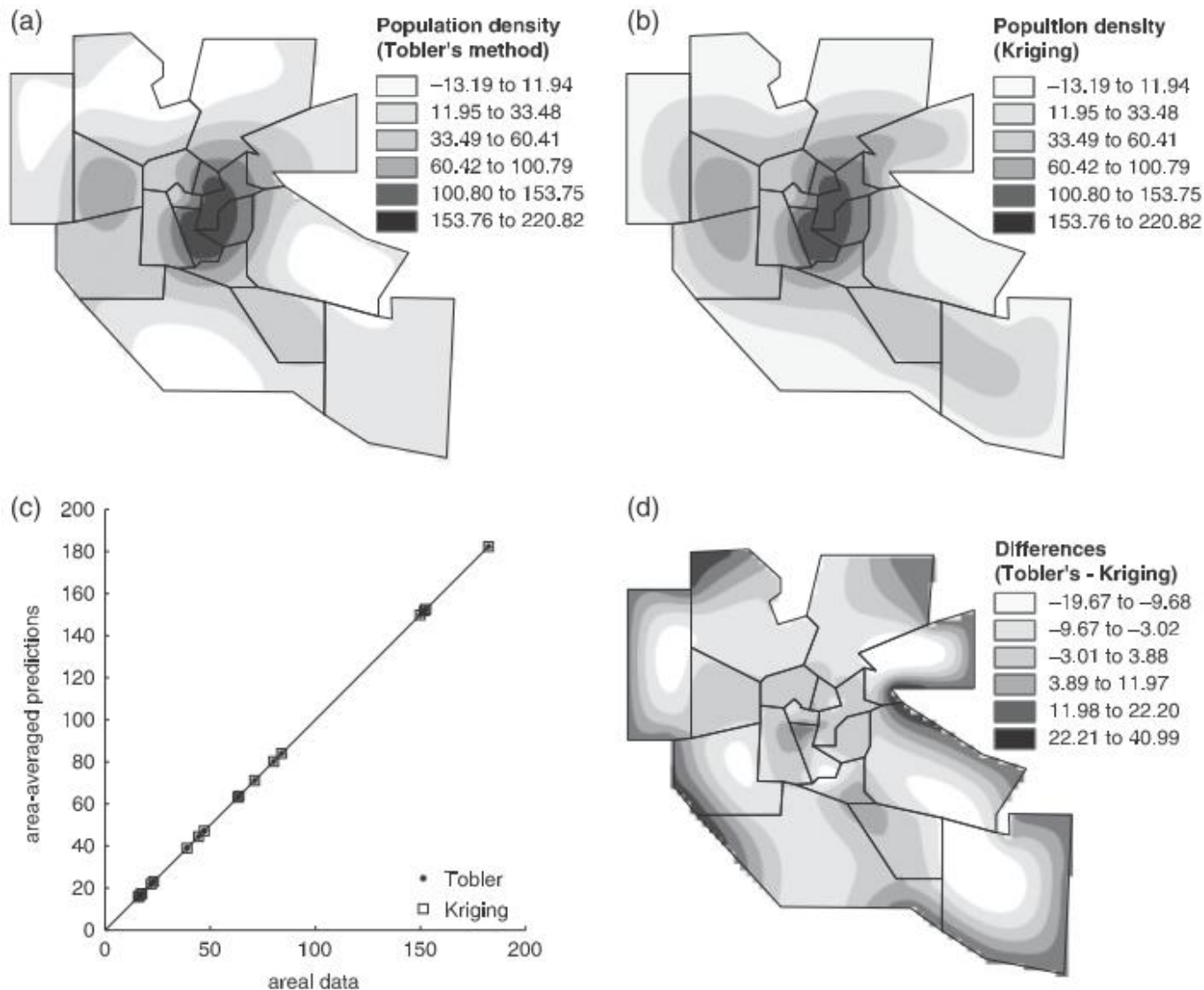


Figure 5. Population density surface derived from Tobler's interpolation method (a) and from area-to-point (ATP) cokriging (b), both under a Dirichlet-type boundary condition. (c) Scatter plot of original areal data (of Fig. 2a) versus area-averaged point predictions derived by Tobler's method (asterisks) and by kriging (squares). (d) Map of differences between point predictions obtained by Tobler's interpolation method (a) and ATP cokriging (b).

Comparação Tobler x ATP Kriging

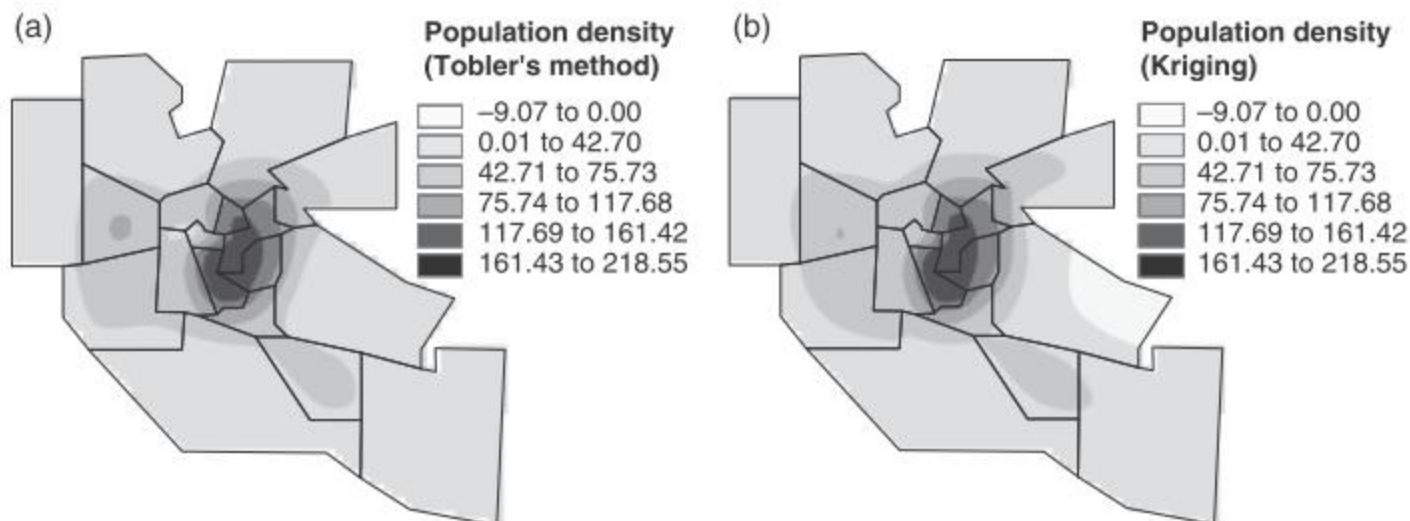


Figure 7. Population density surface predicted from Tobler's interpolation method (a) and from areal-to-point cokriging (b), both subject to a Dirichlet-type boundary condition and non-negativity constraint.

Resultados

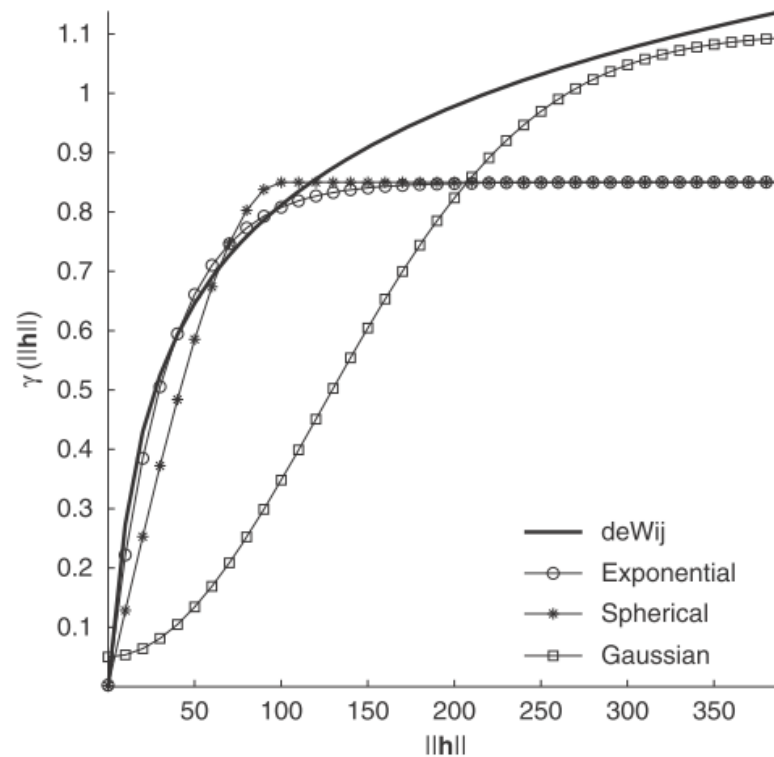


Figure 8. Alternative point variogram models used for area-to-point cokriging: de Wijsian model (solid line), exponential model with effective range 100 and sill 0.85 (solid line with circles), spherical model (solid line with stars) with same parameters as the exponential, and Gaussian model with effective range 300 and sill 1.05 plus a very small nugget effect (solid line with squares).

Table 1 Summary Statistics Comparing Tobler's Solution Surface with Alternative Kriging-Derived Surfaces Constructed Using Different Point Variogram Models

Statistic/Model	de Wijsian	Exponential	Spherical	Gaussian
Correlation coefficient	0.97	0.96	0.95	0.84
Mean absolute error	7.51	7.75	9.74	18.24
Root mean squared error	9.60	9.97	12.21	23.22

Krig x Pycnophylatic



Resultados:

Geoestatística com Modelo de variograma específico – (Wijsian – correspondente a função “free-space Green” da equação de Laplace) e condições e restrições semelhantes a abordagem de Tobler →

Superfície de densidade de população predita é quase idêntica !!!

“results imply that the ATP ordinary kriging with the de Wijsian model provides the closest approximation to Tobler’s solution surface among other point variogram models”

Krig x Pycnophylatic



“we believe it is this Laplacian smoothness criterion that should be the defining feature of Tobler’s method, not its pycnophylactic property”

IMPLICAÇÕES

(1) Superfícies de atributos múltiplos a partir de dados de área podem ser construídas a partir de geoestatística, dependendo do modelo de variograma adotado:

- precisa ser relacionado à solução de Tobler

(2) É responsabilidade do analista justificar se o critério de Tobler empregado é relevante para sua aplicação

(3) Vantagem da geoestatística: permite reportar as incertezas / confiabilidade sobre a região interpolada, com implicações críticas para propagação de incertezas em operações de análise espacial.

Krig & RS – (Area 2 Point)



X. H. LIU, P. C. KYRIAKIDIS and M. F. GOODCHILD Population-density estimation using regression and area-to-point residual kriging.

International Journal of Geographical Information Science,

Vol. 22, No. 4, April 2008, 431–447

- Descreve método de interpolação de área baseado em krigagem → “*area-to-point residual kriging*”
Valores de Densidade de população residual de um modelo de regressão são desagregados usando krigagem área-para-ponto (Kyriakidis 2004)
- Requer apenas um modelo de semivariograma para os pontos residuais, ao invés de um conjunto de modelos de auto- e cross - semivariogramas que envolveria variáveis dependentes e todas as covariáveis.
- **Objetivo: desagregar a população de unidades de censo para zonas de uso da terra, estimando as densidades populacionais equivalentes**

$$P_i = \sum_j d_j A_{ij}$$

P_i = Pop do censo na unidade i ,

d_j = densidade de população da classe de uso da terra j ;

A_{ij} = área da classe de uso j , na unidade do censo i .

(pode ser obtido de diferentes métodos como amostragem empírica)

1. Objetivo: desagregar a população de unidades de censo para zonas de uso da terra, estimando as densidades populacionais equivalentes

Classes de uso da terra

Residencial: (Regressão 1)

- high-density single-unit residential (HSU),
- medium-density single-unit residential (MSU),
- low-density single-unit residential (LSU),
- multiple-unit residential (MU),
- mobile homes (mobile),
- mixed single-and-multiple unit residential (S&M),
- and mixed residential-commercial (mixed)

$$P_i = \sum_j d_j A_{ij}$$



Figure 1. Example of land-use regions in the IKONOS image. Note that the very high spatial resolution of IKONOS enables differentiation in the residential area.

$$P_i = 1024.48A_{\text{LSU}} + 3325.16A_{\text{MSU}} + 5973.49A_{\text{HSU}} + 10180.09A_{\text{MU}} + 5063.15A_{\text{mobile}} + 8826.56A_{\text{S\&M}} + 7593.45A_{\text{mixed}} \quad R^2 = 0.44 \quad (2)$$

1. Objetivo: desagregar a população de unidades de censo para zonas de uso da terra, estimando as densidades populacionais equivalentes

(Regressão 2)

Regressão a partir de métricas de paisagem....

$$\ln(\hat{d}) = 8.819 + 1.772p_1 - 2.612p_2 + 0.0632p_3 \quad R^2 = 0.55 \quad (3)$$

Onde

$\hat{d}$ é a densidade de população estimada para a zona de land-use

p_1 é a proporção de área construída

p_2 é a proporção de vegetação

p_3 é a densidade de manchas (patches) de área construída

- métrica de paisagem computada pelas manchas formadas por pixels contíguos: o n de patches é dividido pela área total da zona de land-use

Equações Regressão (2) and (3) by Liu and Herold (2007).

2. Krigagem de resíduos área-para-ponto (co-kriging variant)

- Suponha que a zona de uso da terra está referenciada por seu centroide **u**.
- Sua densidade de população seria escrita por:
 - **d(u)** é a densidade de pop não conhecida para a zona **u**,
$$d(u) = \hat{d}(u) + r(u)$$
 - $\hat{d}(u) = m(u)$ é o valor estimado pelo modelo de regressão na equação de regressão (2) (3)
 - **r(u)** é o resíduo da densidade de população .

A tarefa é interpolar $r(u)$ de modo que $d(u)$ possa ser estimado...

Area-to-point residual kriging foi aplicado para estimar **r(u)**

Os semivariogramas dos resíduos foram computados para as zonas de land-use cujas pop são conhecidas

A densidade de população dos resíduos destas zonas foram calculadas.

3. Investigaram duas abordagens (e.g. Bracken and Martin 1989, Mennis 2003)

3.1. **population-weighted centroid** para representar a zona de land-use

- Como a densidade populacional de uma zona é assumida como uniforme, o centroide ponderado pela população é equivalente ao centroide geométrico.
- Area-to-point kriging foi aplicada para cada centroide, o valor residual obtido foi considerado representativo da zona como um todo.

3.2. Rasterizou as zonas de land-use para uma **grade** e tratou cada célula como um ponto

- the interpolated values of the cells were aggregated to the land-use zone level by averaging the estimates of the cells in the same land-use zone.

Area-to-point residual kriging

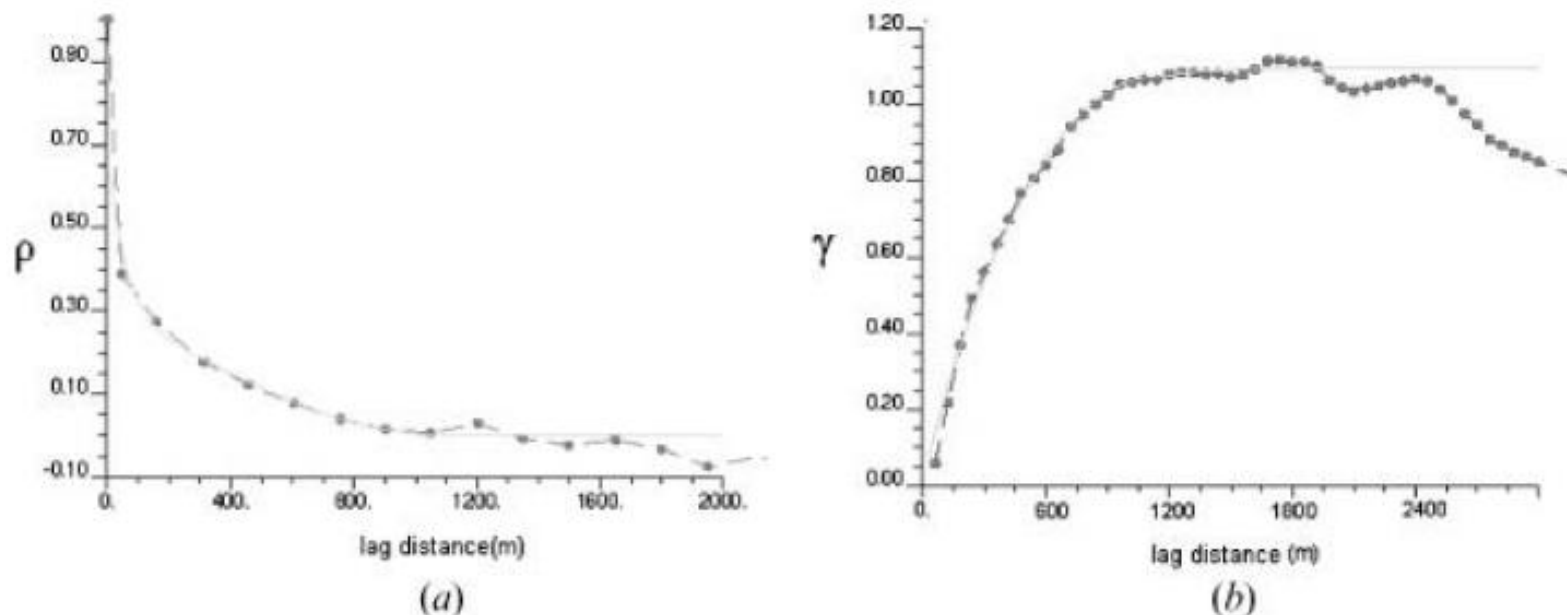


Figure 2. Modelling the spatial autocorrelation of the residual population density: (a) correlogram estimated from the centroid approach; (b) semivariogram estimated from the grid approach using 30-m resolution.

Table 1. Theoretical models for the centroid-based correlogram and grid-based semivariogram.

Centroid				Grid			
Type	Nugget	Range (m)	Sill	Type	Nugget	Range (m)	Sill
S ^a	0.01	0.1	0.30	S ^a	0.01	1080	0.54
S ^a		150	0.40				
E ^a		1100	0.31	E ^a		1080	0.55

^aS: spherical model; E: exponential model.

Então...



- Os semivariogramas dos centróides e grades foram aplicados para estimar a densidade de pop residual para cada zona
- A densidade pop estimada para os resíduos por krigagem foi combinada com a densidade pop estimada pela regressão para obter a população total de cada zona.

(Como densidade de população não pode ser menor que 0, estimativas negativas foram ajustadas para 0)

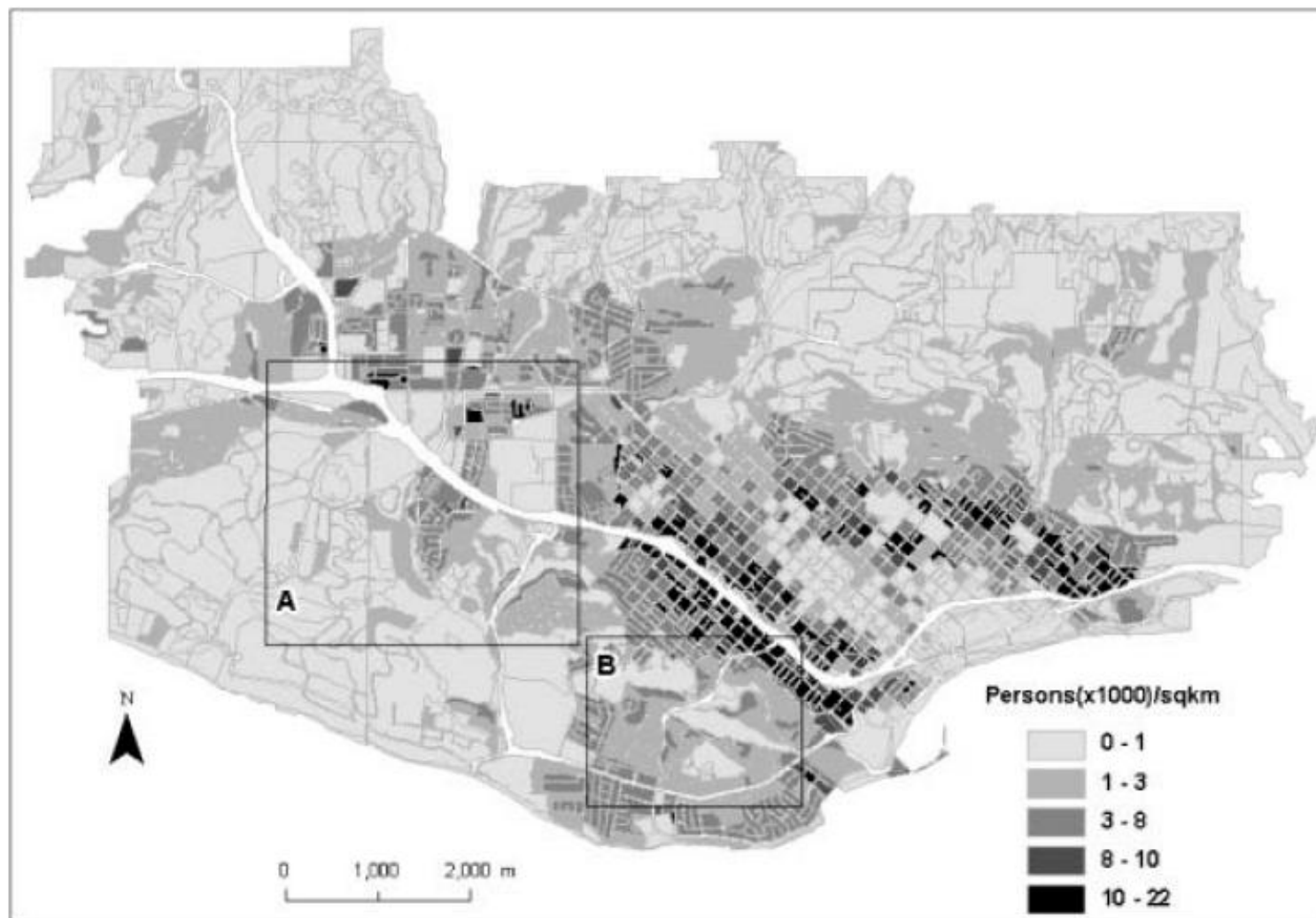


Figure 3. Population density estimated by regression supplemented by area-to-point residual kriging using the grid approach.

A

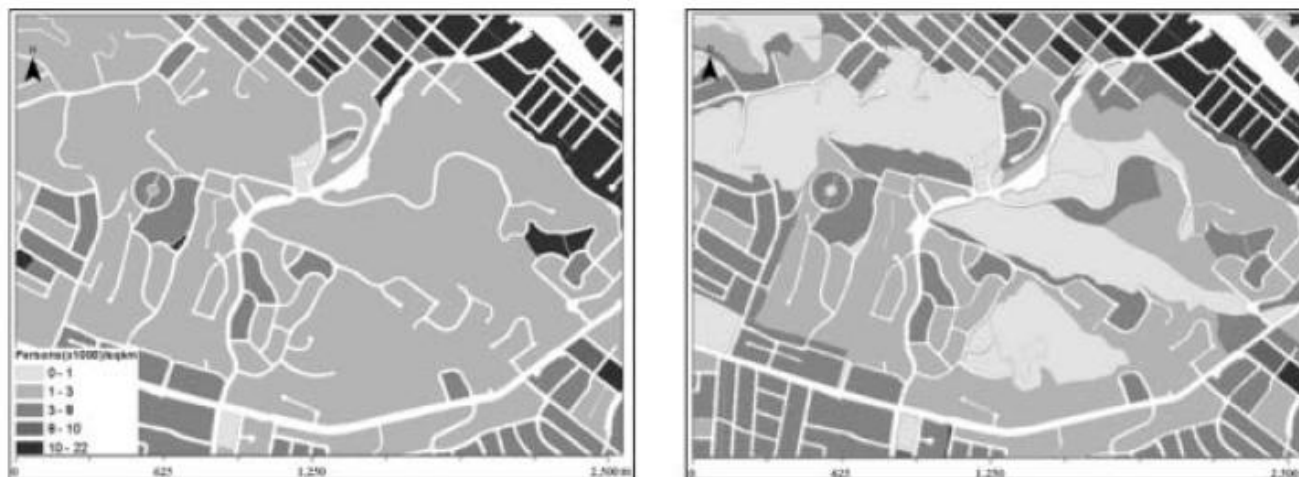


Figure 4. Details of the low-density area, specified as rectangle A in figure 3, showing the difference between the population density reported by the census (left) and that estimated by areal interpolation (right).

B



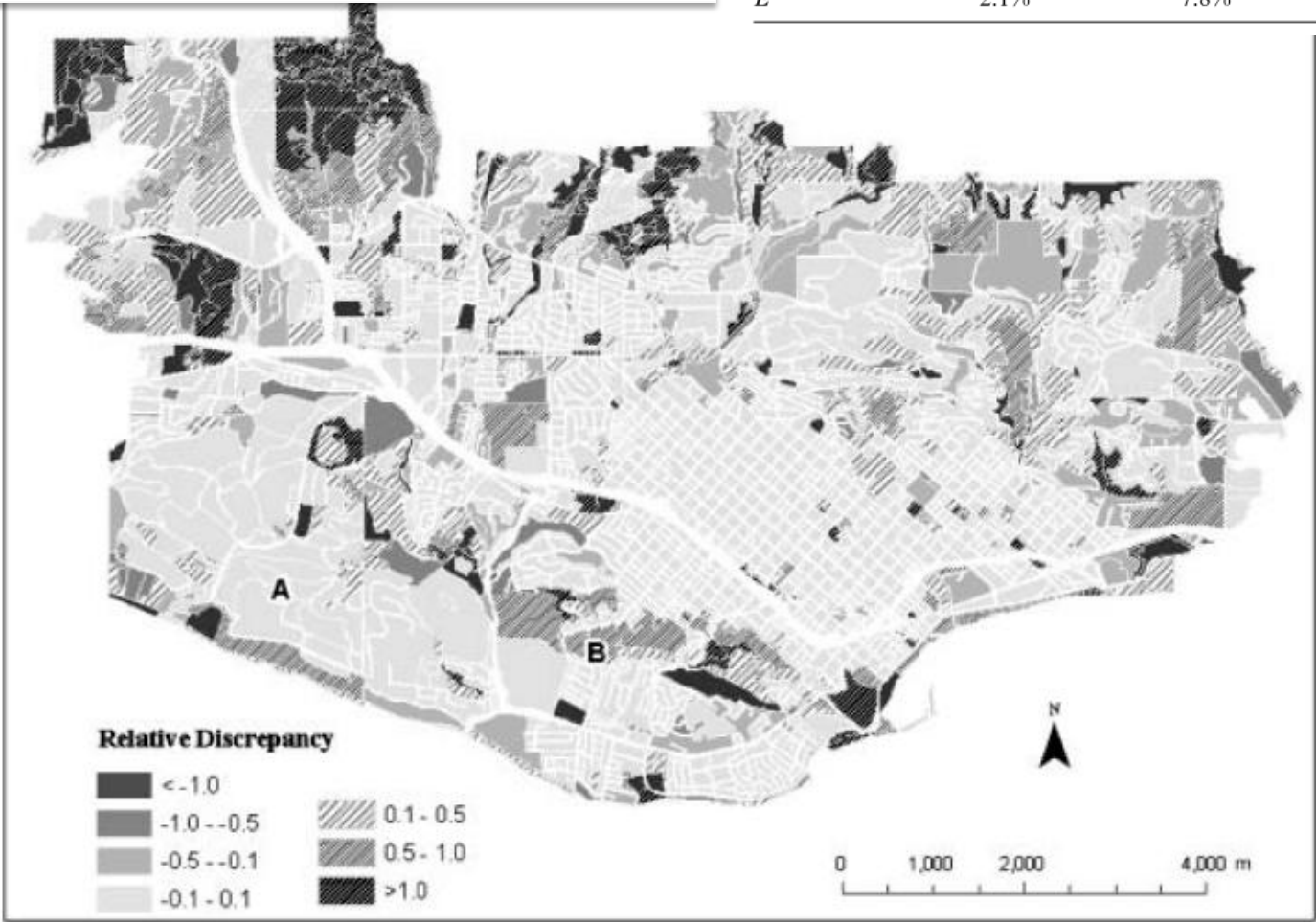
Figure 5. Details of a study area, specified in figure 3 as rectangle B, showing the difference between the population density reported by the census (left) and that estimated by areal interpolation (right).

Table 2. Accuracy assessment of regression-based predictions with and without the supplement of area-to-point kriging.

Accuracy assessment – Precisão

global da estimativa da densidade populacional foi avaliada pela agregação da população das zonas de landuse para cada unidade de censo.

	Regression		Regression supplemented by cokriging	
	Land use (equation (2))	Spatial metrics (equation (3))	Centroid	Grid
ρ	0.70	0.72	0.98	0.98
RMSE	104.8	107.6	21.1	22.4
MAE	48.5%	45.6%	2.9%	4.3%
E	2.1%	-7.8%	-2.6%	2.2%



“ In theory, regression supplemented by area-to-point residual kriging reproduces the population of the census blocks.

However, this mass-preserving property was lost in this study because the negative estimates from kriging were reset to 0 “

Figure 6. Relative discrepancy between population-density estimates obtained by the centroid and grid approaches.

O método **Area-to-point residual Kriging**

- Usa o fato de que densidade de população dos resíduos de uma zona de origem é uma média ponderada linear de uma densidade de população dos resíduos de zonas de destino.
 - Simplificou o procedimento de modelagem por semivariograma para os resíduos (em pontos)
 - Explorou semivariograma a partir de centroide e células da grade, e recomenda o procedimento por células (grid approach)
- Na real: area-to-point residual kriging pode ser usado para complementar qualquer outro método desde que a densidade de resíduos encontrada seja espacialmente variável e auto-correlacionada

Básicas – Krigeagem:

- Camargo, E.C.G.,; Fucks, S. ; Camara, G. Análise Espacial de Superfícies <http://www.dpi.inpe.br/gilberto/livro/analise/cap3-superficies.pdf>
- Tobler, W.R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 519-530.
- Tobler, W.R., Deichmann, U., Gottsegen, J.& Maloy, K. (1995). *The Global Demography Project*. Santa Barbara, CA: National Center for Geographic Information and Analysis.
- Wolf-Dieter Rase, 2001, “Volume-preserving interpolation of a smooth surface from polygon-related data”, *J. Geograph. Syst.*, 3: 199-213

Exemplos de aplicação:

- JONES E.E.; DeWITTE, S.N. (2012) Using spatial analysis to estimate depopulation for Native American populations in northeastern North America, AD 1616–1645. *Journal of Anthropological Archaeology*, 31 83–92.
- LIU, X. H.; KYRIAKIDIS, P. C.; GOODCHILD, M. F. Population-density estimation using regression and area-to-point residual kriging (2008) *International Journal of Geographical Information Science*. Vol. 22, No. 4, April 2008, 431–447
- YOO, E.; KYRIAKIDIS, P.C.; TOBLER, W. (2010) Reconstructing Population Density Surfaces from Areal Data: A Comparison of Tobler's Pycnophylactic Interpolation Method and Area-to-Point Kriging. *Geographical Analysis*., 42: 78–98.