

Geoestatística aplicada à AP

José P. Molin
ESALQ/USP
jpmolin@usp.br



www.agriculturadeprecisao.org.br

Objetivo

Abordar os conceitos fundamentais relacionados à geoestatística aplicada à agricultura de precisão visando a modelagem de dependência espacial e interpolações

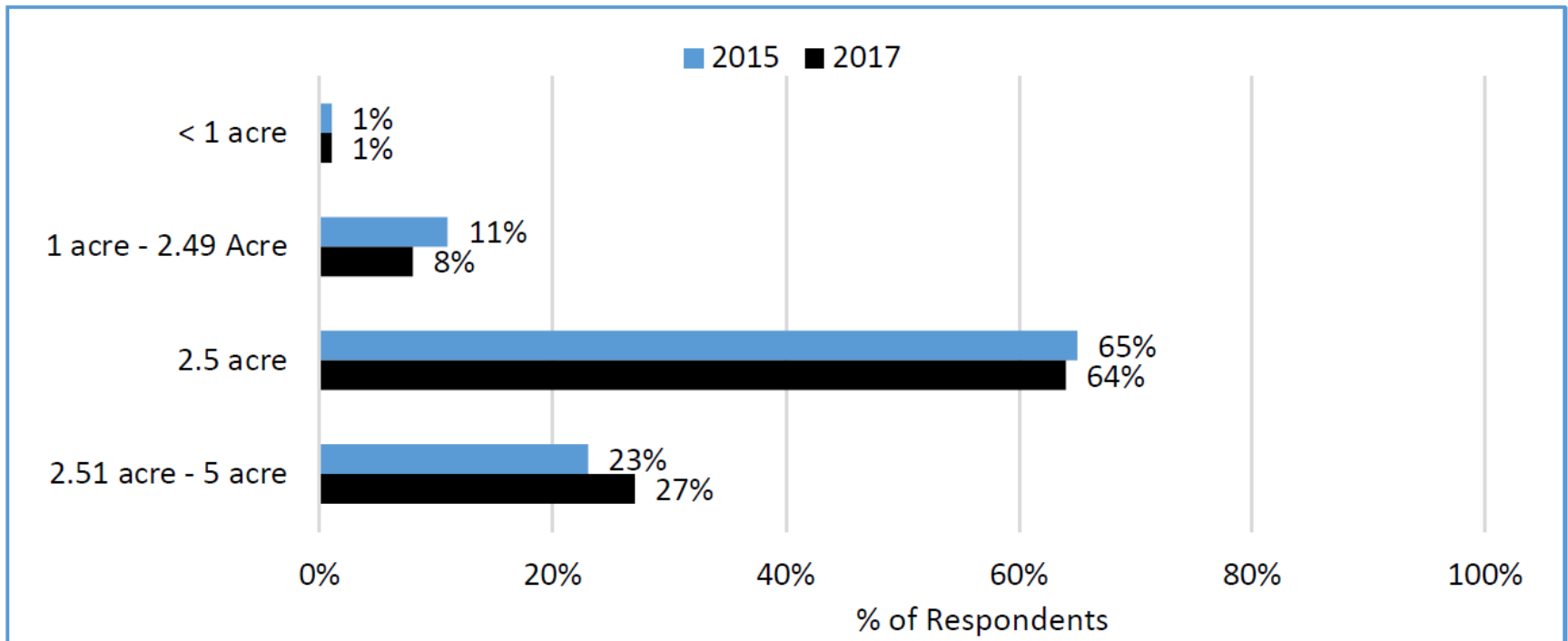
Como se define o “tamanho” mais correto de
uma grade amostral?

ou

Como se produz mapas confiáveis?

A forma mais recomendada
é a partir da **Geoestatística**

Tamanhos de grades amostrais praticadas

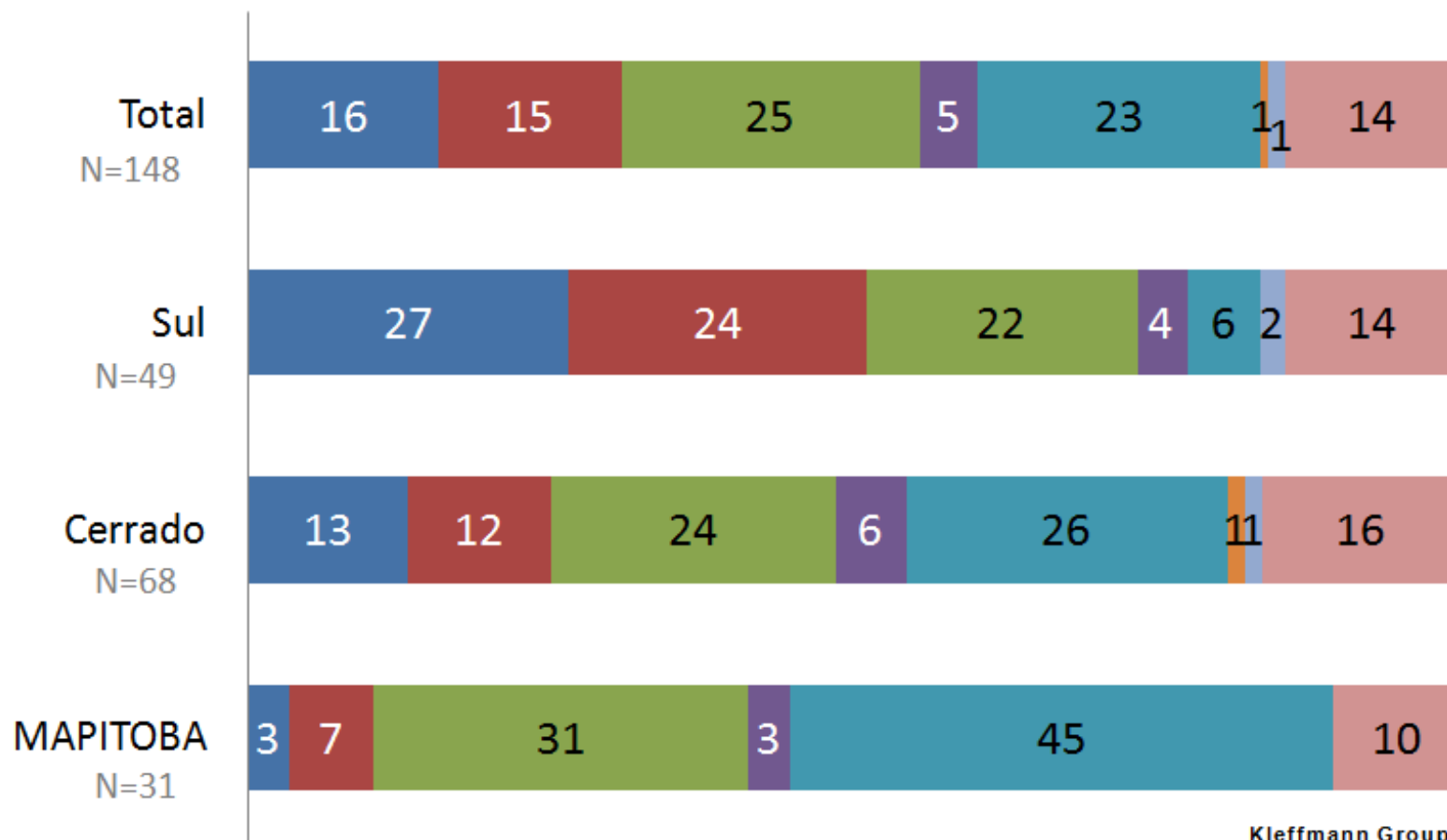


B.Erickson, J. L. Lowenberg-DeBoer & J. Bradford (2017)

Tamanho das grades amostrais, em hectares, praticadas em cada região do Brasil

Todas as indicações em %. Base: Entrevistados que realizam o mapeamento da fertilidade do solo por grades N = 148

■ Até 1 ■ 1,1 - 2 ■ 2,1 - 3 ■ 3,1 - 4 ■ 4,1 - 5 ■ 6,1 - 7 ■ 7,1 - 9 ■ Maior que 9,1



Kleffmann Group ®

J. P. Molin (2017)

A estatística...

- Permite a caracterização de parâmetros da população a partir de parâmetros da amostra
- Pressupõem:
 - independência entre as observações;
 - distribuição normal dos dados;
 - variância e coeficiente de variação constantes.

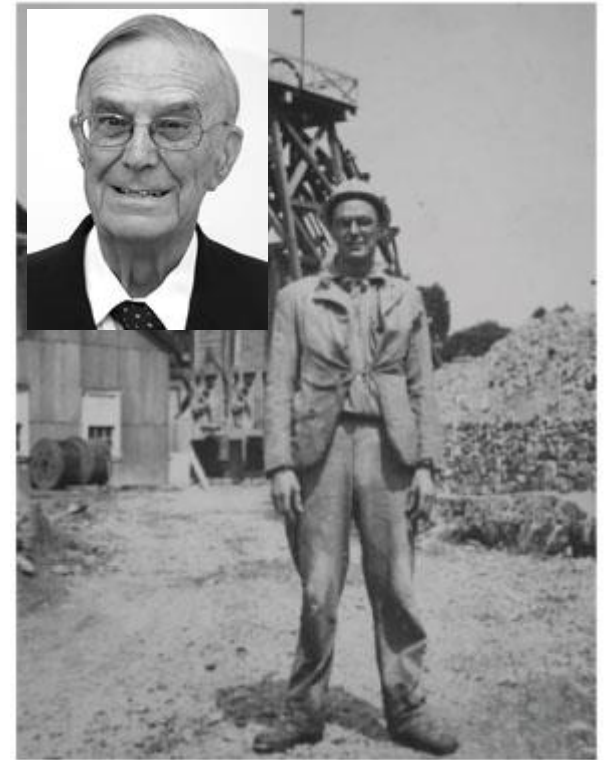
E a geoestatística?

A geoestatística tem por objetivos:

- Identificar, na aparente desordem das amostras, uma **medida da correlação espacial**;
- **modelar e quantificar a dependência espacial**;
- identificar padrões de amostragem adequados;
- permite estimativas de valores em locais não amostrados a partir das amostras existentes (**interpolação por krigagem**).

A origem...

- Daniel G. Krige (1951), na África do Sul, estudou teor de ouro em amostras;
- importância de ligar as variâncias com as distâncias entre as amostras;

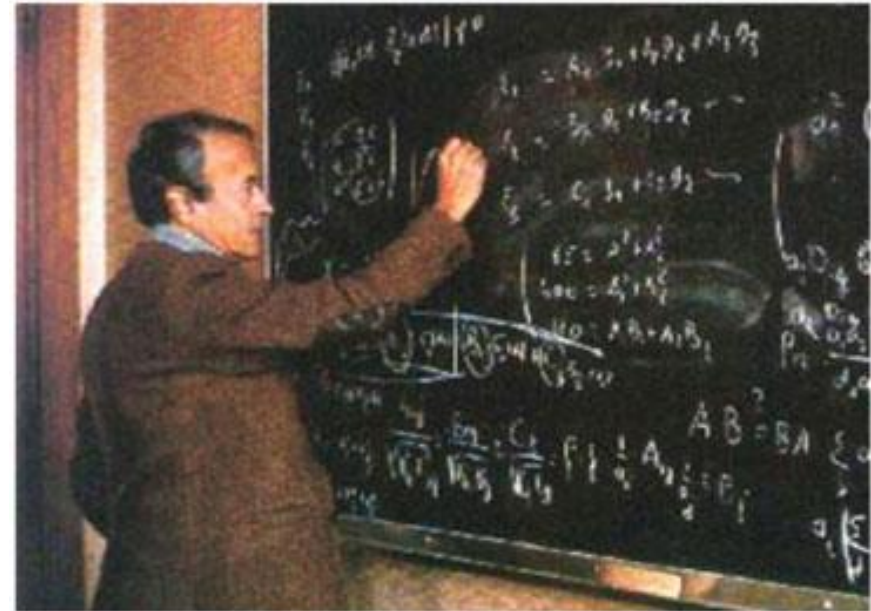


Danie Krige learning the trade, 1939.

Minnitt & Assibey-Bonsu (2014)

A origem...

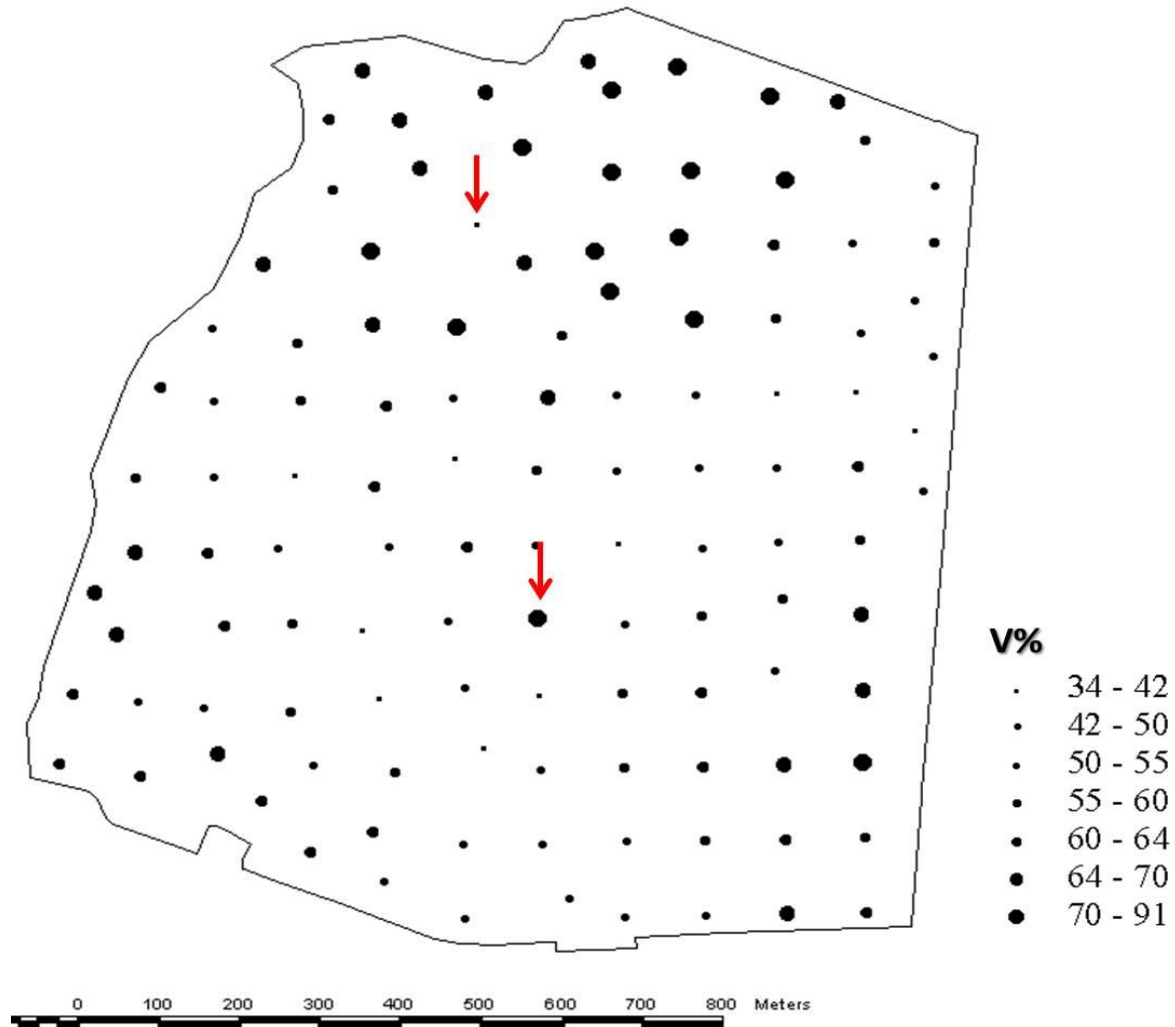
- Matheron (1961), na França, elaborou a Teoria das Variáveis Regionalizadas: “esperança de que, na média, as amostras próximas, no tempo e espaço, sejam mais similares entre si do que as que estiverem mais distantes”



Professor Georges Matheron.

Minnitt & Assibey-Bonsu (2014)

Primeiro passo: análise exploratória dos dados

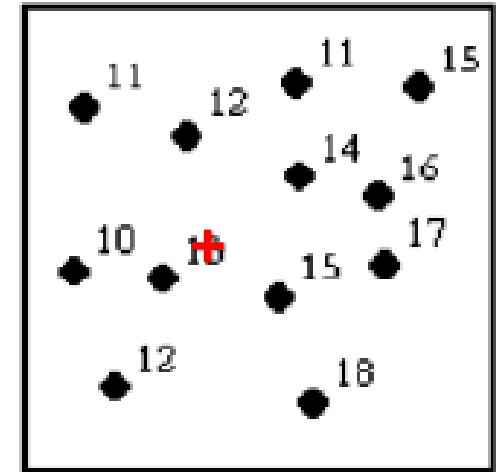


L.R. AMARAL, 2014

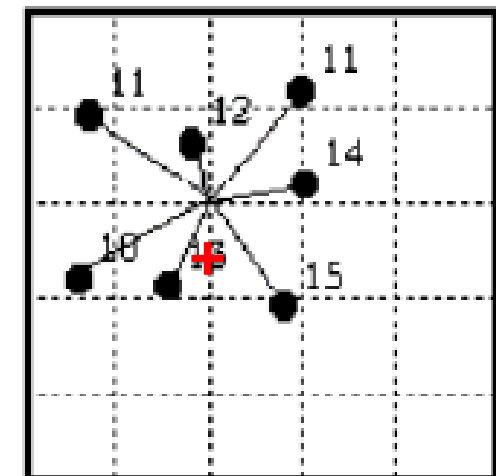
Segundo passo:

interpolação
ou
análise da dependência espacial

Segundo passo:
se interpolação...



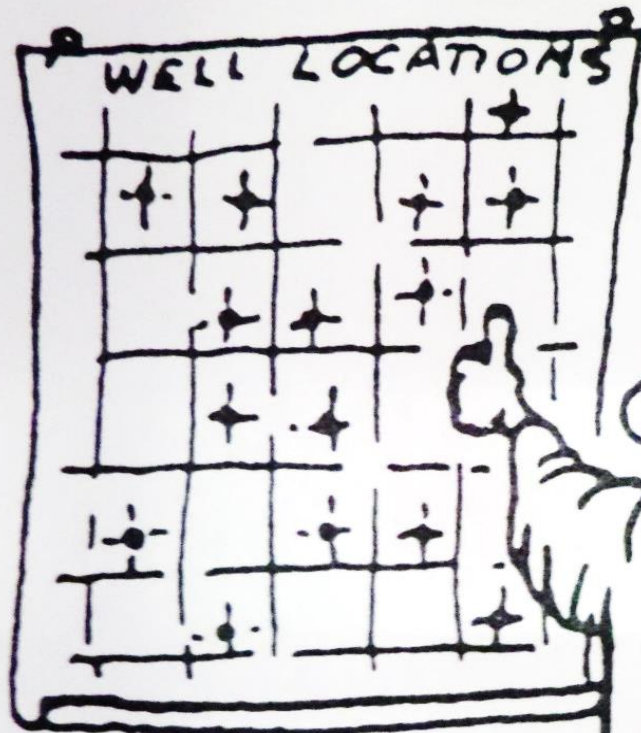
(a)



L.R. AMARAL, 2014

O processo (ou a arte?)
de interpolar...

Structure
Lab

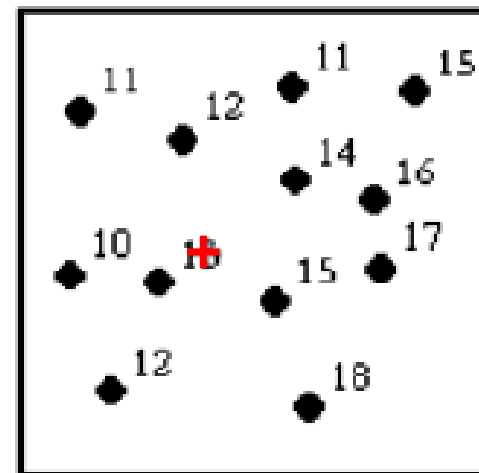


- MTR -

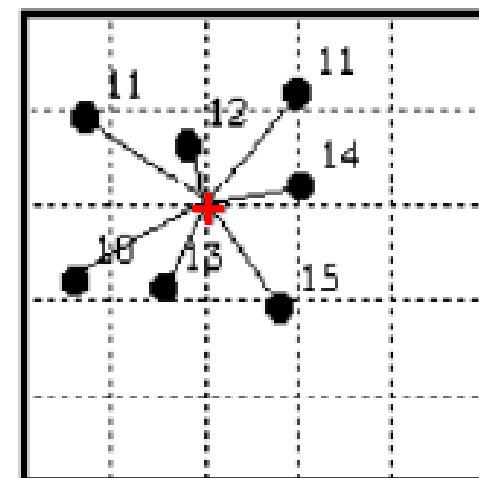
Segundo passo: se interpolação...

Escolha do interpolador matemático:

- vizinho mais próximo
- média simples
- mínima curvatura
- triangulação
- **média ponderada**



(a)



L.R. AMARAL, 2014

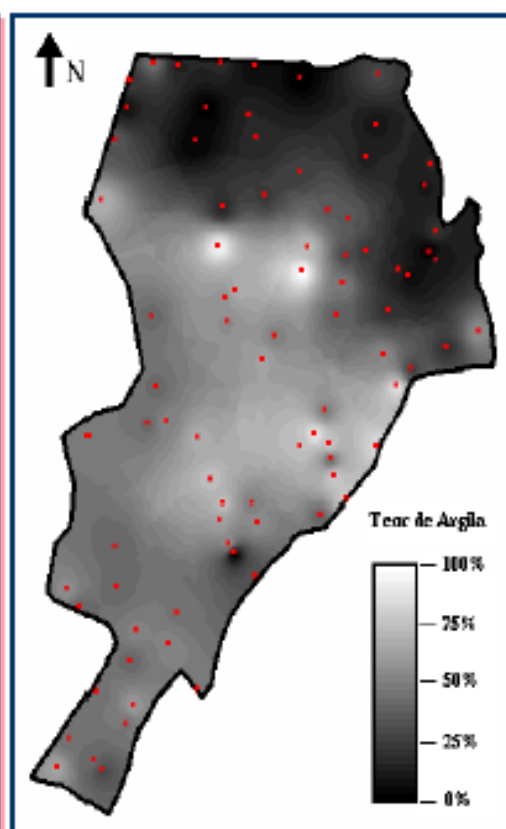
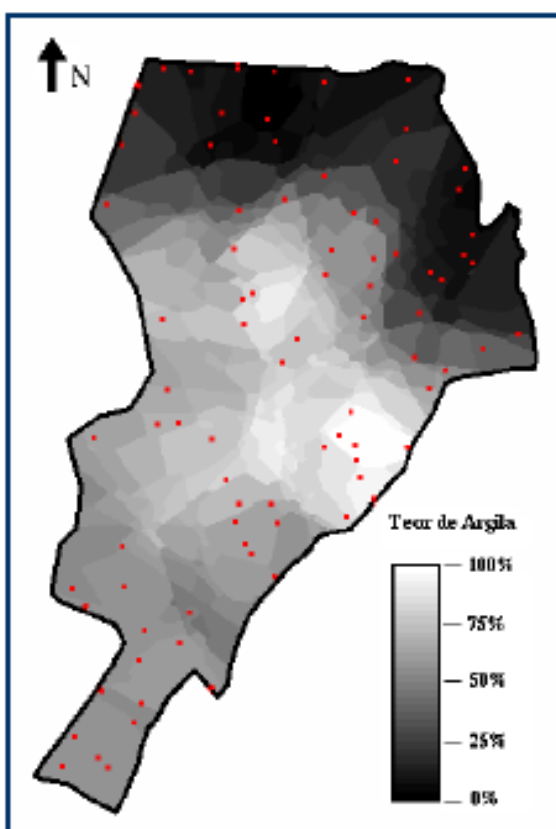
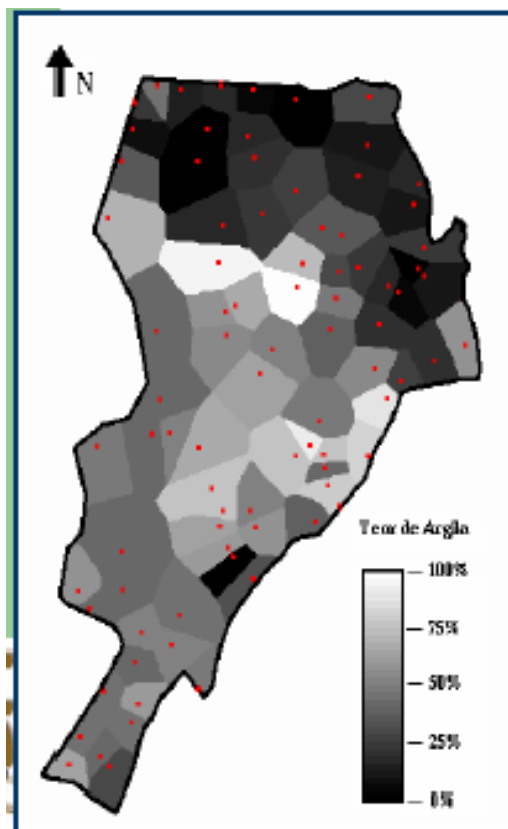
Segundo passo: se interpolação...

→ Inverso da distância ao quadrado

Vizinho mais
próximo

Média Simples

Média Ponderada
(inverso da distância)



L.R. AMARAL, 2014

Segundo passo:

interpolação
ou
análise da dependência espacial

~~Segundo passo:~~

interpolação → 3º passo
ou

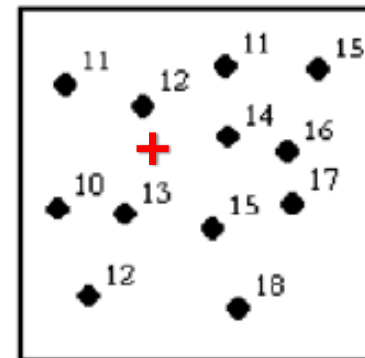
análise da dependência espacial



2º passo



Geoestatística



Geoestatística e dependência espacial

- Valores de uma mesma variável, próximos um ao outro, têm a probabilidade de serem mais semelhantes do que dois valores mais distantes entre si.
- COMO INVESTIGAR ESSA DEPENDÊNCIA ESPACIAL?
- Calculando a variância entre pares de pontos e representando-a em um gráfico como função da distância entre os dois pontos – o **semivariograma**

O que é o semivariograma?

- é um modelo que mede e descreve a dependência espacial entre as amostras.
- é do que um gráfico que representa a semivariância dos dados $\gamma(h)$ em relação à distância que os separa (h).

Como se calcula a semivariância?

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

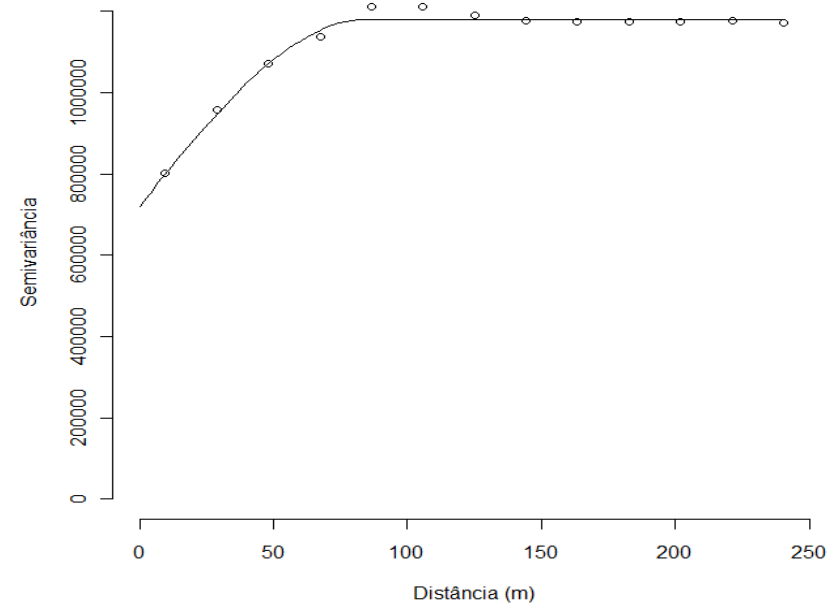
em que:

$\gamma(h)$ é a semivariância

z é a variável em estudo

$z(s+h)$ e $z(s)$ são os valores separados por um vetor h

$N(h)$ é o número de pares de valores $[z(s+h)-z(s)]$ separados por um vetor h



Como se calcula a semivariância?

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

onde:

$\gamma(h)$ é a semivariância;

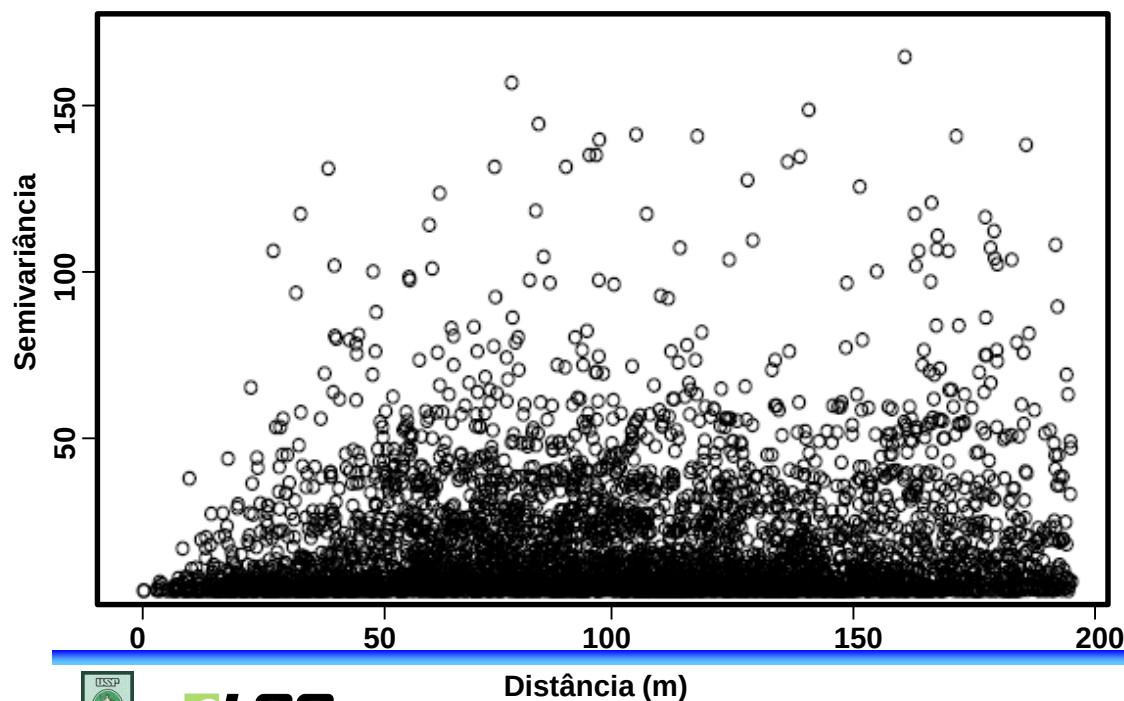
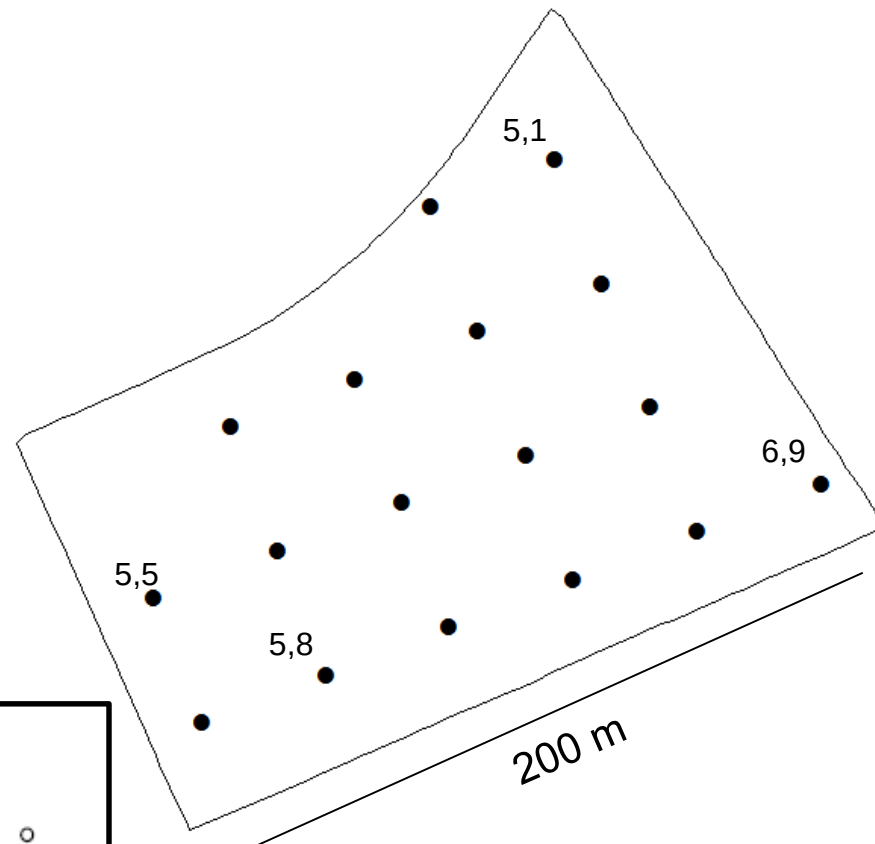
$z(s+h)$ e $z(s)$ são os valores separados por um vetor h ;

$N(h)$ é o número de pares de valores $[z(s+h)-z(s)]$ separados por um vetor h ;

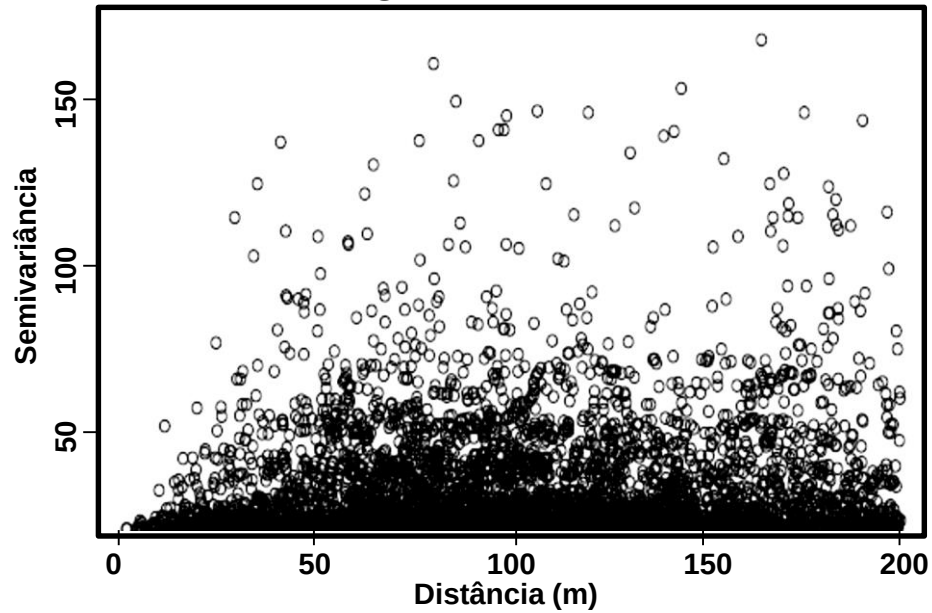
z é a variável em estudo.

Cálculo das semivariâncias:

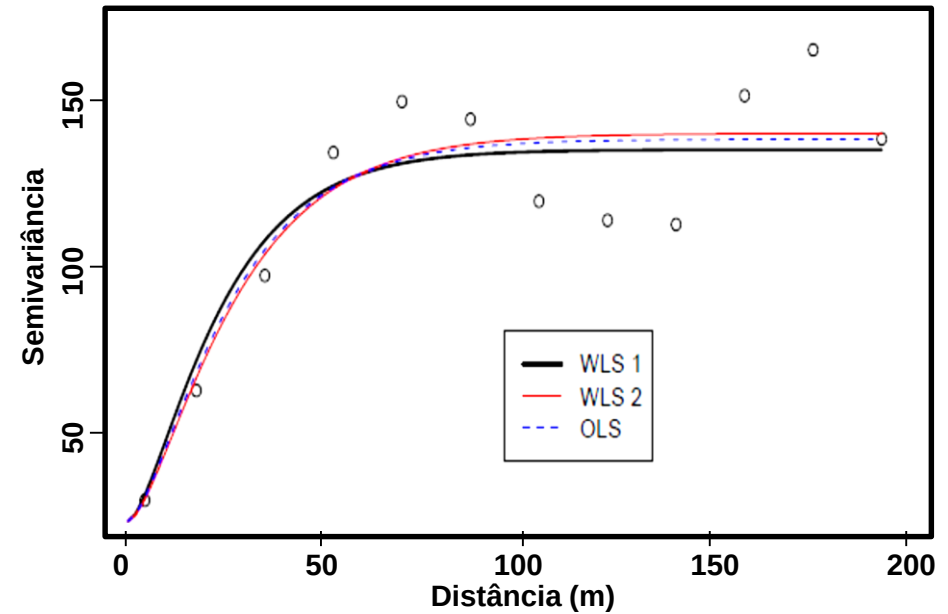
$$\text{Semivariância} = \frac{(p_1 - p_2)^2}{2}$$



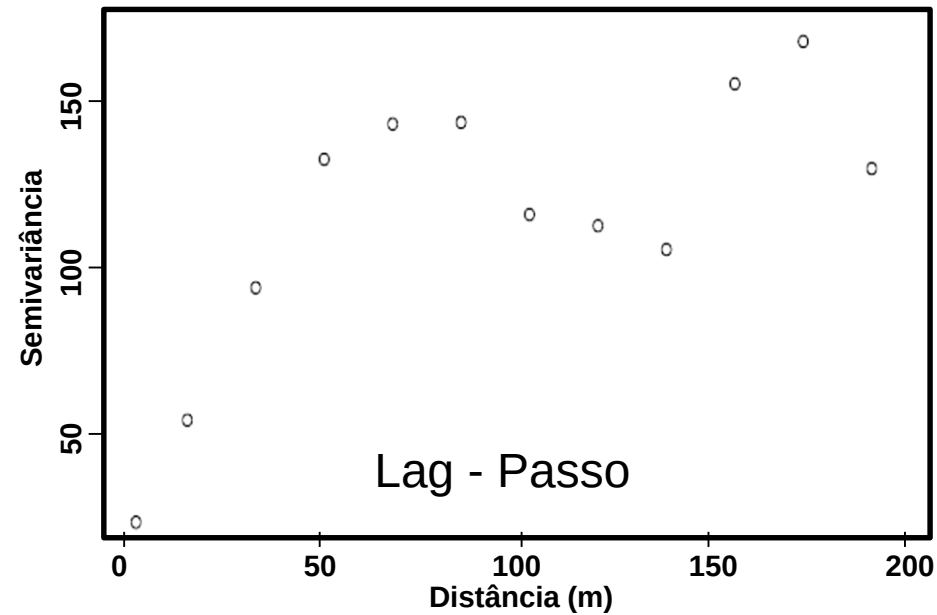
Nuvem variográfica



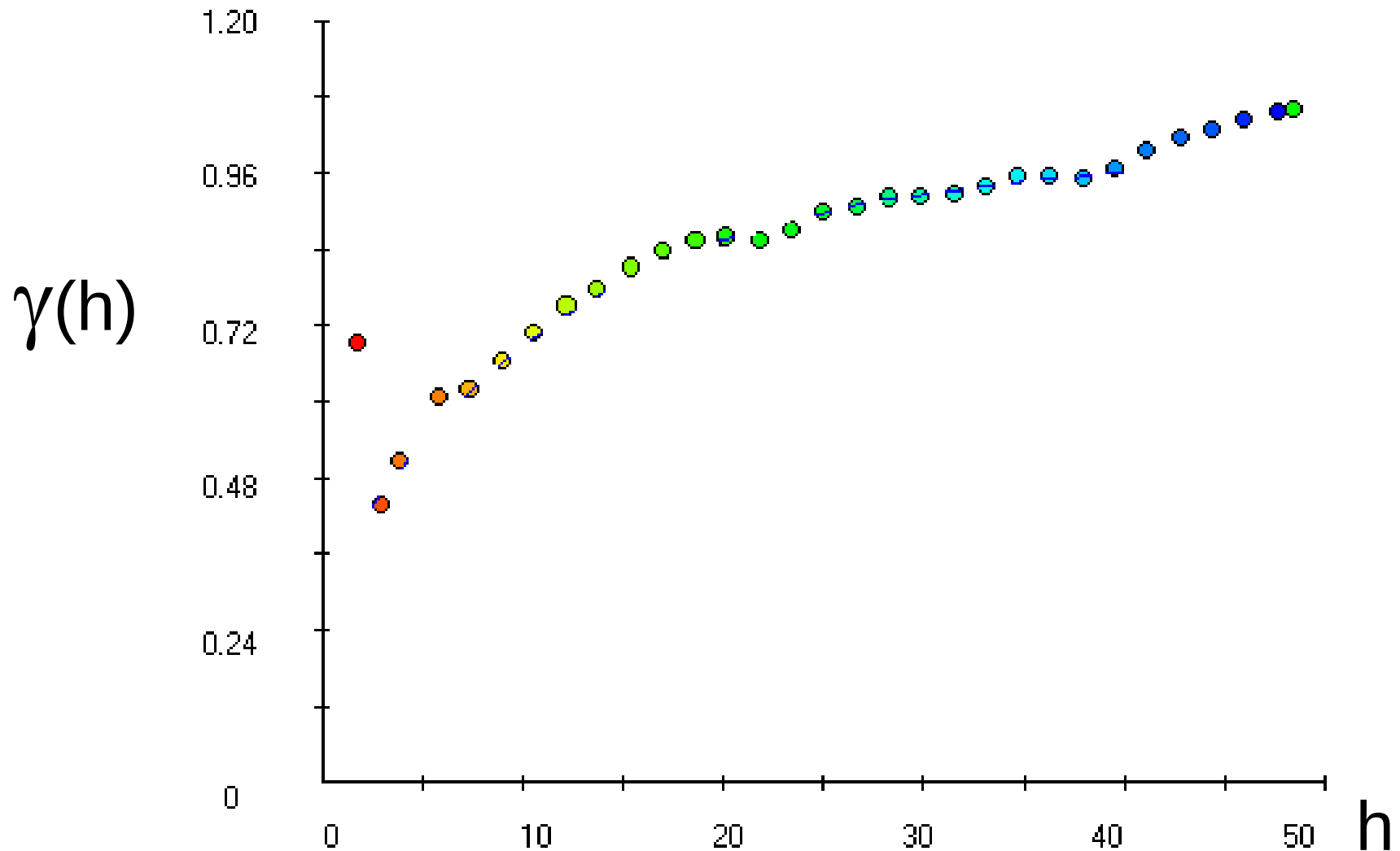
Semivariograma teórico (modelado)



Semivariograma experimental



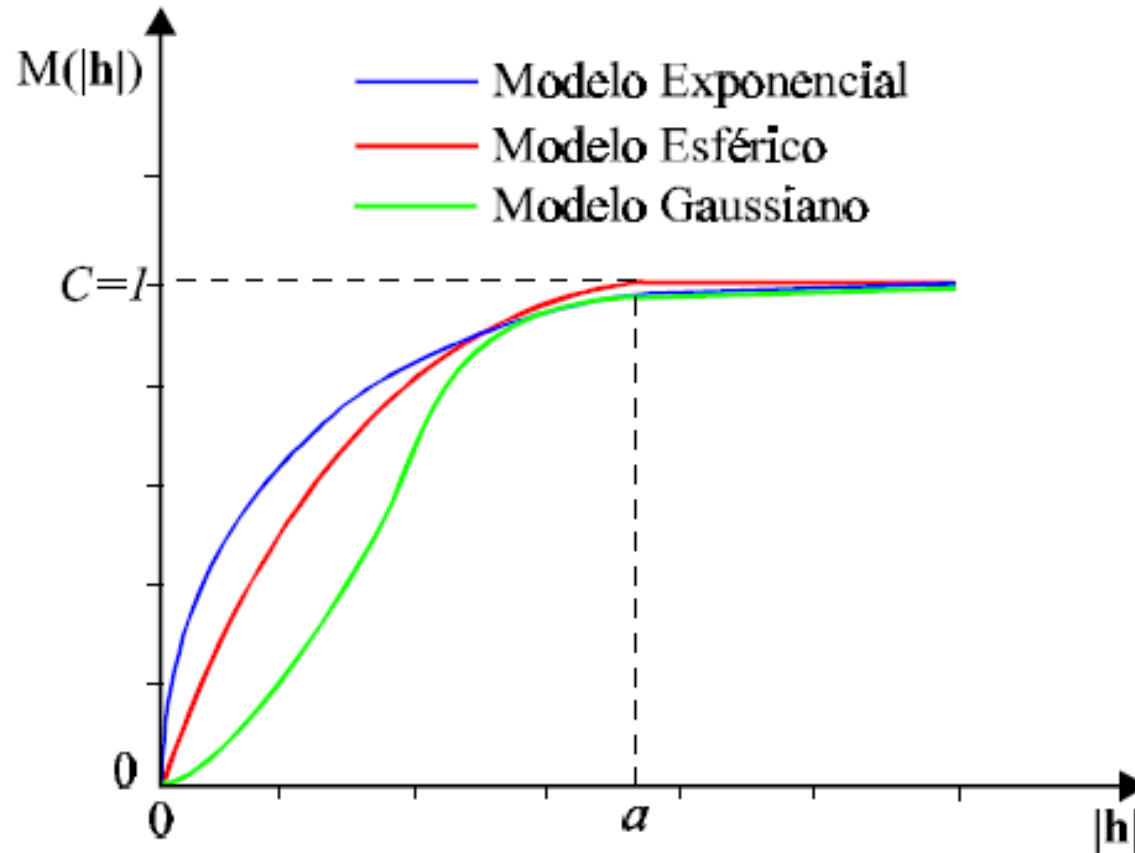
Semivariograma experimental



Ajuste de modelo a um semivariograma

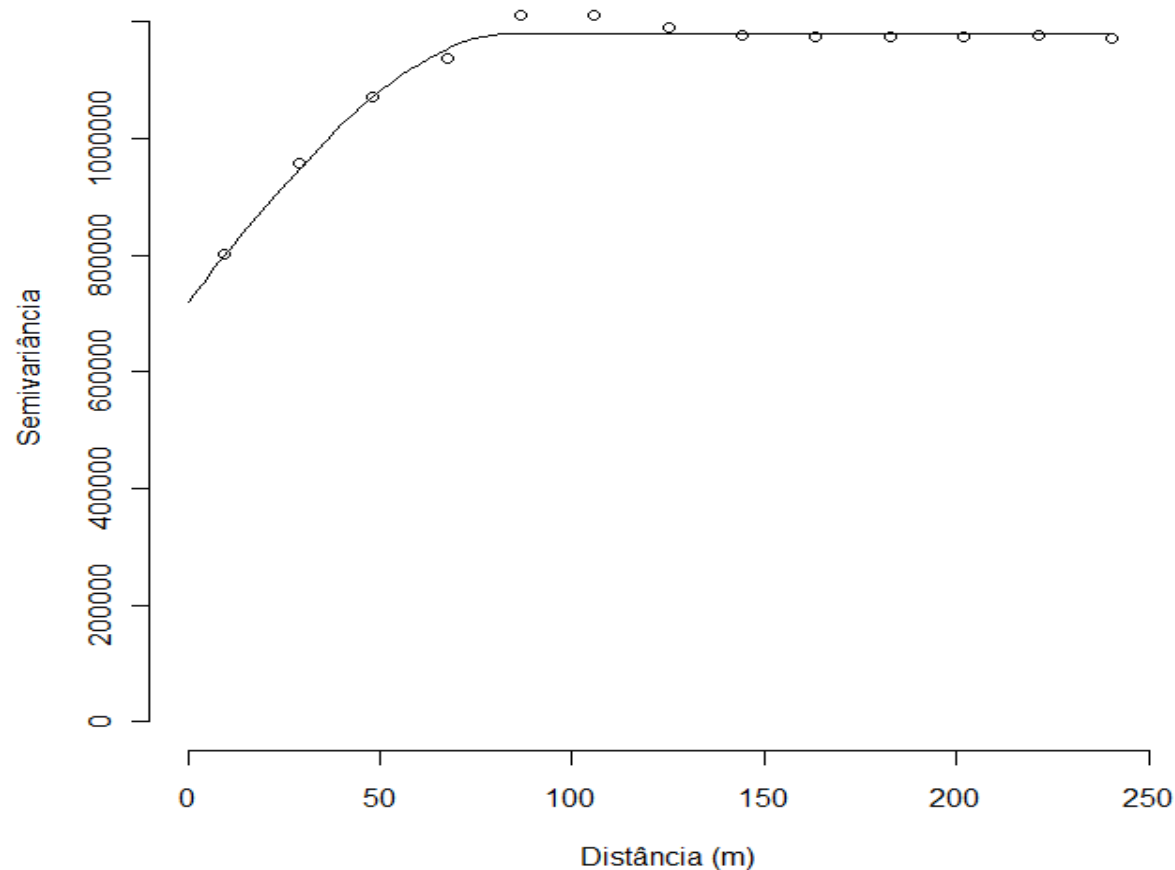
- O ajuste de um modelo ao semivariograma experimental é uma etapa importante pois durante o processo de krigagem será necessário ter disponíveis valores de semivariância para qualquer distância dentro do limite do alcance.
- Existem diversos métodos para ajuste do modelo do semivariograma, dentre eles:
 - ajuste visual
 - minimização da soma dos erros quadrados

Principais modelos de ajuste matemáticos ao semivariograma

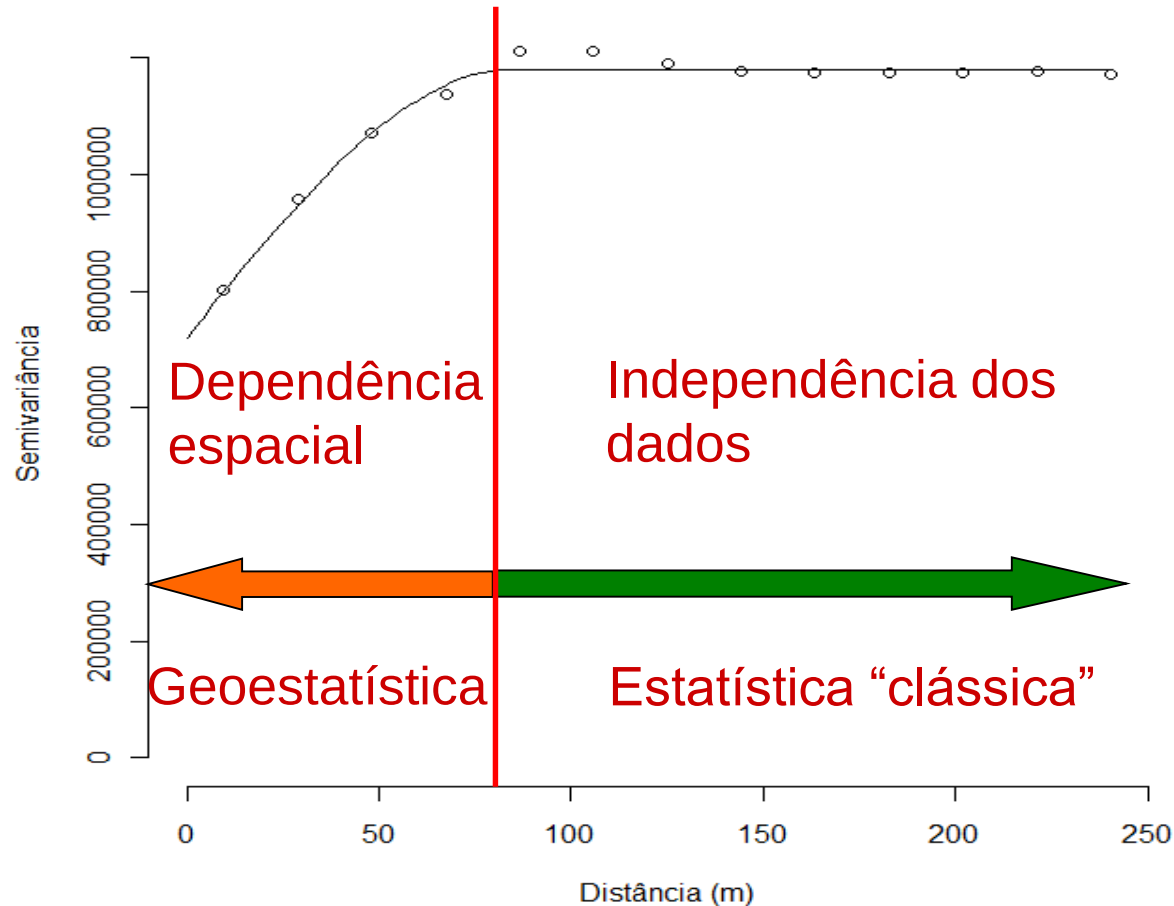


Camargo (1998)

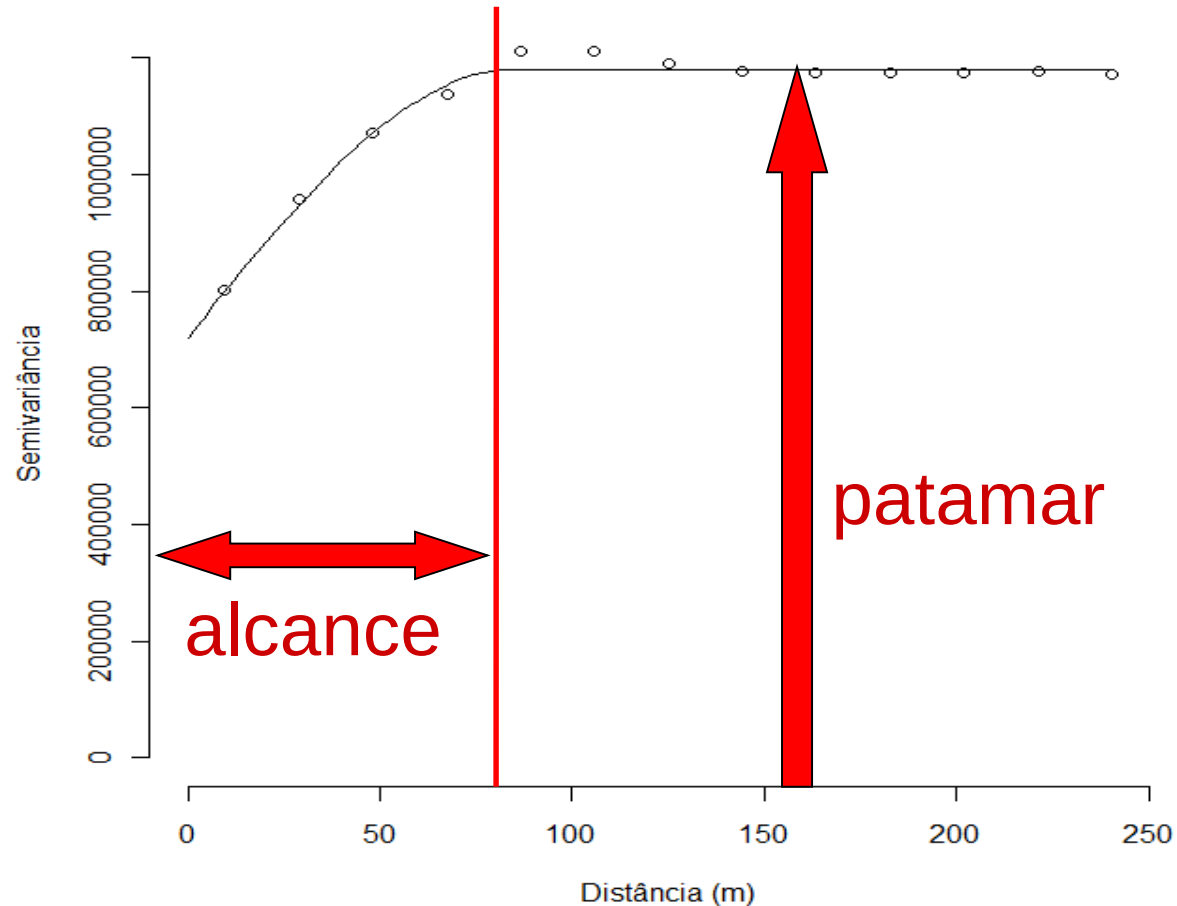
Ajuste de modelo a um semivariograma



Componentes do semivariograma



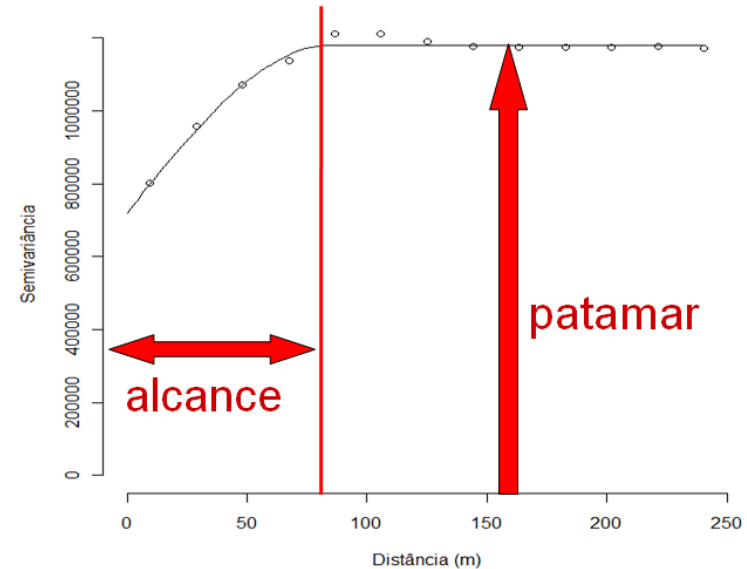
Componentes do semivariograma



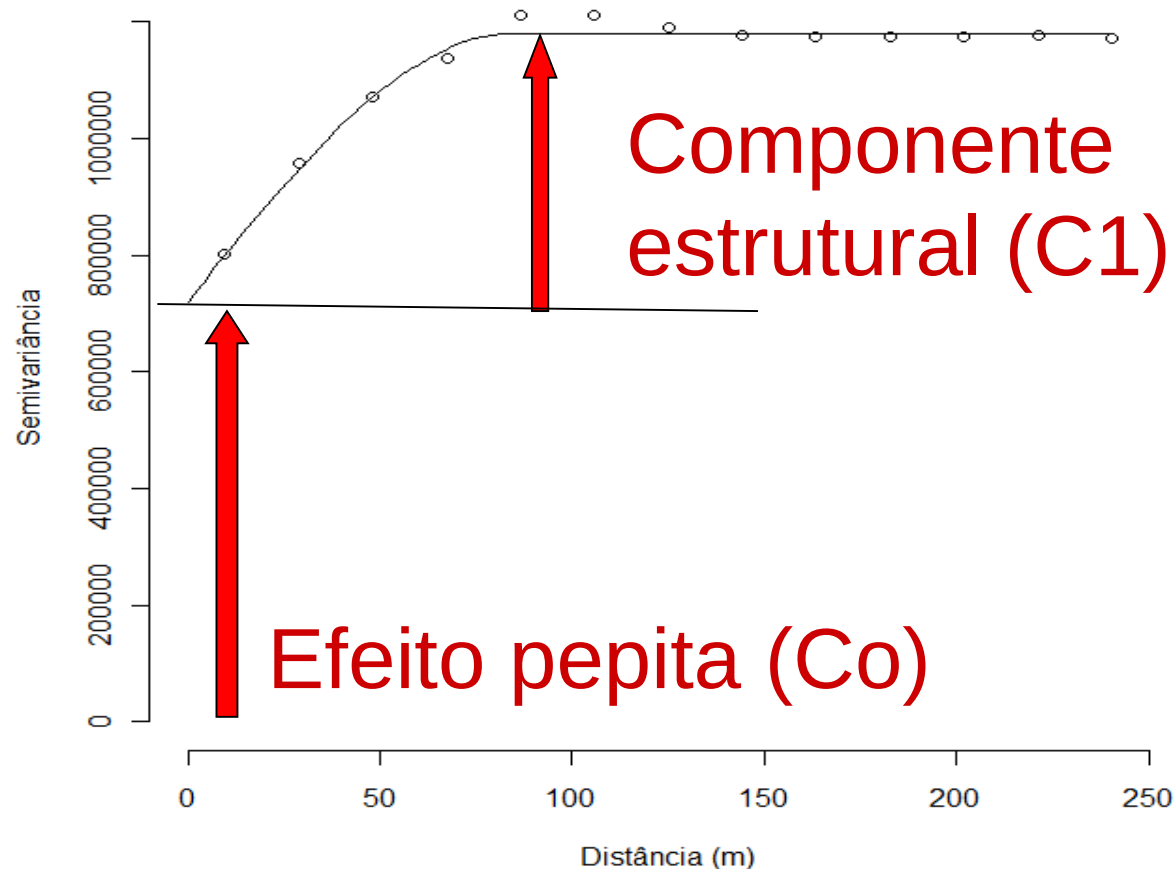
Componentes do semivariograma

- Alcance e Patamar:

A semivariância cresce em função de h até um ponto onde não se observa mais a continuidade espacial. Este ponto no eixo h é chamado de alcance do semivariograma (a) e no eixo $\gamma(h)$ é chamado de patamar).



Componentes do semivariograma



Componentes do semivariograma

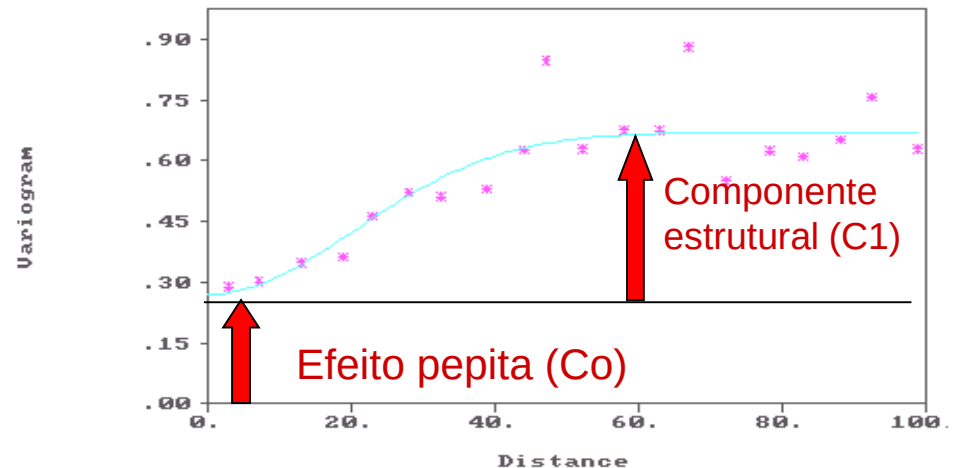
Efeito pepita (C_0):

À medida em que h tende para zero a semivariância $\gamma(h)$ tende para um valor positivo chamado efeito pepita (C_0); o efeito pepita engloba a variação em escala menor que a amostrada e/ou os erros ocorridos nas medições.

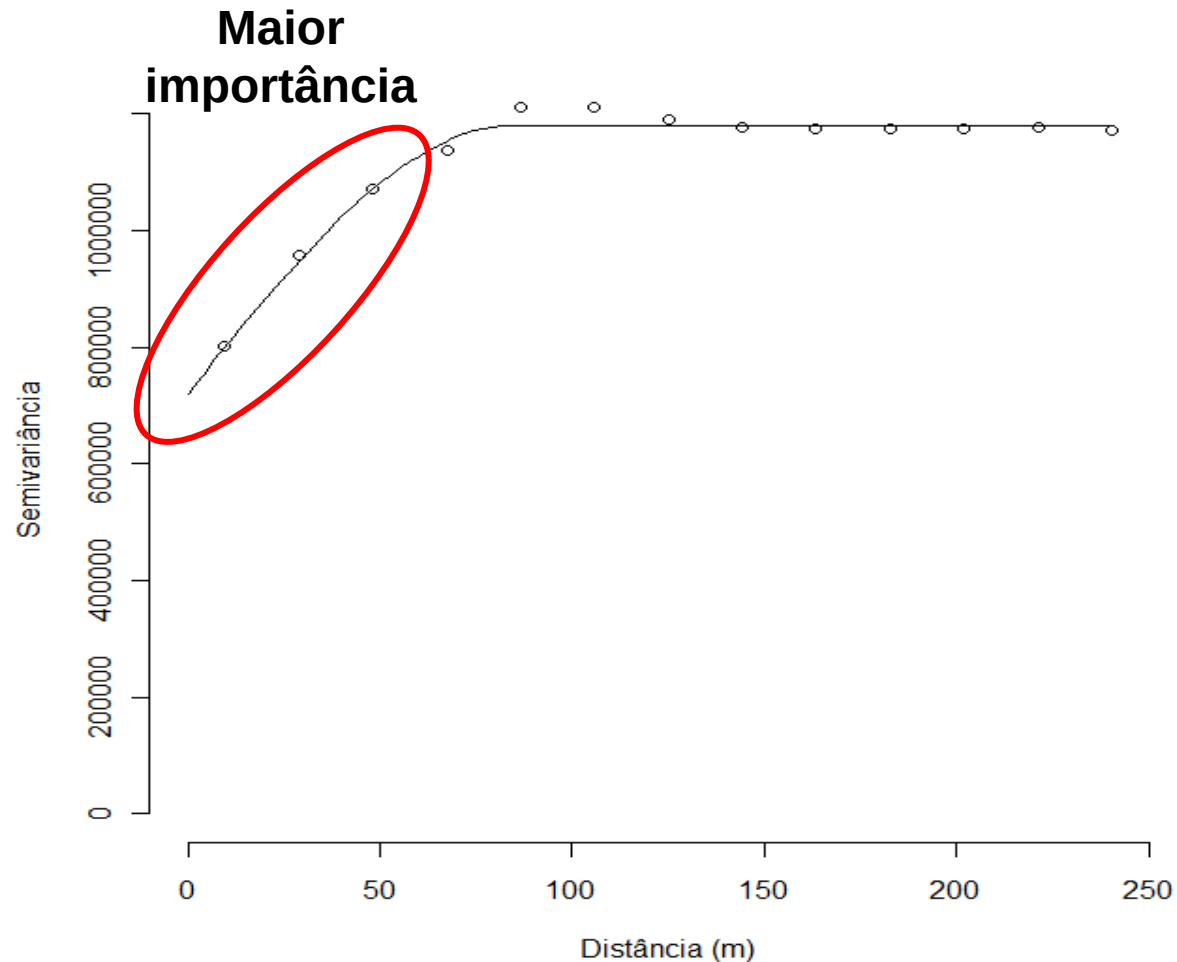
Componente estrutural (C_1):

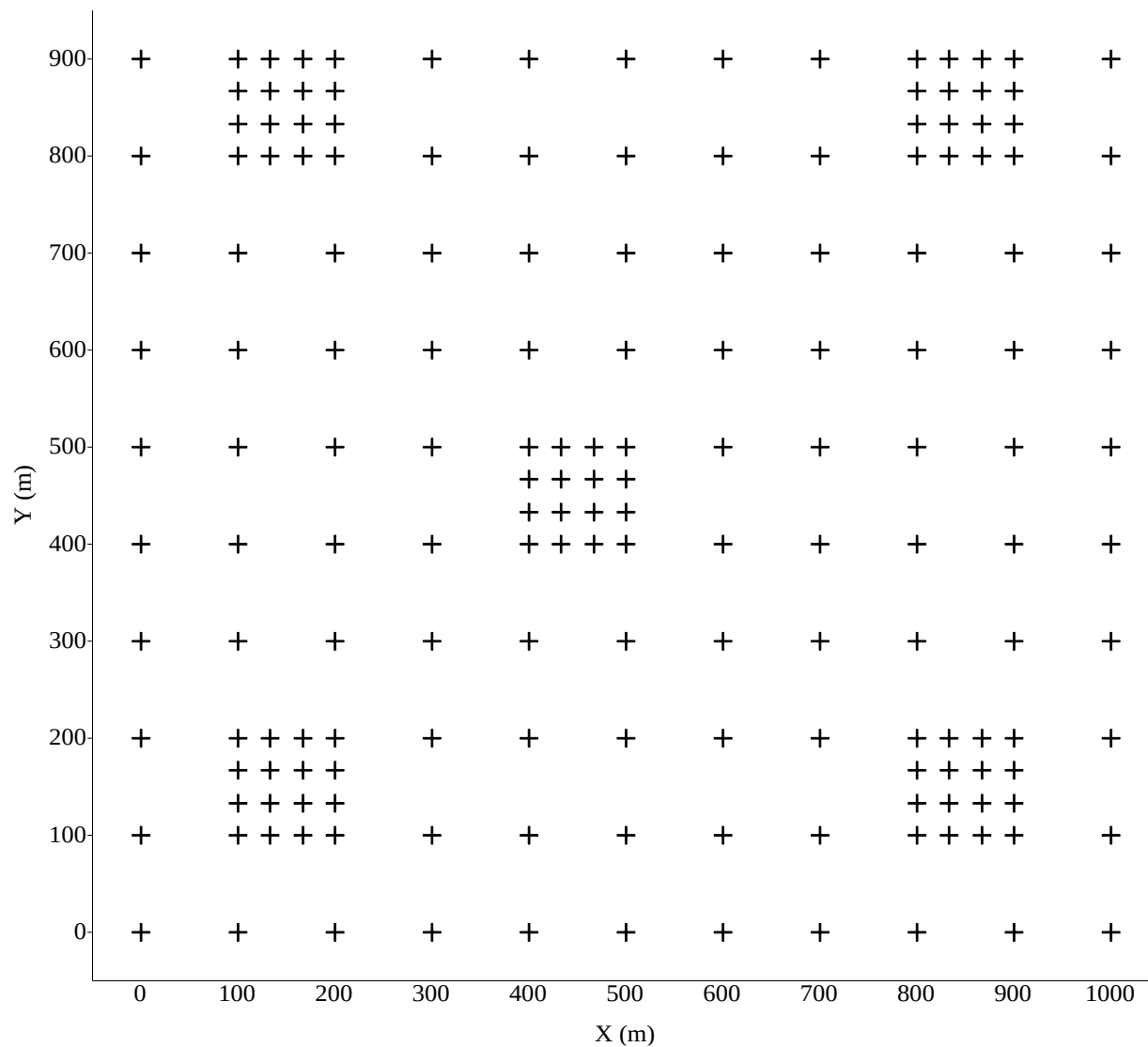
Porção da variação que é explicada pela continuidade espacial.

$C_0 + C_1 = \text{patamar}$



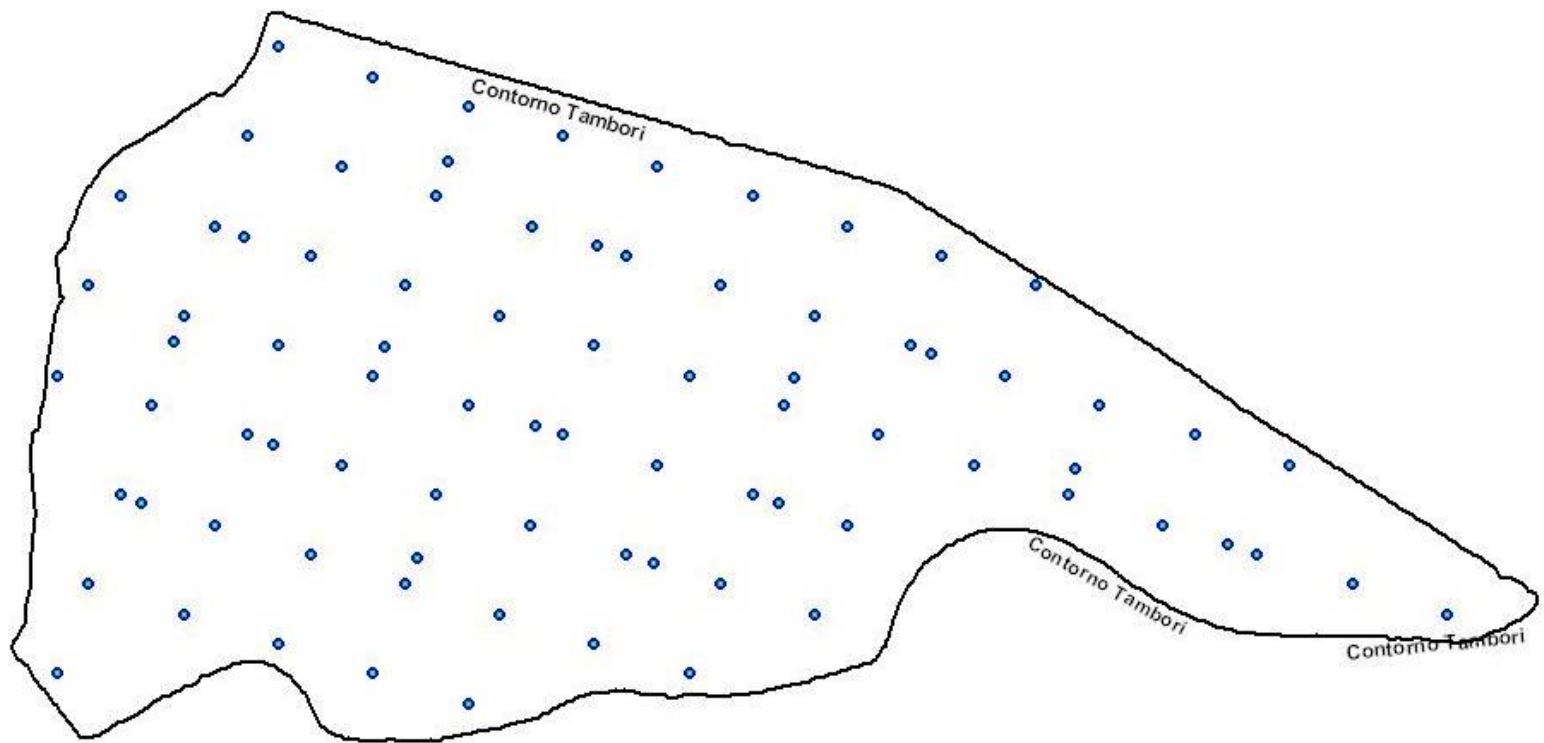
Ajuste de modelo a um semivariograma





Esquema amostral em grade regular acrescida de “ilhas” com espaçamento adensado

MOTOMIYA (2007)

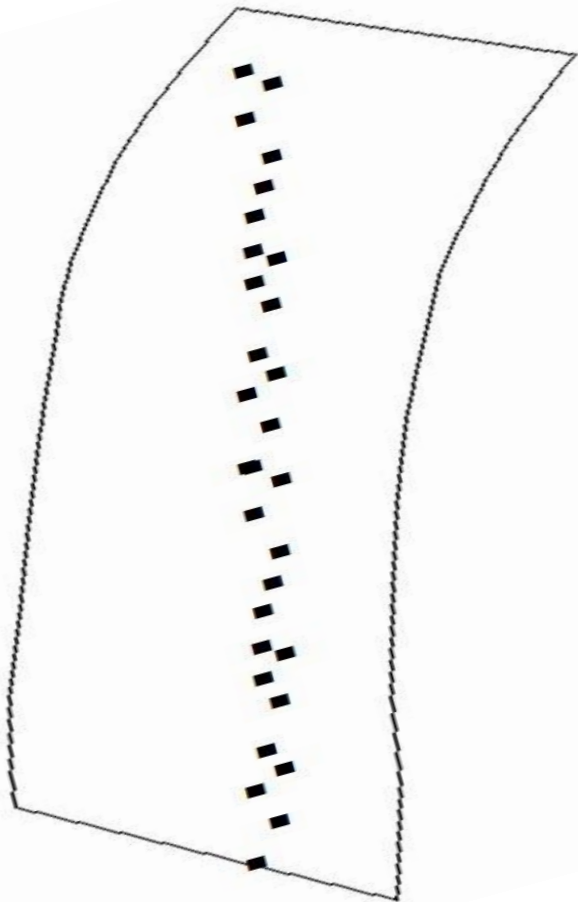


0 100 200 m

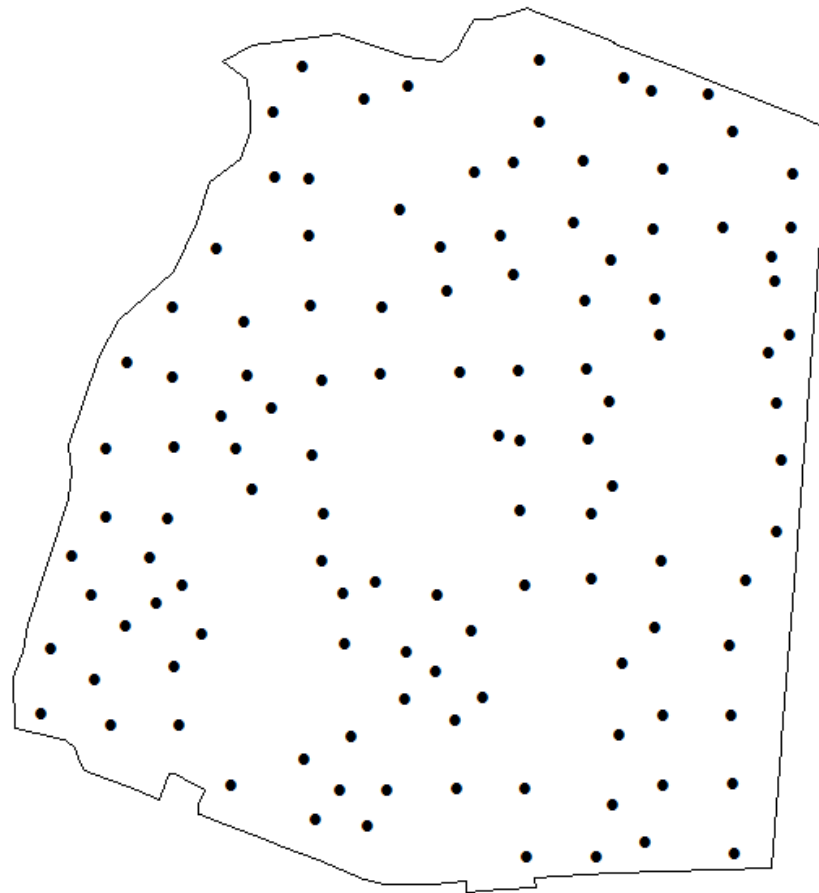
Esquema amostral em grade regular acrescida de pontos com espaçamento adensado

gMAP (2010)

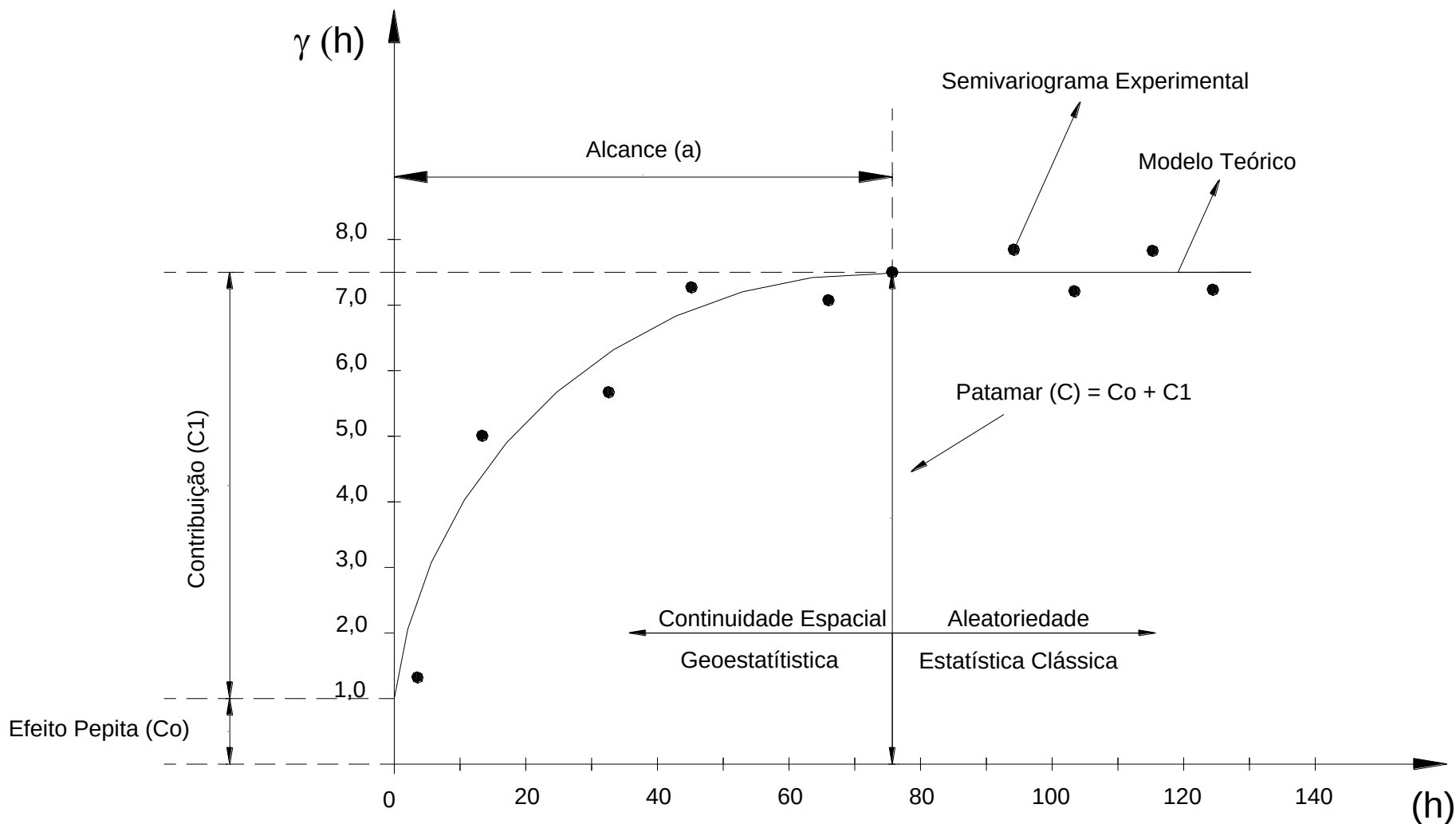
Transversal



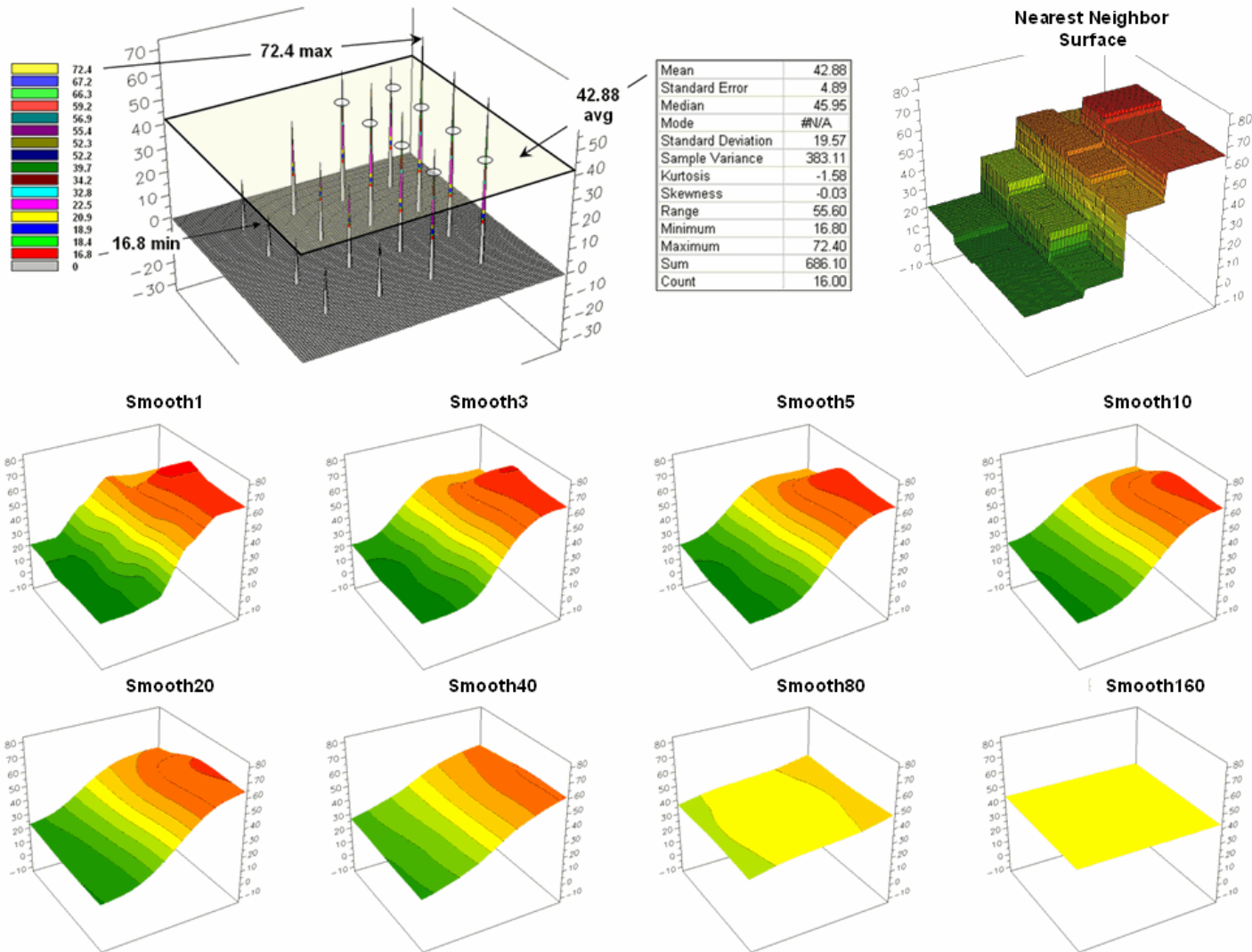
Aleatorização



Componentes do semivariograma (síntese)



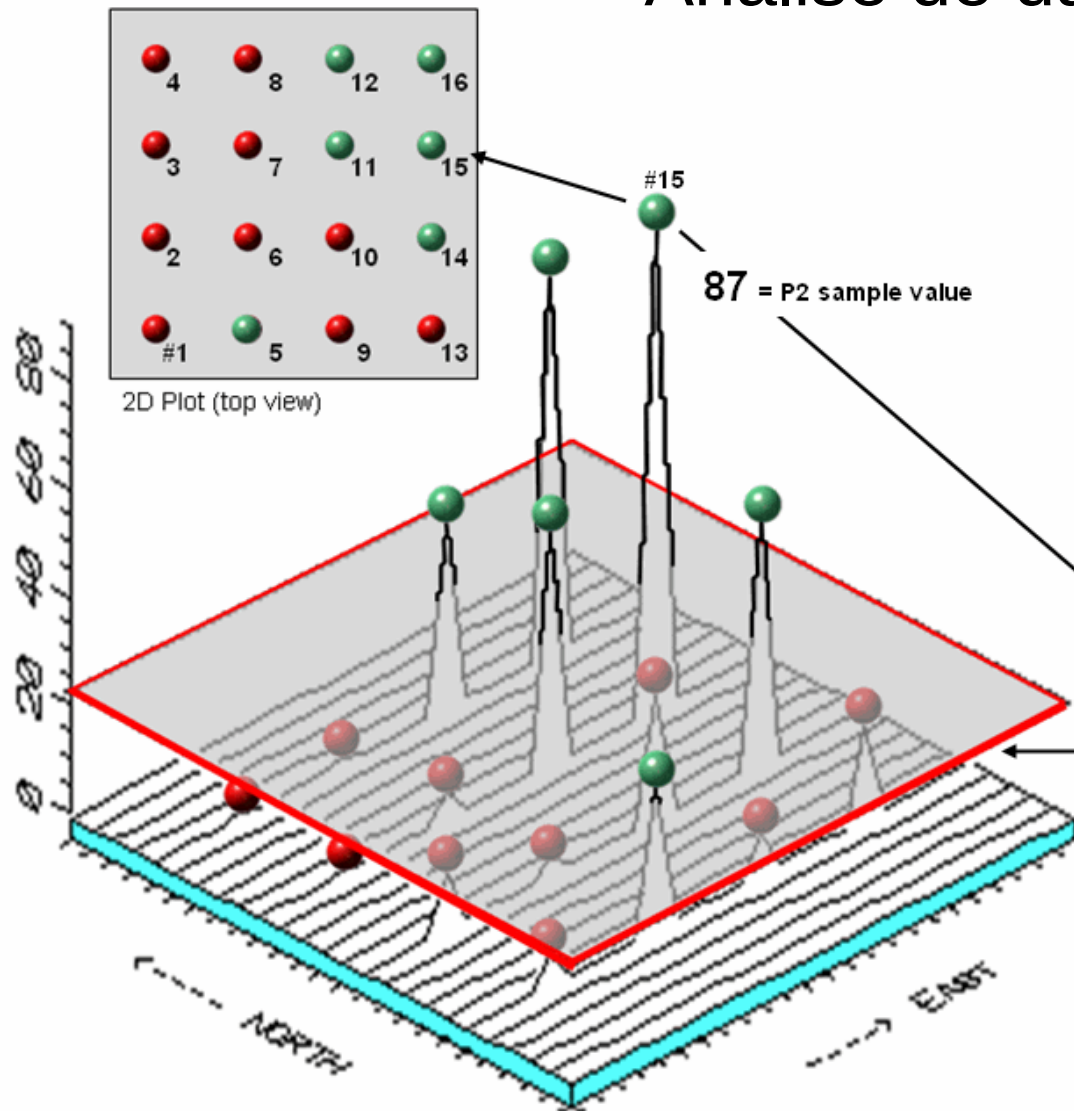
O processo de interpolação



The spatial distribution implied by a set of discrete sample points can be estimated by iterative smoothing of the point values.

© 2007, Joseph K. Berry—permission to copy granted

Análise de dados



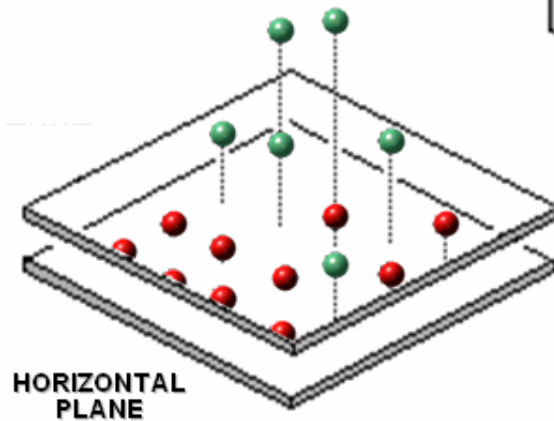
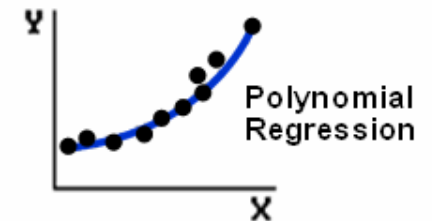
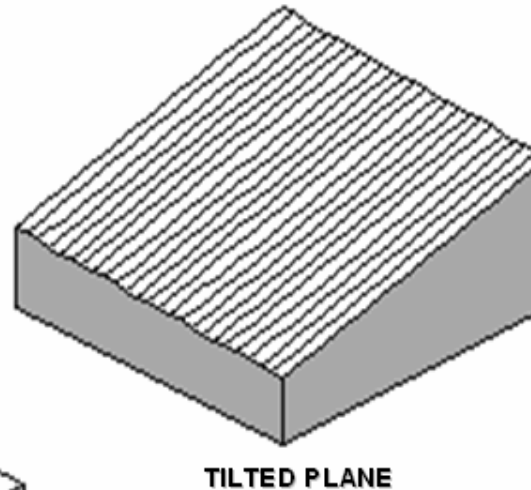
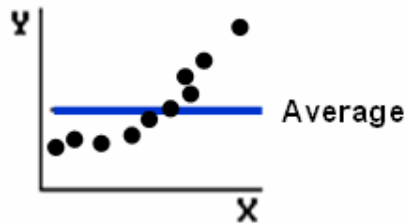
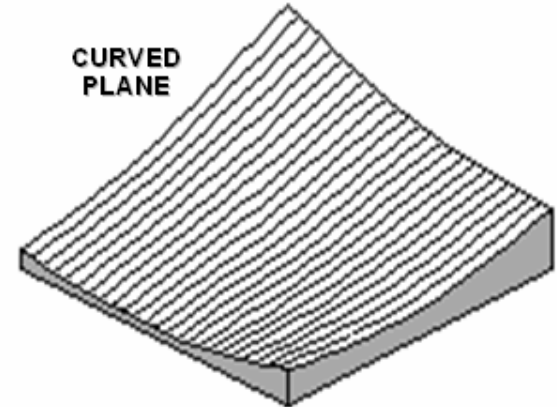
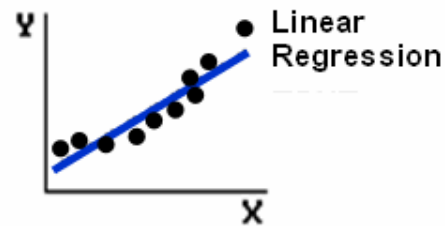
Field Collected Data

Sample	X	Y	P1	P2
#1	1000	1000	11	4
2	1000	1500	19	9
3	1000	2000	8	0
4	1000	2500	0	0
5	1500	1000	27	25
6	1500	1500	12	2
7	1500	2000	14	4
8	1500	2500	2	0
9	2000	1000	10	6
10	2000	1500	17	22
11	2000	2000	34	42
12	2000	2500	22	33
13	2500	1000	20	16
14	2500	1500	28	43
15	2500	2000	42	87
16	2500	2500	34	68
Total			300	361
Average			19	23
StDev			12	26

Arithmetic Average – plot of the data average is a horizontal plane in 3-dimensional geographic space with some of the data points balanced above (green) and some below (red) the “typical” value
(uniform estimate of the spatial distribution)

A comparison shows the spatial distributions of field samples (floating balls) and their arithmetic average (horizontal plane).

© 2007, Joseph K. Berry—permission to copy granted

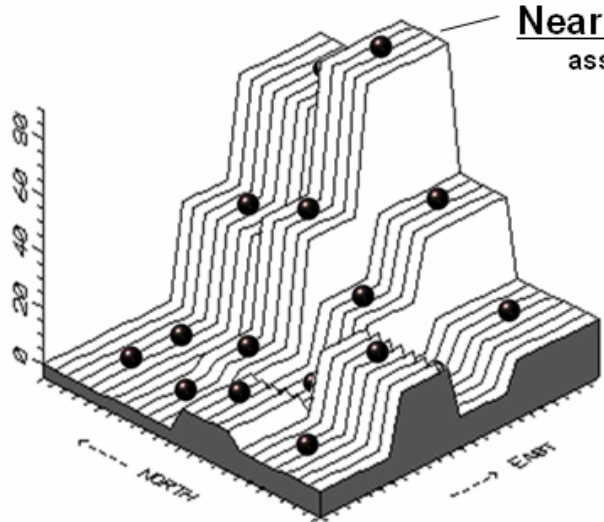


Map Generalization – fits standard functional forms to the data, such as a N^{th} order polynomial for curved surfaces with several peaks and valleys

(similar to curve fitting techniques in traditional statistics)

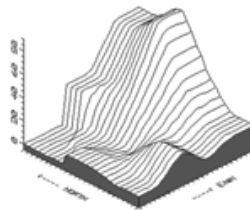
Map generalization techniques can be used to approximate geographic trends.

© 2007, Joseph K. Berry—permission to copy granted

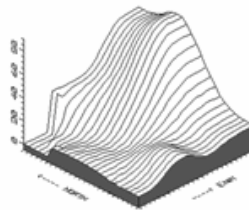


Nearest Neighbor Surface – all locations are assigned the value of the closest sample
(first order estimate of the spatial distribution)

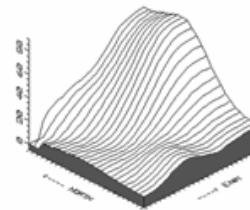
Smooth 1



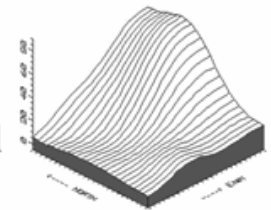
Smooth 2



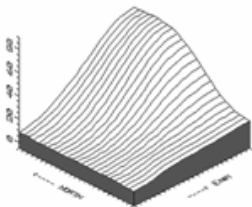
Smooth 3



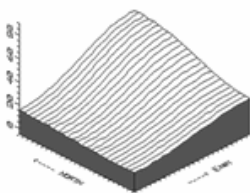
Smooth 4



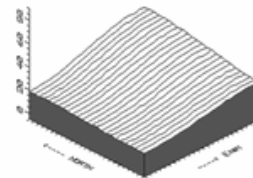
Smooth 9



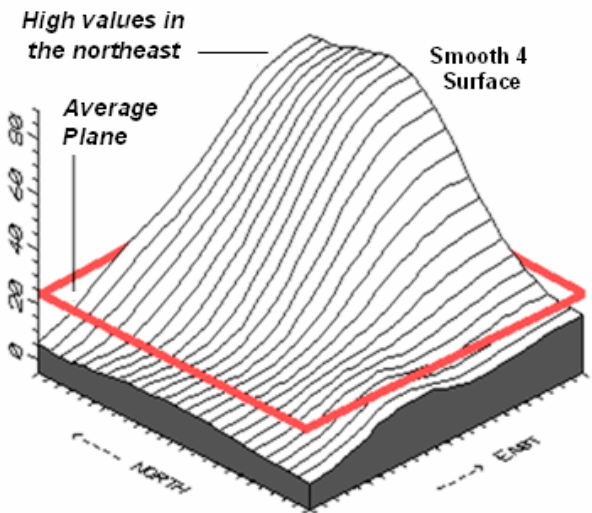
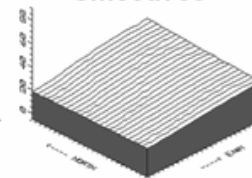
Smooth 29



Smooth 49



Smooth 99

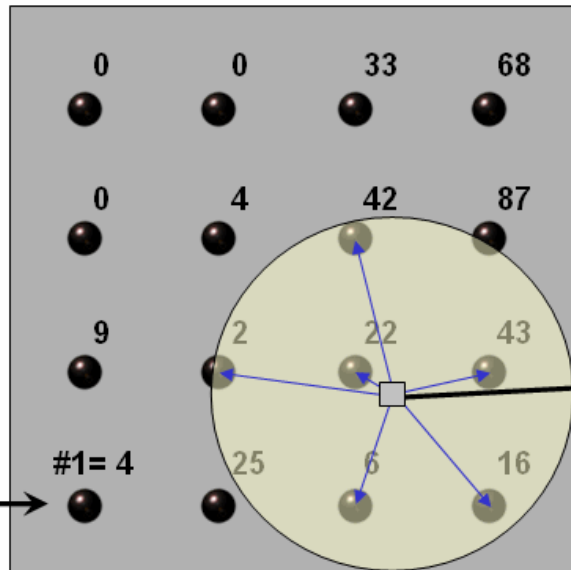


Iteratively smoothed approximations show localized geographic distribution.

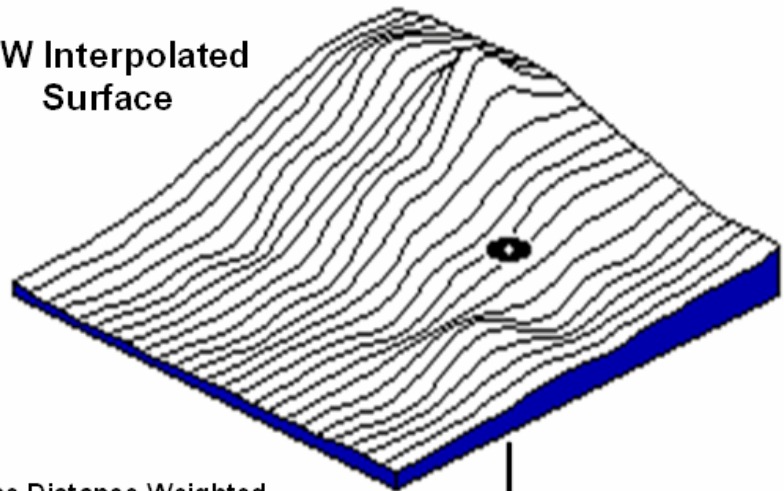
© 2007, Joseph K. Berry—permission to copy granted

Sample	X	Y	P1	P2
#1	1000	1000	11	4
2	1000	1500	19	9
3	1000	2000	8	0
4	1000	2500	0	0
5	1500	1000	27	25
6	1500	1500	12	2
7	1500	2000	14	4
8	1500	2500	2	0
9	2000	1000	10	6
10	2000	1500	17	22
11	2000	2000	34	42
12	2000	2500	22	33
13	2500	1000	20	16
14	2500	1500	28	43
15	2500	2000	42	87
16	2500	2500	34	68

Field Collected Data



IDW Interpolated Surface

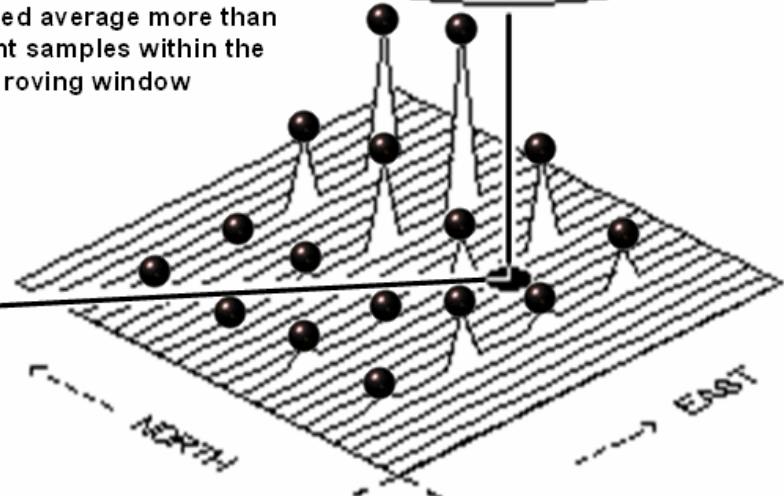


Inverse Distance Weighted Interpolation for...

Location X= 2200, Y= 1300

22.14

... closer samples influence weighted average more than distant samples within the roving window



A roving window is used to identify and summarize nearby sample values.

© 2007, Joseph K. Berry—permission to copy granted

			Distance	Weight (1/D ²)	Value	Weight * Value
Location	2200	1300				
#6	1500	1500	728.01	0.00000189	2	0.0000038
#9	2000	1000	360.56	0.00000769	6	0.0000462
#10	2000	1500	282.84	0.00001250	22	0.0002750
#11	2000	2000	728.01	0.00000189	42	0.0000792
#13	2500	1000	424.26	0.00000556	16	0.0000889
#14	2500	1500	360.56	0.00000769	43	0.0003308
			Sum =	0.00003721		0.0008238
Interpolated Value (Sum W*V / Sum W) = .0008238 / .00003721 =						22.14

Location X= 2200, Y= 1300

Sample #10 X= 2000, Y= 1500

Distance = SQRT [(2200 – 2000)² + (1300 - 1500)²] = SQRT [(40,000) + (40,000)] = **282.84**

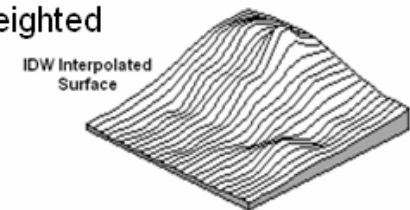
Inverse Distance Squared = 1 / D² = 1 / (282.84)² = **0.00001250**

Sample #10 Value = 22

Weight₁₀ * Value₁₀ = 0.00001250 * 22 = **0.0002750**

...repeat for other sample points, then calculate the **Sum of the Weights** and the **Sum of the Products** and divide for the distance-weighted average **Interpolated Value** at the location = **22.14**

...repeat for all of the other locations in the project area to generate the **IDW Interpolated Surface**

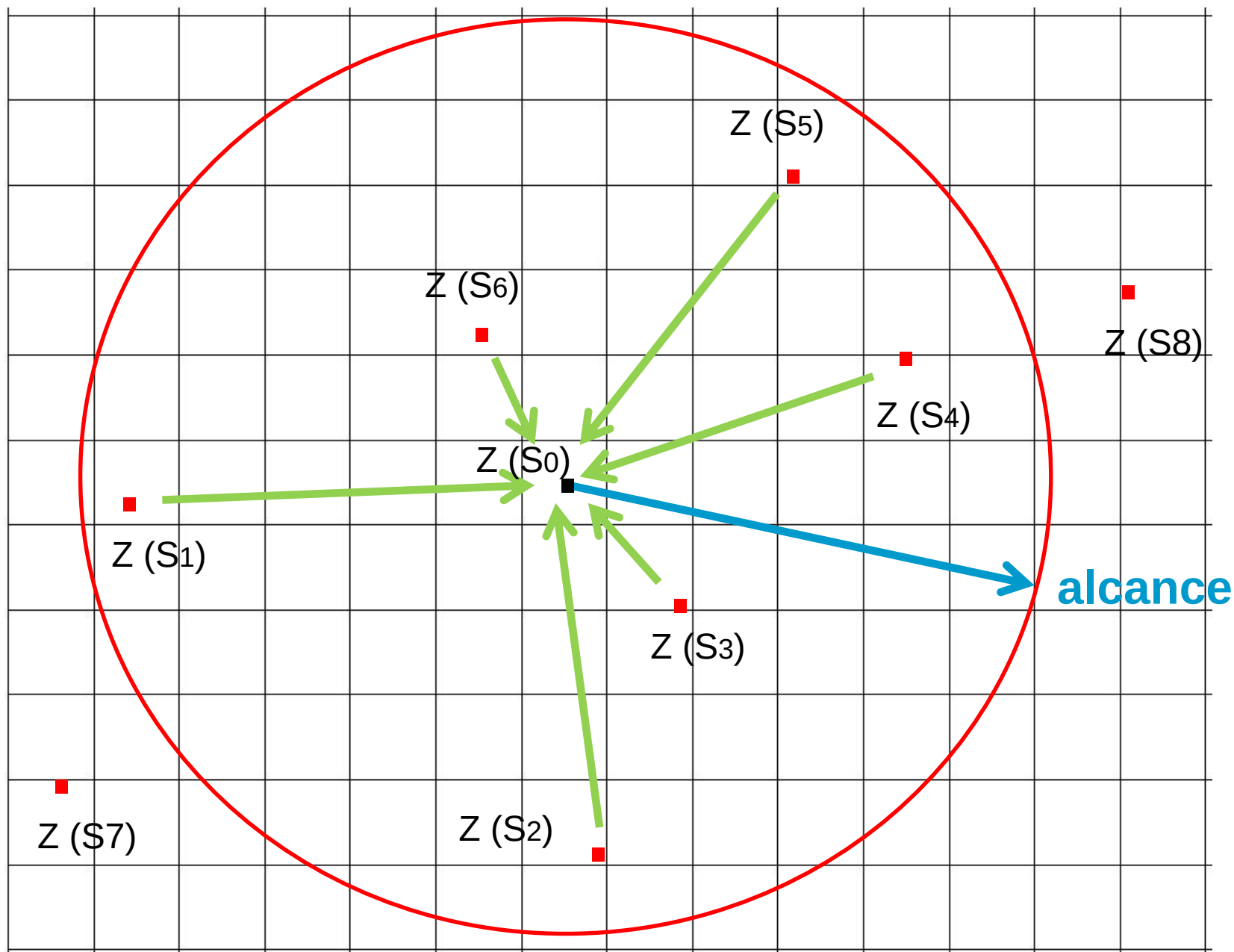


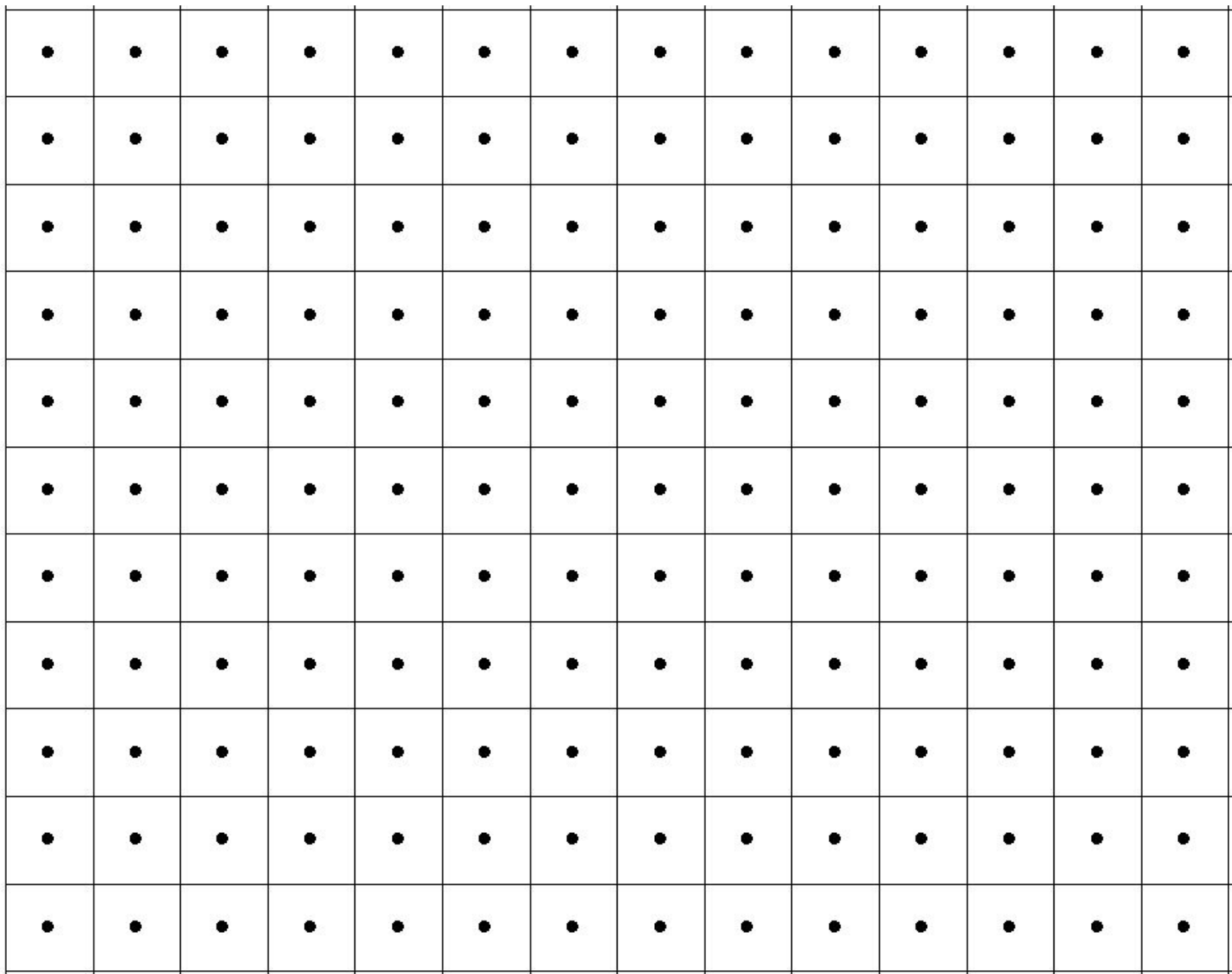
IDW calculations use simple geographic distance to weight the average of samples within a roving window.

© 2007, Joseph K. Berry—permission to copy granted

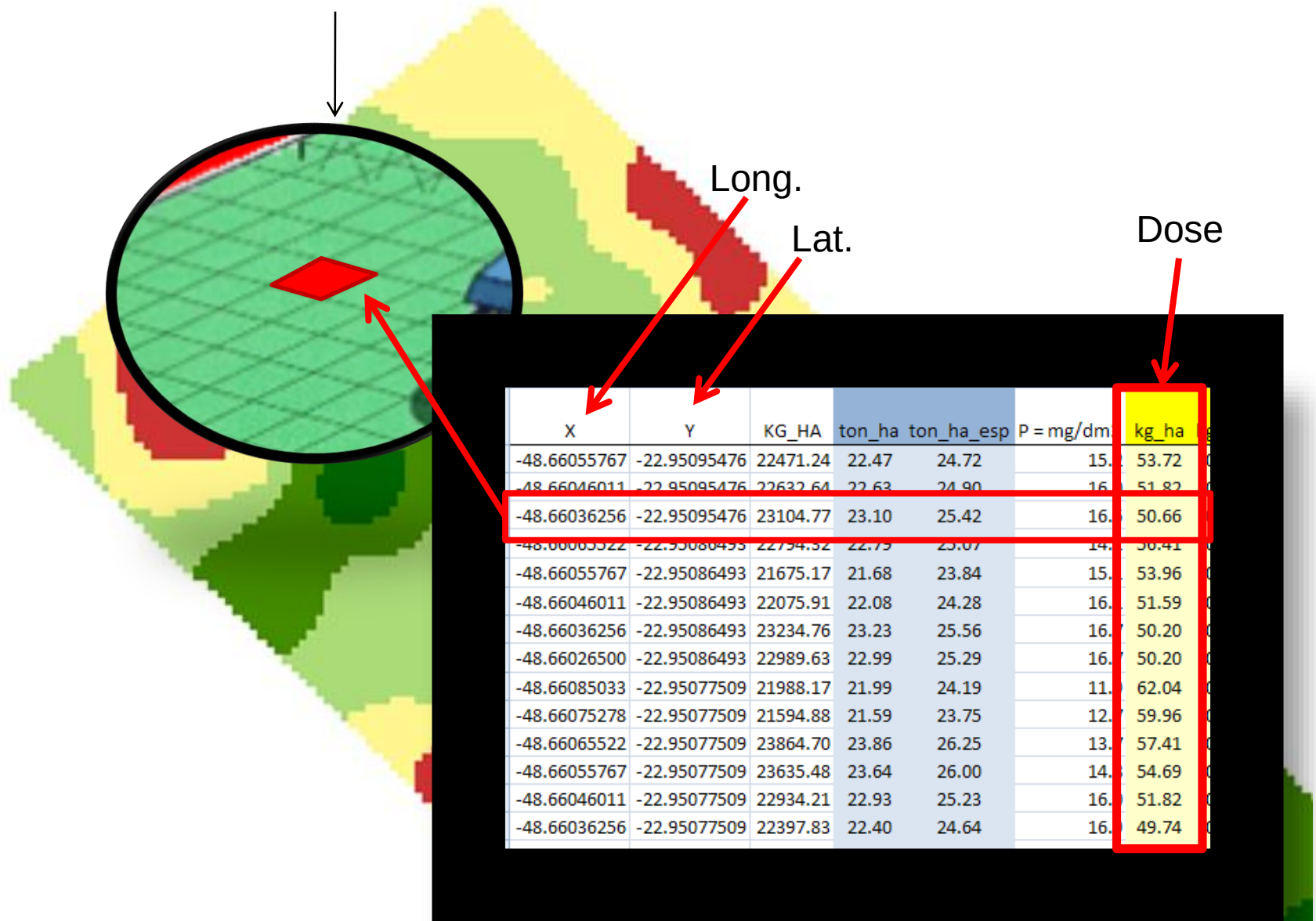
Krigagem

- É um estimador geoestatístico (o único).
- Leva em consideração a continuidade espacial da variável no processo de estimação da mesma.

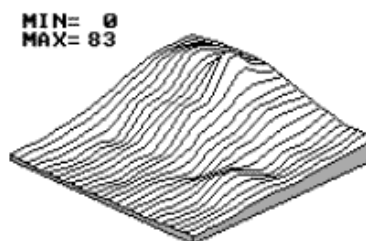




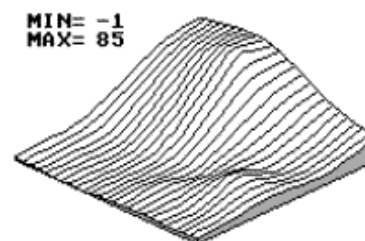
Cada pixel com
um valor



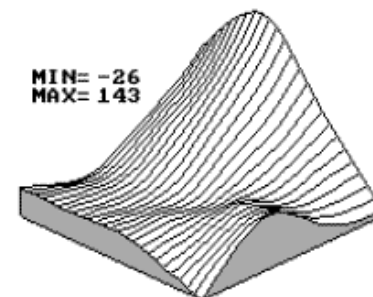
... the best map is the one that has the best "guesses"



INVERSE DISTANCE



KRIGING



MINIMUM CURVATURE

Sample	Col	Row	Actual	Average	Inverse	Kriging	MinCurve
#17*	1	1	0	23 (23)	8 (8)	2 (2)	6 (6)
18*	18	2	48	23 (-25)	42 (-6)	46 (-2)	28 (-20)
19*	23	2	64	23 (-41)	52 (-12)	65 (1)	65 (1)
20*	19	4	65	23 (-42)	54 (-11)	56 (-9)	48 (-17)
21	15	6	34	23 (-11)	33 (-1)	30 (-4)	31 (-3)
22	5	7	0	23 (23)	2 (2)	-1 (-1)	1 (1)
23	9	8	6	23 (17)	7 (1)	1 (-5)	1 (-5)
24	19	11	79	23 (-56)	67 (-12)	70 (-9)	69 (-10)
25	23	13	64	23 (-41)	52 (-12)	68 (4)	90 (26)
26*	4	16	8	23 (15)	8 (0)	7 (-1)	6 (-2)
27	16	17	19	23 (4)	22 (3)	19 (0)	17 (-2)
28*	2	20	6	23 (17)	8 (2)	3 (-3)	-6 (-12)
29	13	22	12	23 (11)	15 (3)	14 (2)	19 (7)
30	22	22	17	23 (6)	19 (2)	20 (3)	7 (-10)
31*	2	24	9	23 (14)	8 (-1)	6 (-3)	-16 (-25)
32	19	24	14	23 (9)	19 (5)	11 (-3)	-7 (-21)
Test Set Average= 28							
Average Estimate= 23					26	26	22
Sum of the Residuals=				(-77)	(-29)	(-28)	(-86)
Average Unsigned Residual=				(22.2)	(5.1)	(3.3)	(10.5)
Normalized Residual Index=				(.80)	(.18)	(.12)	(.38)

Residual table identifying actual and estimated values.

© 2007, Joseph K. Berry—permission to copy granted

Como saber se a interpolação foi acertada?

A validação cruzada é a técnica de avaliação de erros de estimativa que permite comparar valores preditos com os amostrados.

O valor da amostra, no local $Z(s_i)$, é temporariamente descartado do conjunto de dados e, então, é feita uma previsão por krigagem no local $\hat{Z}(s_{(i)})$ usando-se as amostras restantes.

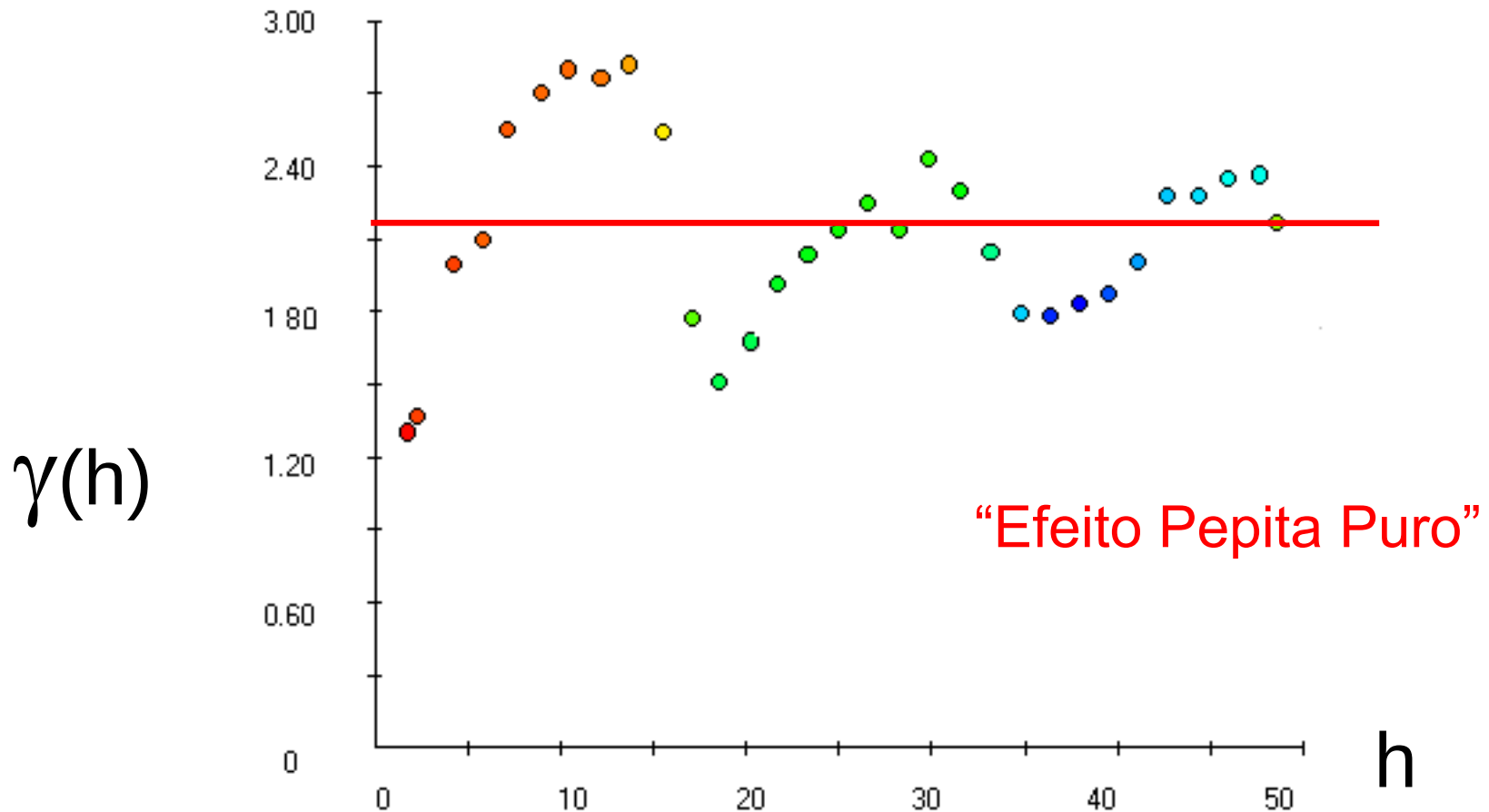
Desta forma é possível retirar alguns valores e estimar, por exemplo, o Erro Médio (EM):

$$EM = \frac{1}{n} \sum_{i=1}^n \left(Z(s_i) - \hat{Z}(s_{(i)}) \right)$$

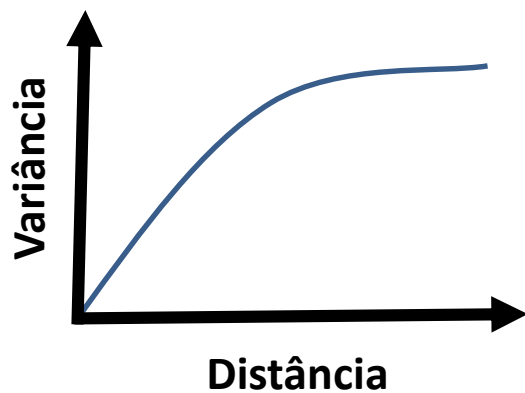
onde n é o número de dados; $Z(s_i)$, valor observado no ponto s_i ; e $\hat{Z}(s_{(i)})$ valor predito por krigagem ordinária no ponto s_i , sem considerar a observação $Z(s_i)$

E quando não é possível obter os parâmetros do semivariograma?

- E quando não é possível obter os parâmetros do semivariograma?

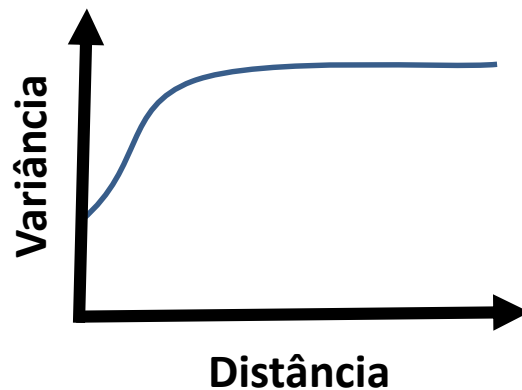


Argila



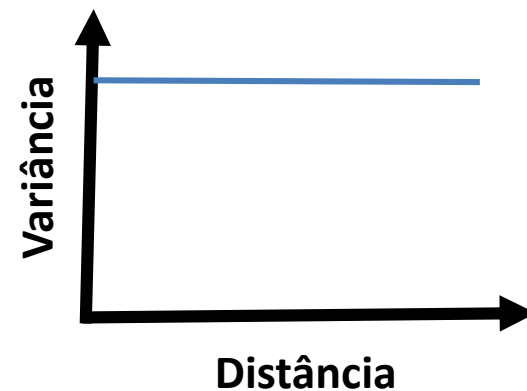
**Alta
dependência
espacial**

pH

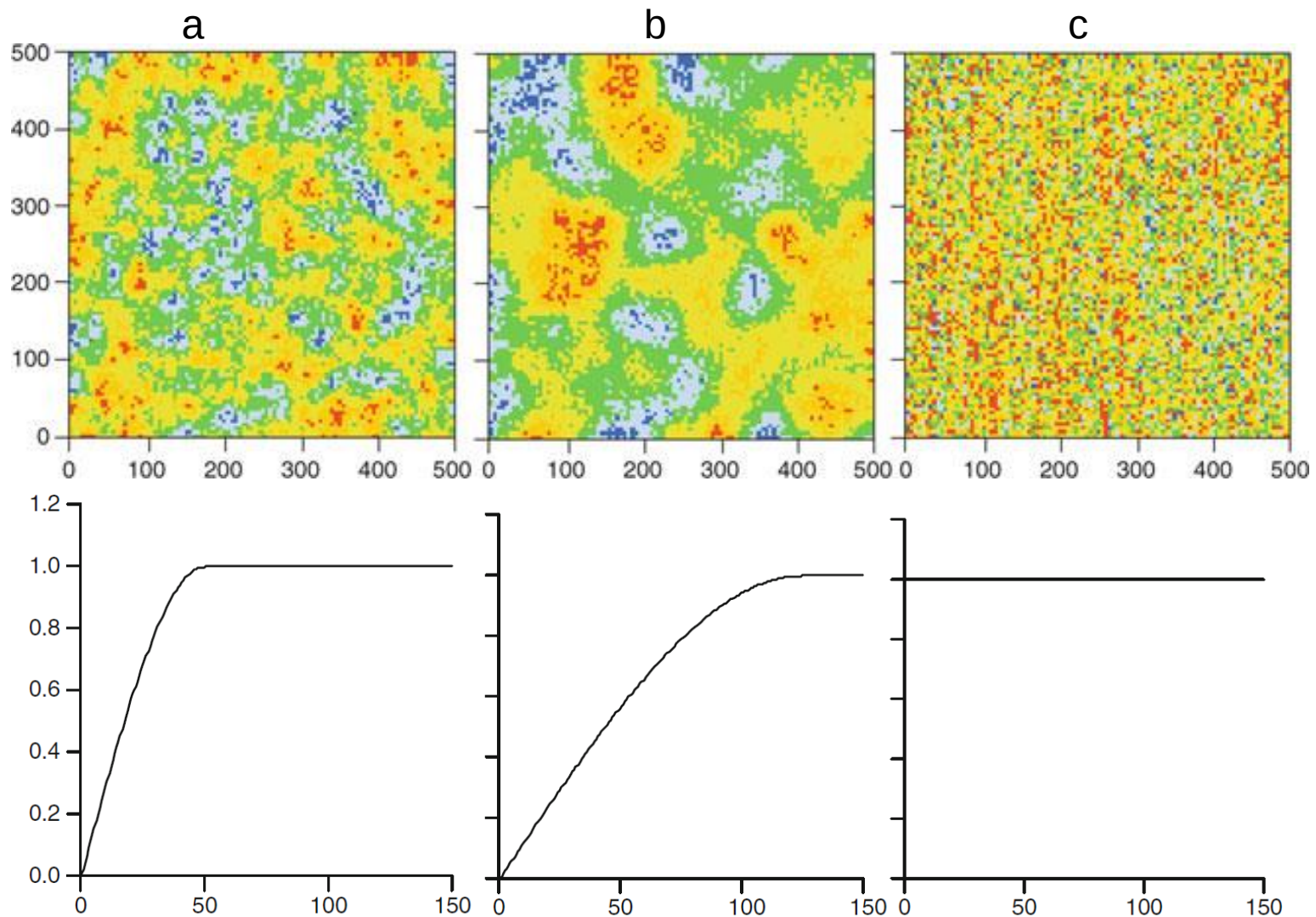


**Dependência espacial
+
Efeito pepita**

Fósforo

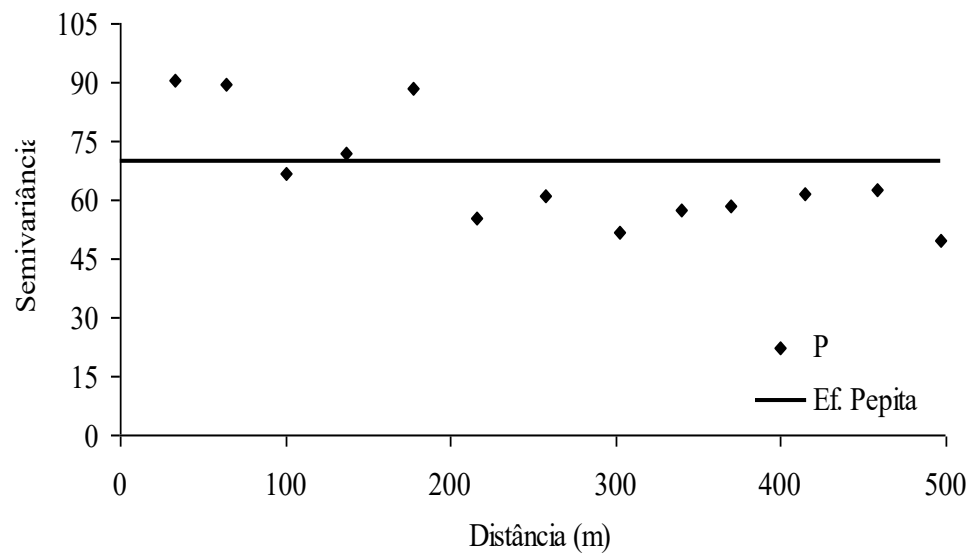
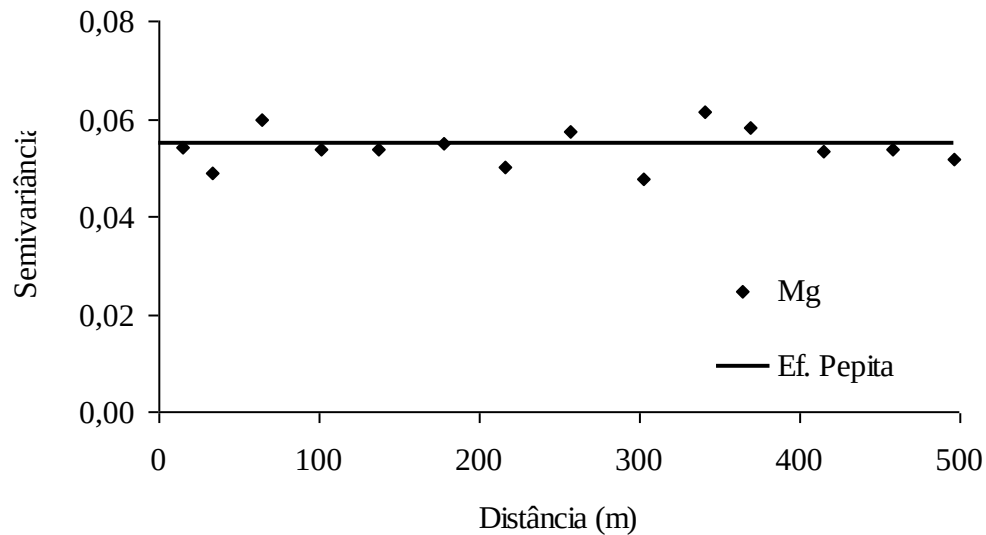
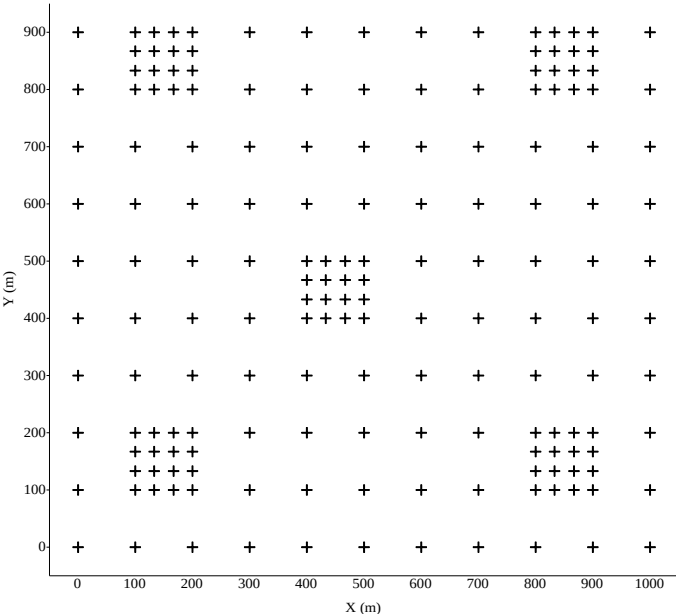


Efeito pepita puro



Distância entre amostras menor do que “manchas”

Talhões simulados com funções esféricas, com efeito pepita zero, patamar 1 e alcances de: (a) 50 m e (b) 125m; (c) variograma com efeito pepita puro (Kerry et al., 2010)



MOTOMIYA (2007)