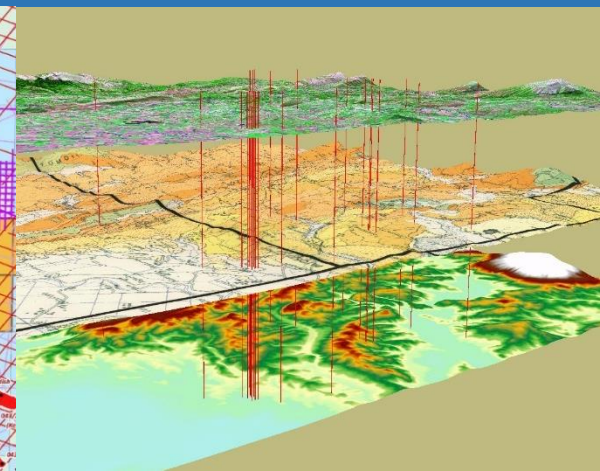
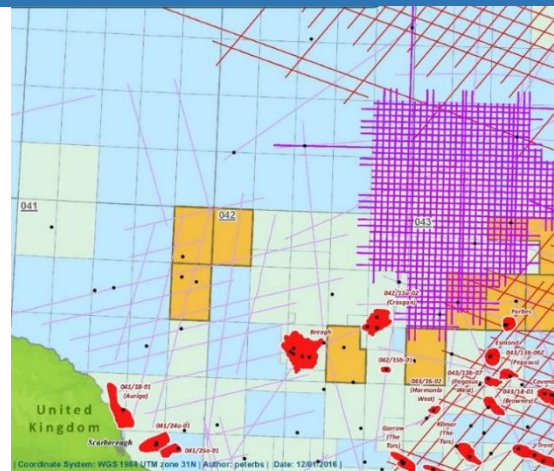
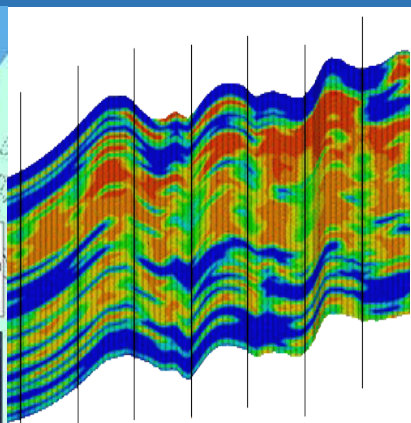


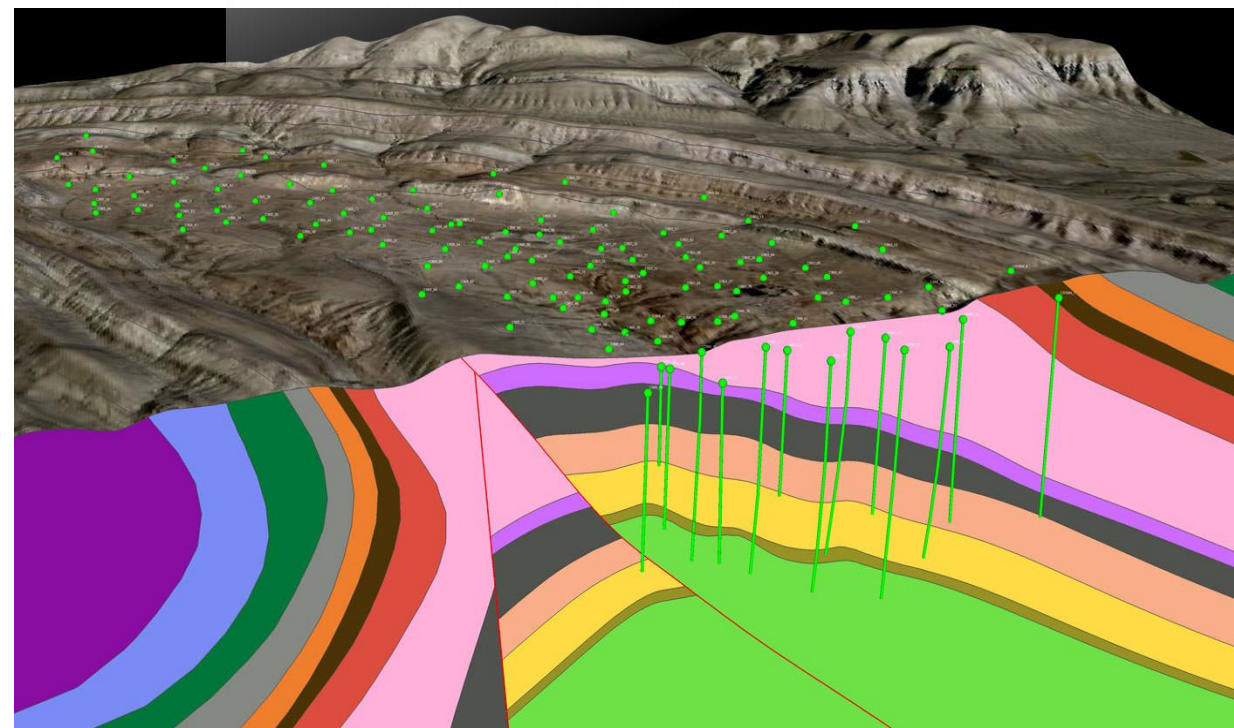
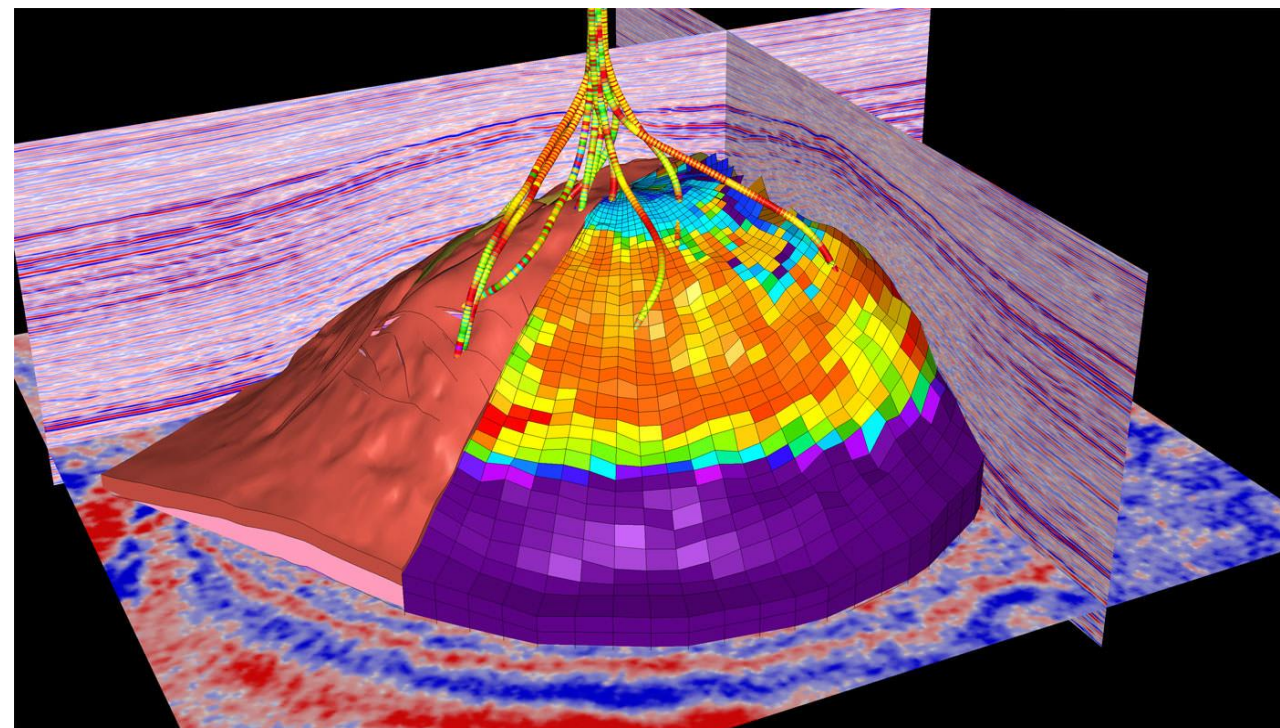
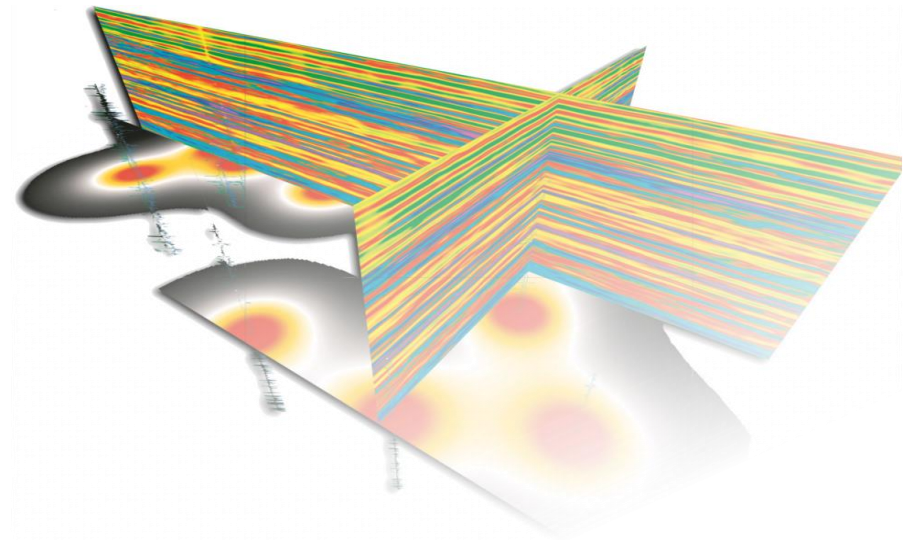


PMI 3331 – GEOMÁTICA APLICADA À ENGENHARIA DE PETRÓLEO





# ANÁLISES E ESTATÍSTICAS EM DADOS ESPACIAIS

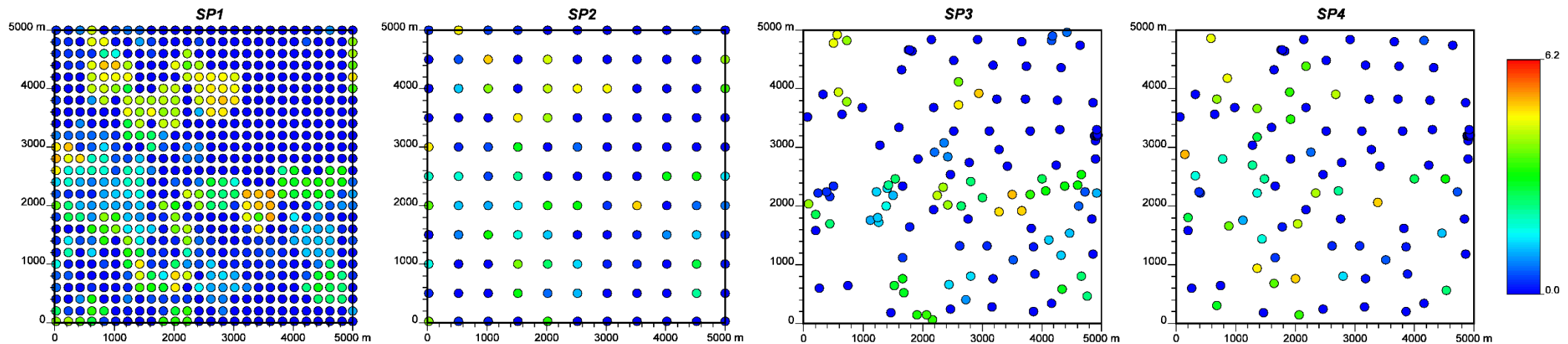




# Amostra e Métodos de Amostragem

*Amostra é um subconjunto de valores do fenômeno espacial que, se representativa, deve reproduzir a distribuição e variabilidade espaciais tanto em tamanho, isto é, número de pontos de dados, como em termos de distribuição dos pontos no domínio a ser estudado.*

*Yamamoto e Landim, 2013*



*Fonte Imagem: Martinez et al., 2010*



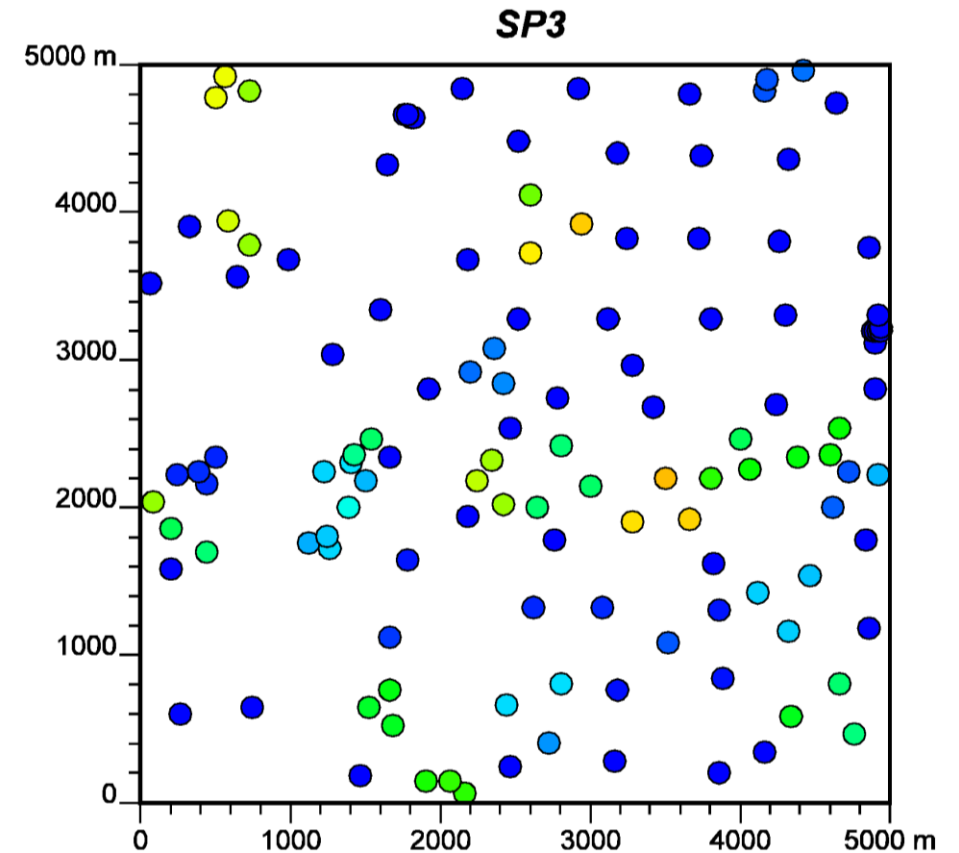


# Amostragem Aleatória Simples

A componente aleatória é o número de  $N$  unidades, onde  $n$  destas são sorteadas sem reposição. Em estudos geoestatísticos os pontos de amostragem estão localizados no espaço e representam, portanto, as coordenadas espaciais.

Comp. Aleatória  $\rightarrow$  núm. sequencial entre 1 e  $N$ .

Yamamoto e Landim, 2013



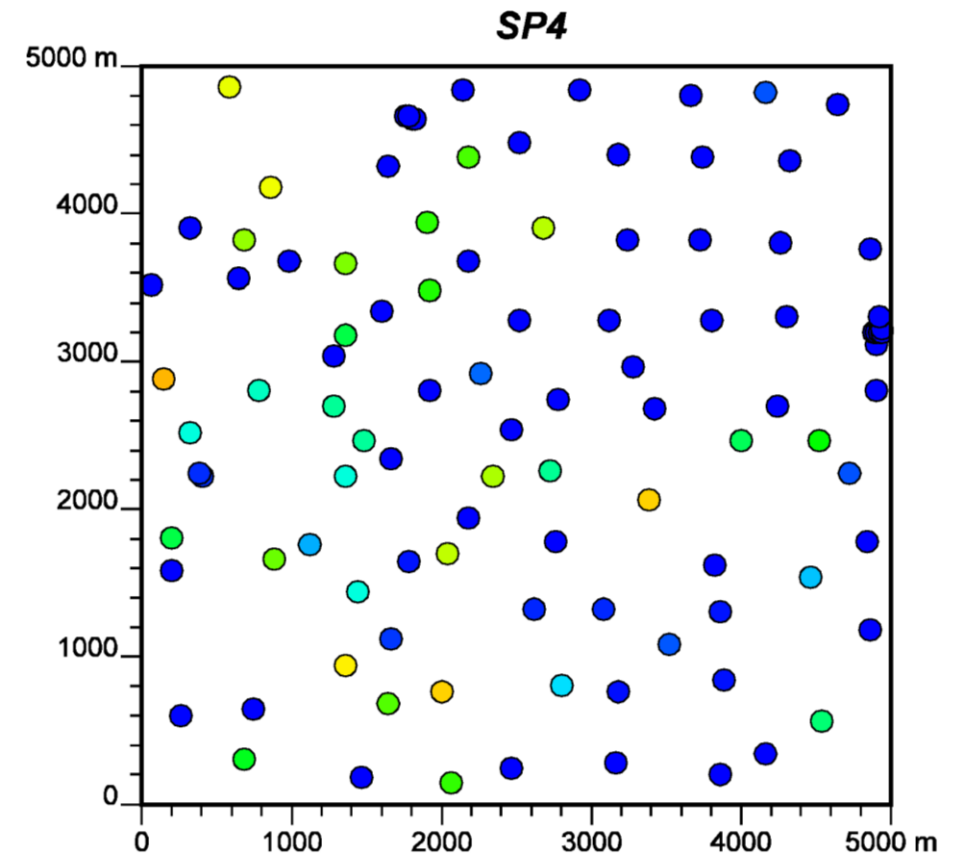
Fonte Imagem: Martinez et al., 2010



# Amostragem Aleatória Estratificada

*É desenvolvida aleatoriamente em estratos, ou seja, subdivisões da região em **células com dimensões fixas**. Cada célula deverá conter um **ponto selecionado aleatoriamente**.*

*Yamamoto e Landim, 2013*



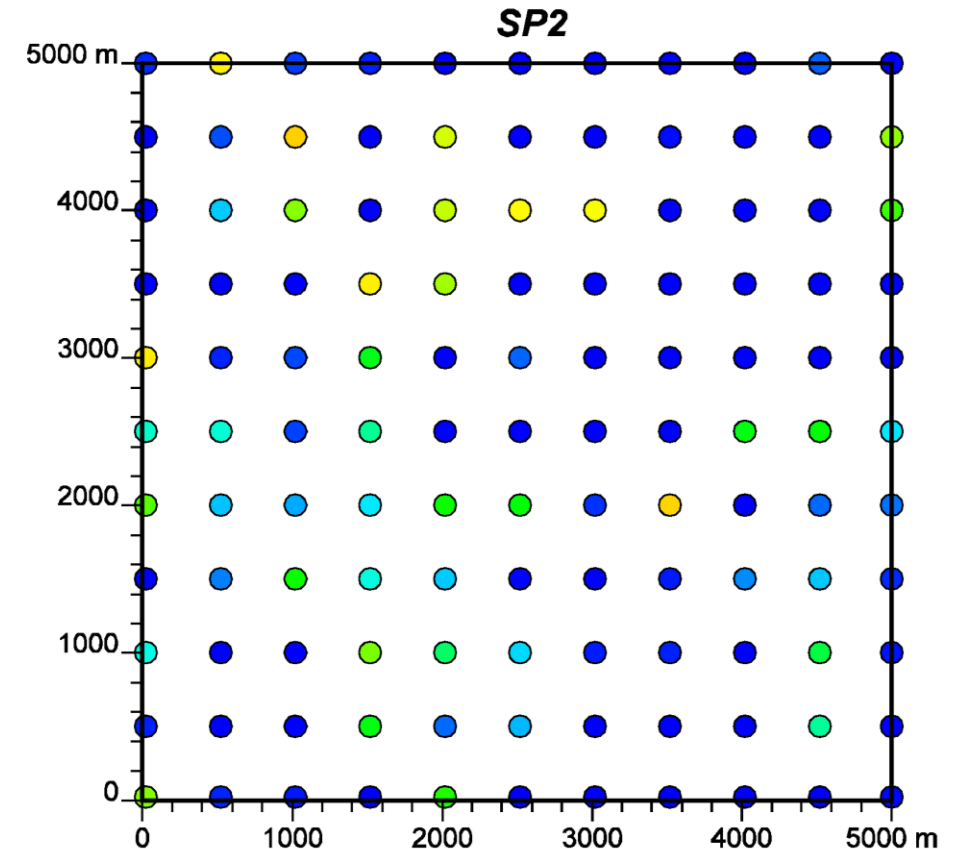
*Fonte Imagem: Martinez et al., 2010*



# Amostragem Sistemática

É desenvolvida sobre os nós de uma **malha regular** definida com base em uma origem escolhida aleatoriamente.

*Yamamoto e Landim, 2013*

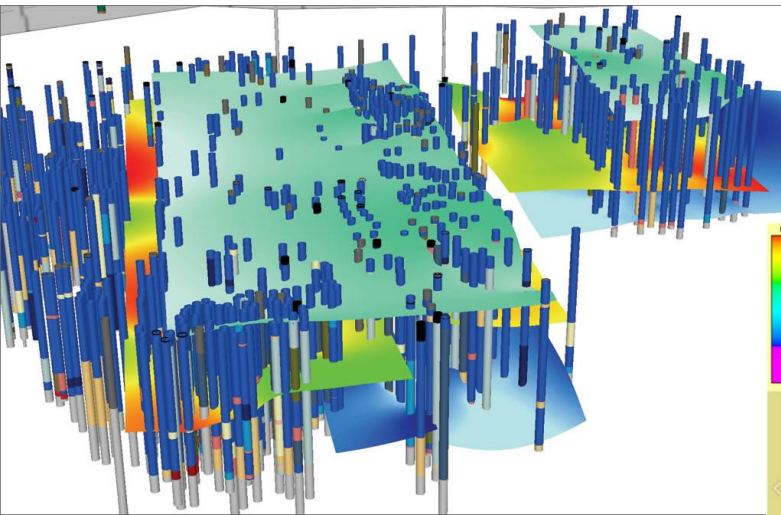


*Fonte Imagem: Martinez et al., 2010*

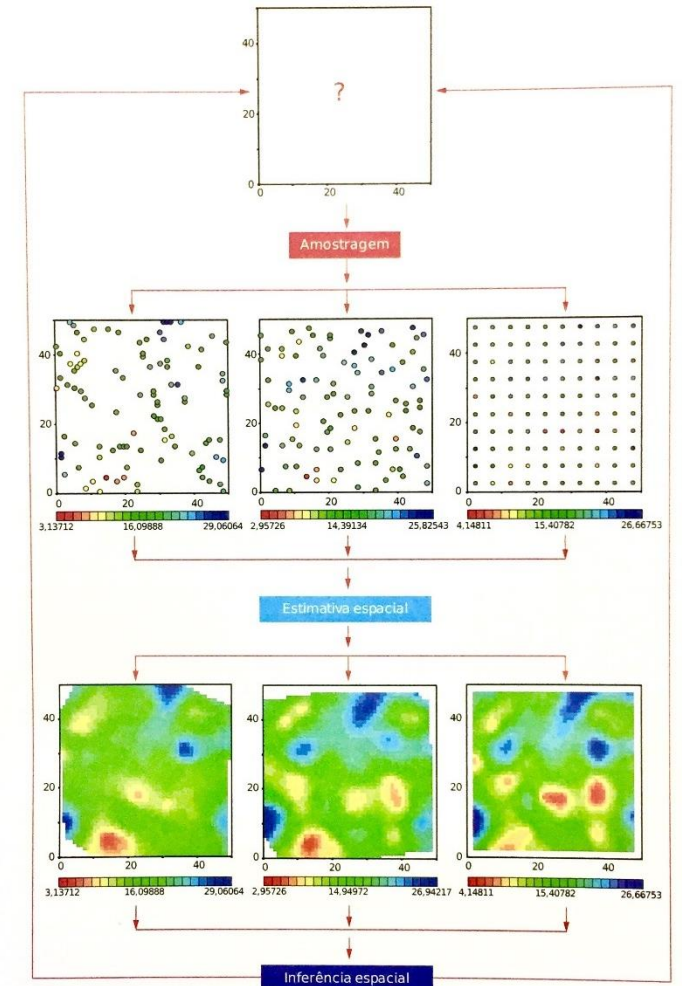
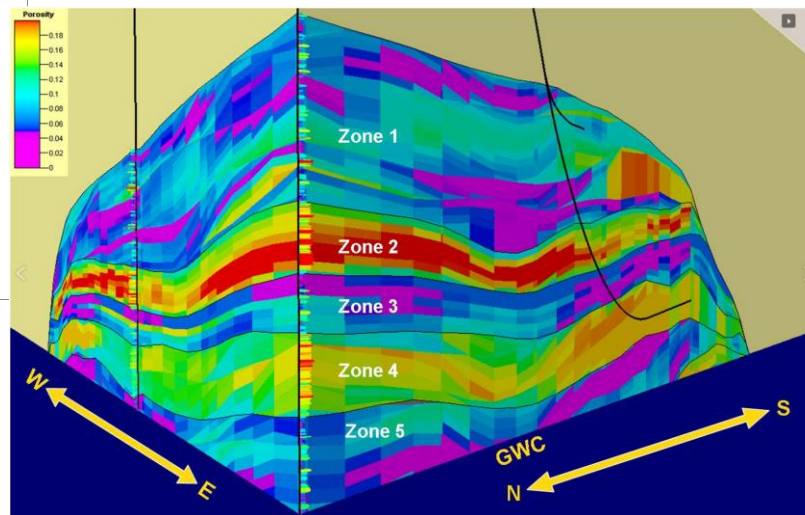


# Inferência Espacial

- **Interpolação ou estimativa:** processo de reprodução das características do fenômeno espacial baseado em pontos amostrais.



<http://www.esri.com/news/arcuser/0312/modeling-the-terrain-below.html>



Fonte Imagem: Yamamoto e Landim, 2013





# Métodos Estimadores

- **Método DETERMINÍSTICO:** chance de ocorrência das variáveis envolvidas no processo é ignorada. Segue uma lei definida que não a lei das probabilidades.
- **Método ESTOCÁSTICO:** chance de ocorrência das variáveis é levada em conta. Conceito de probabilidade é introduzido na formulação do modelo.







# GEOESTATÍSTICA

É uma técnica de análise de dados dentro da estatística espacial, que possui a característica de tratar de variáveis regionalizadas.

Uma subárea da estatística que estuda variáveis regionalizadas (Yamamoto e Landim, 2013).



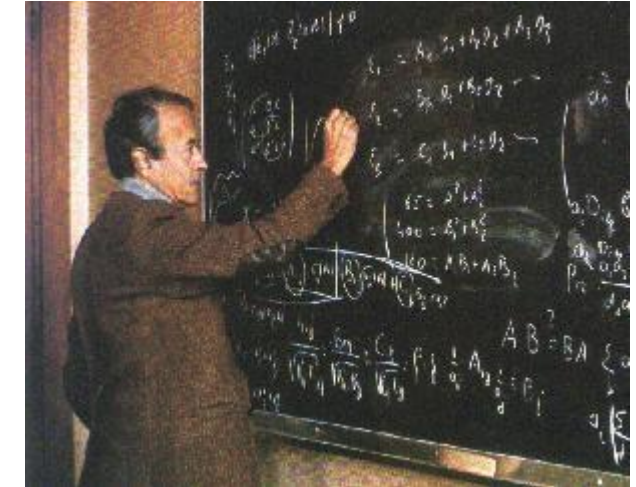
# GEOESTATÍSTICA - Histórico

Década de 1960 → Georges Matheron (Paris) inspirado nos trabalhos de De Wijs (professor na Holanda) e Danie Krige (eng. de minas de ouro na África do sul) gerou publicações que o tornaram o criador da geoestatística.

Década de 1970 → Formalização dos conceitos e difusão na geologia e engenharia de minas.

Década de 1980 → Início da aplicação em outras áreas do conhecimento além da engenharia de minas: Agricultura, Cartografia, Climatologia, Ecologia, etc.

Década de 1990 → Popularização e Crescimento rápido devido ao desenvolvimento da computação.



(Fonte: Juliana Almeida Kolling)



# Estatística Comum

Série 1	Série 2
1	1
7	3
3	5
6	7
2	9
9	8
4	6
8	4
5	2

Série 1  
 $m = 5$   
 $\sigma^2 = 6,67$

Série 2  
 $m = 5$   
 $\sigma^2 = 6,67$

(Fonte: Jorge Kazuo Yamamoto)



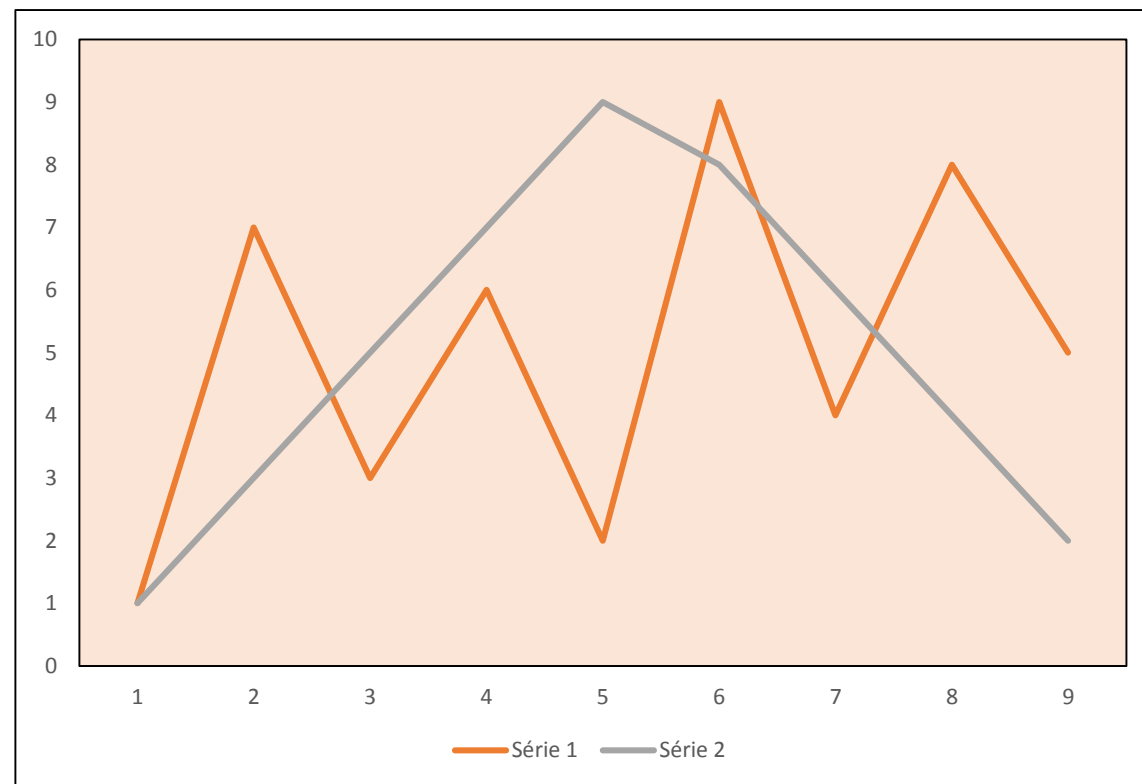


# Estatística Comum

Linha	Série 1	Série 2
1	1	1
2	7	3
3	3	5
4	6	7
5	2	9
6	9	8
7	4	6
8	8	4
9	5	2

Série 1  
 $m = 5$   
 $\sigma^2 = 6,67$

Série 2  
 $m = 5$   
 $\sigma^2 = 6,67$



(Fonte: Jorge Kazuo Yamamoto)



# Variáveis Regionalizadas

## Primeira Lei da Geografia:

"Todas as coisas estão relacionadas entre si, no entanto, as coisas mais próximas estão mais relacionadas que as distantes."

*(Fonte: Tobler W., 1970)*





# Variáveis Regionalizadas

**Variável aleatória:** recebe um certo número de valores de acordo com uma certa distribuição de probabilidades.

**Variável regionalizada:** é uma realização particular de uma variável aleatória em um ponto definido (georreferenciado) na região de interesse.





# Variáveis Regionalizadas

As variáveis regionalizadas apresentam **duplo aspecto contraditório**:

- **Aleatório**: apresenta irregularidades e variação imprevisível de um ponto para outro;
- **Estrutural**: apresenta relação entre os pontos no espaço. (determinístico no espaço).

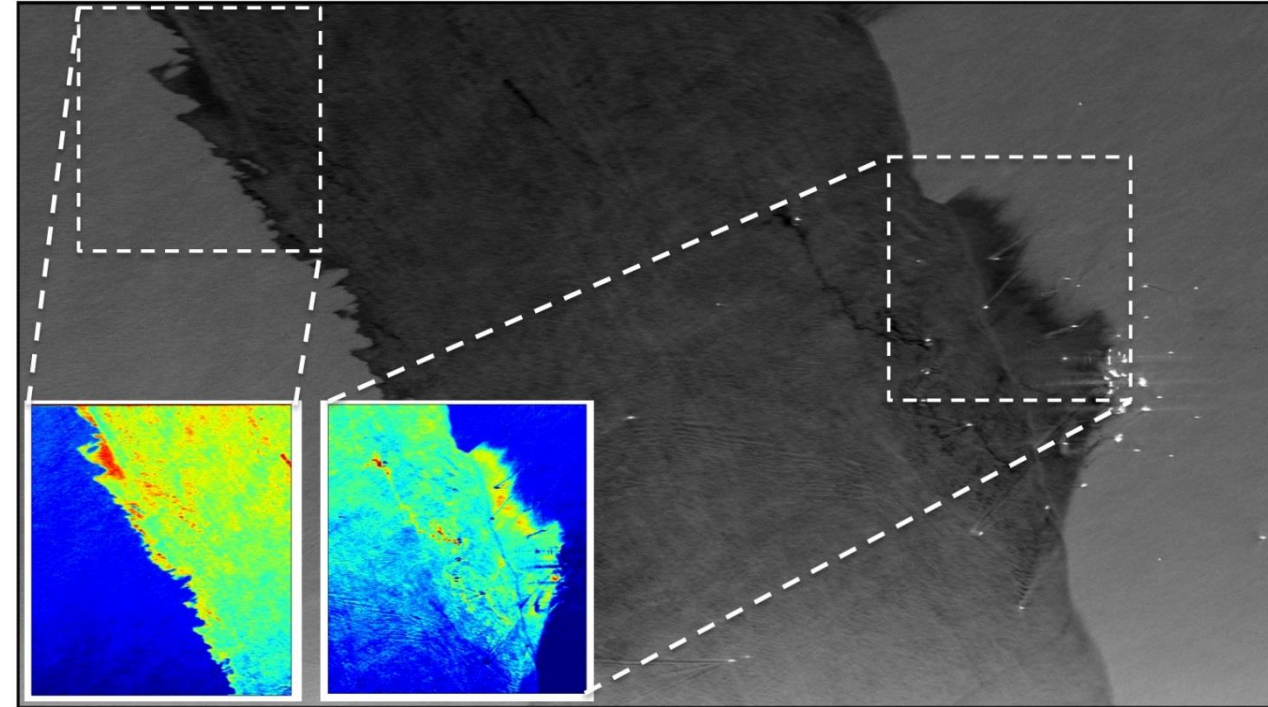
(Fonte: INPE)





# Variáveis Regionalizadas

- **Aleatório:** é impossível de prever a concentração exata de óleo na pluma de derramamento;
- **Estrutural:** teores de concentração em pontos próximos entre si são provavelmente semelhantes.



*Imagem UAVSAR NASA de um derrame de petróleo em Deepwater Horizon, 23/6/2010.*

(Fonte: <http://www.jpl.nasa.gov/news/news.php?release=2012-337>)



# Métodos de Interpolação: Inverso do Quadrado da Distância

Dados pontuais são ponderados durante a interpolação de tal forma que a **influência de um determinado ponto em relação a outro diminui com a distância**. Normalmente este método de geração de superfícies se comporta como um interpolador exato e a **velocidade de processamento dependerá da quantidade de pontos envolvidos**.

$$Z = \frac{\sum_{i=1}^n \frac{Z_i}{h_{ij}^\beta}}{\sum_{i=1}^n \frac{1}{h_{ij}^\beta}}$$

Sem parâmetro de suavização

$$Z = \frac{\sum_{i=1}^n \frac{Z_i}{(h_{ij} + \delta)^\beta}}{\sum_{i=1}^n \frac{1}{(h_{ij} + \delta)^\beta}}$$

Com parâmetro de suavização

Onde:

$Z$  = medida da interpolação;  
 $Z_j$  = vizinho do ponto a ser interpolado;

$h_{ij}$  = medida de distância;

$\beta$  = fator de ponderação;

$\delta$  = parâmetro de suavização.





# Métodos de Interpolação: Krigagem

“É um processo de estimativa de valores de variáveis distribuídas no espaço, com base em valores adjacentes, quando considerados interdependentes pela análise variográfica”.

Variograma mostra dependência espacial → **KRIGAGEM**

Variograma não mostra dependência espacial → **INTERPOLAÇÃO DETERMINÍSTICA**

*(Fonte: Juliana Almeida Kolling)*



# Distribuição de Pontos

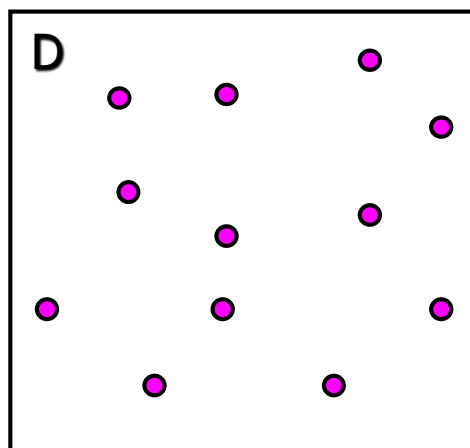
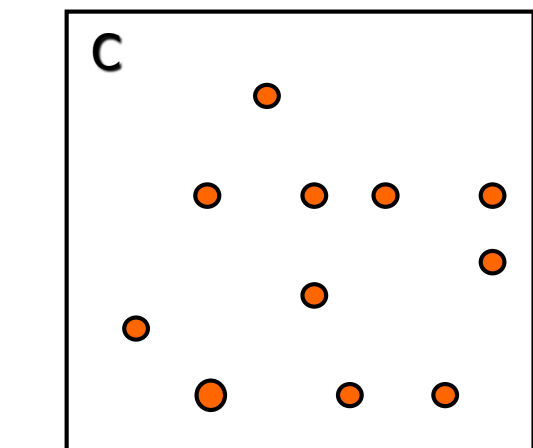
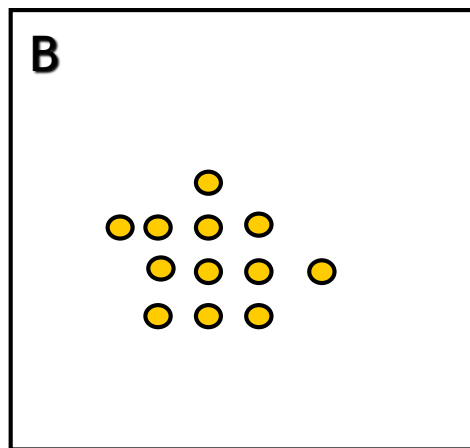
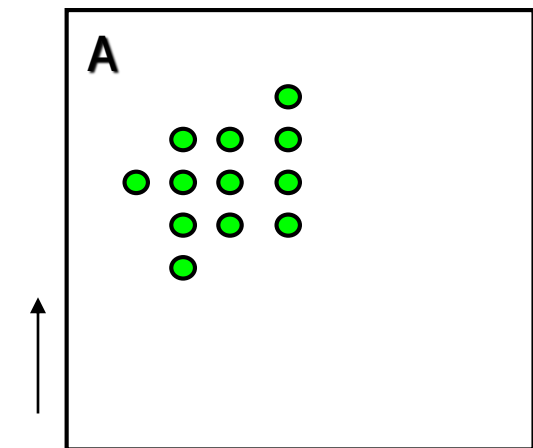
- A distribuição de características pontuais pode ser descrita pela:
  - Frequência
  - Densidade (frequência/área)
  - Centro Geométrico (média das coordenadas)
  - Dispersão Espacial (desvio padrão de cada média)
  - Arranjo Espacial \*
- Com exceção do arranjo espacial, a avaliação das propriedades espaciais pontuais pode ser realizada através de estatísticas descritivas básicas.

*Fonte: Mariana Giannotti, adaptado de DPI/INPE*



# Distribuição de Pontos

(12, 15)



LOCALIZAÇÕES			
A	B	C	D
(2, 7)	(3, 6)	(2,4)	(3,4)
(3, 5)	(4, 4)	(3,10)	(5,2)
(3, 6)	(4, 5)	(4,7)	(5,8)
(3, 7)	(4, 6)	(5,2)	(7,11)
(3, 8)	(5, 4)	(7,4)	(8,5)
(4, 6)	(5, 5)	(7,6)	(8,8)
(4, 7)	(5, 6)	(7,9)	(9,2)
(4, 8)	(5, 7)	(10,2)	(10,8)
(5, 6)	(6, 4)	(11,6)	(12,2)
(5, 7)	(6, 5)	(11,10)	(13,4)
(5, 8)	(6, 6)	(13,4)	(13,6)
(5, 9)	(7, 5)	(13,8)	(13,8)

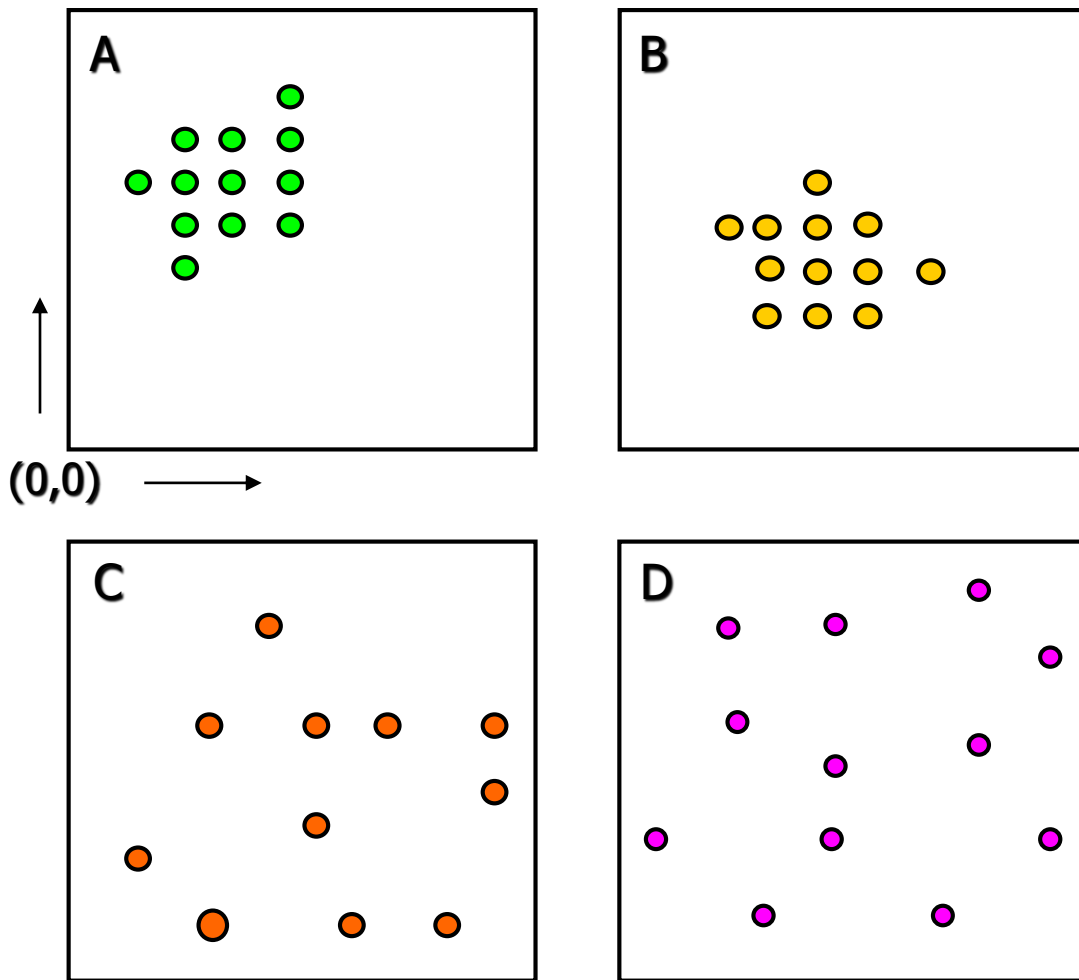
Fonte: Mariana Giannotti,  
adaptado de DPI/INPE



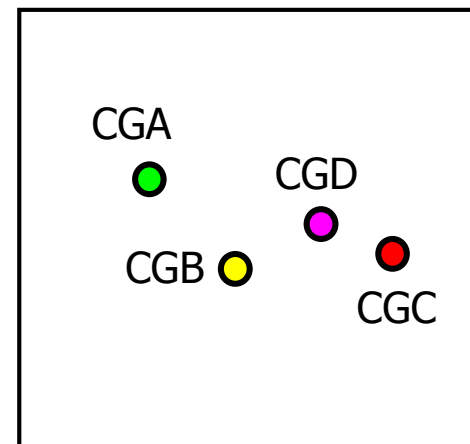


# Distribuição de Pontos

(12, 15)



- Frequência = 12 em A, B, C, D
- Densidade =  $180/12$  em A, B, C, D
- Centro Geométrico



*Nota:* CGA e CGB representam bem a tendência central porque ambas distribuições estão concentradas em torno dos respectivos centros. Por outro lado, CGC e CGD não são bons indicadores para suas respectivas distribuições.

Fonte: Mariana Giannotti, adaptado de DPI/INPE



# Densidade de Pontos

- Requer a adoção de um zoneamento e contagem do número de eventos em cada zona.
- Dificuldade: densidade variará substancialmente a depender da zona adotada (grade, polígono, etc) e quão variável é o tamanho e os atributos associados a esta zona. Alternativas:
  - definir zonas baseadas em uma co-variável de interesse (ex. petróleo: porosidade, concentração, teores etc);
  - usar técnica para suavizar o cálculo da densidade

*Fonte: Smith, Goodchild e Longley, 2013 & DPI/INPE*





# Densidade de Pontos

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
						X	X	X			X		X						
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20

Alternativa para o cálculo da densidade, basear-se em técnicas desenvolvidas para análise estatística univariada.

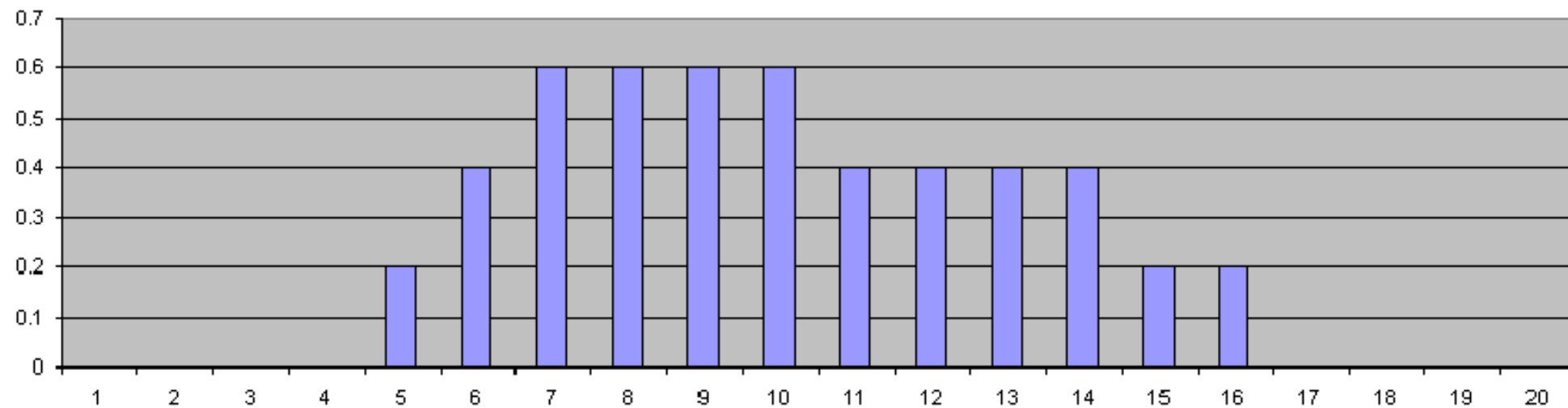
- Pontos 7, 8, 9, 12 e 14 ao longo de uma linha. Densidade ao longo da linha:  $5/20 = 0,25$  para cada seção.
- Se dividido em duas metades, novas densidades.
- O tamanho da linha que partimos tem um efeito importante. Como isso pode ser arbitrário, algum método para tirar a dependência do tamanho da linha é desejável.
- Se dividirmos em “pedaços” discretos, quebras na densidade ocorrerão e teremos áreas com valor zero (fronteiras), o que é indesejável.
- Se tivermos muitas partes todas as seções terão valor 1 ou 0, voltando onde começamos.

Fonte: Mariana Giannotti

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
						X	X	X			X		X						
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20

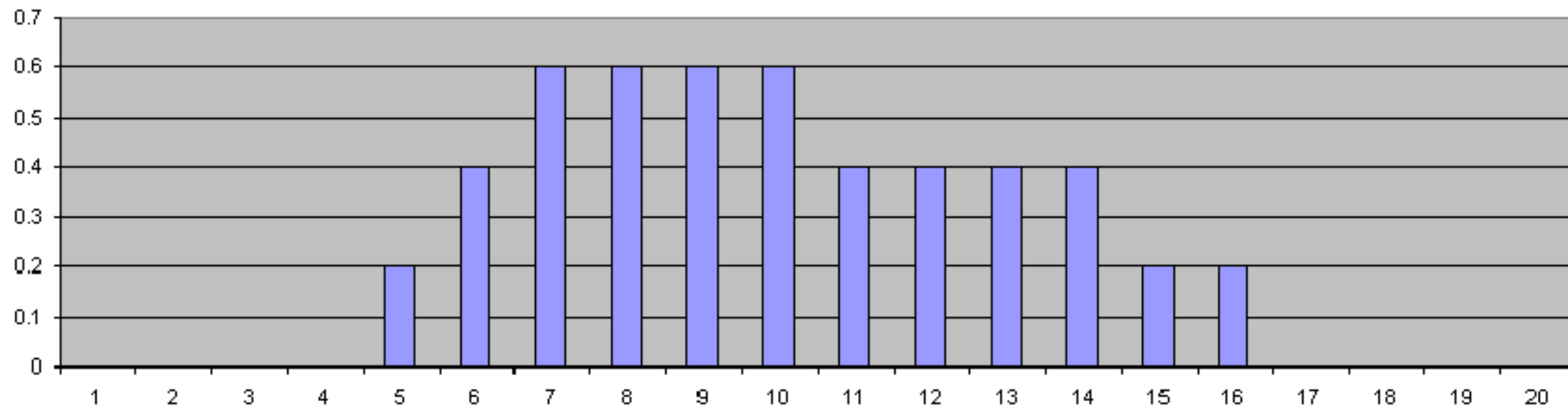
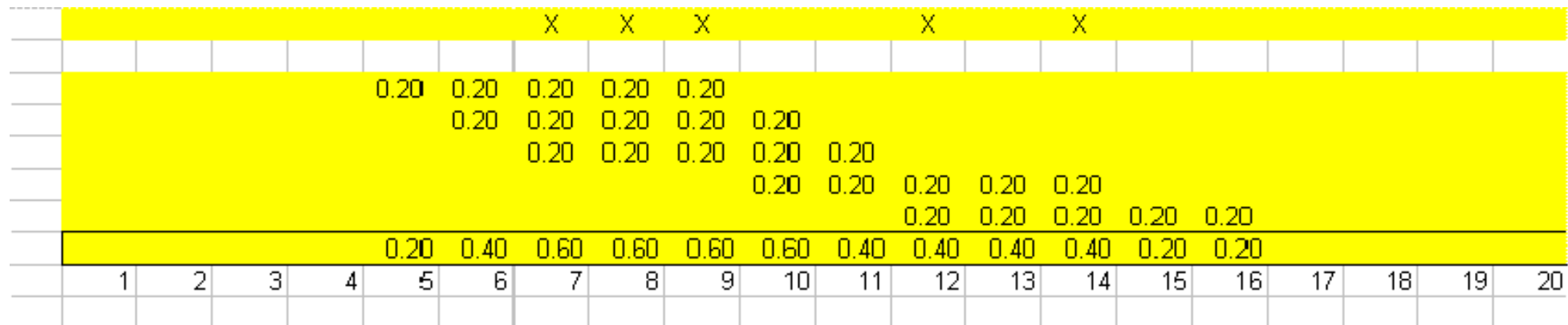
- Tratar cada ponto do conjunto original como se estivessem espalhados em um intervalo, adicionar zonas de sobreposição, checar se o total equivale ao valor original.

						X	X	X			X		X						
				0.20	0.20	0.20	0.20	0.20											
					0.20	0.20	0.20	0.20	0.20										
						0.20	0.20	0.20	0.20	0.20									
								0.20	0.20	0.20	0.20	0.20							
										0.20	0.20	0.20	0.20	0.20					
												0.20	0.20	0.20	0.20				
														0.20	0.20				
																0.20	0.20		
																		0.20	0.20
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20



Fonte: Mariana Giannotti

- Não há valores de densidade nas bordas.
- Valores de densidade “pulam” abruptamente de um valor para outro.
- Valores eventualmente espalhados entorno dos pontos iniciais, quando o desejável seria ter um peso maior para os valores do ponto central.

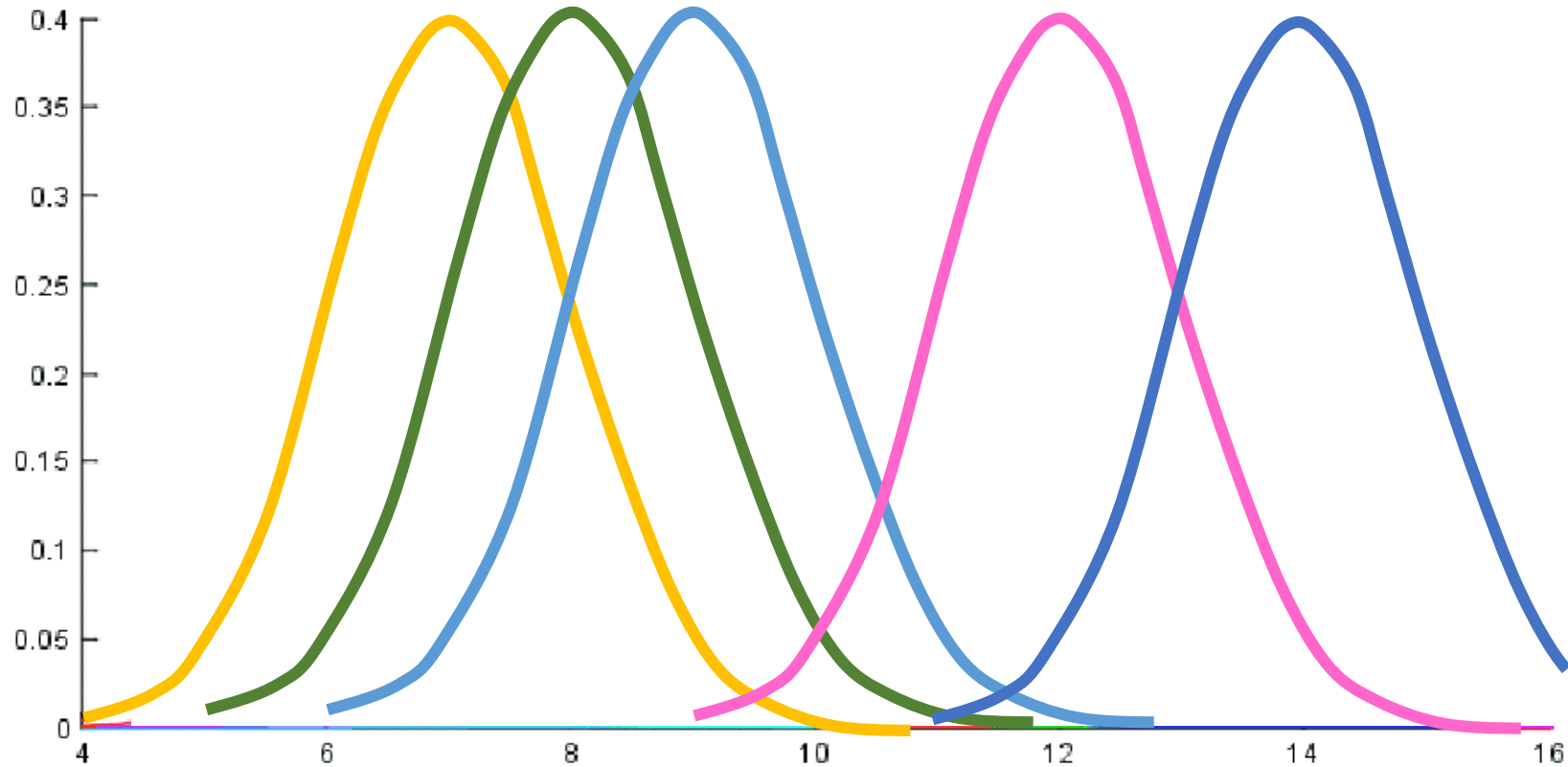


Fonte: Mariana Giannotti





# Kernel

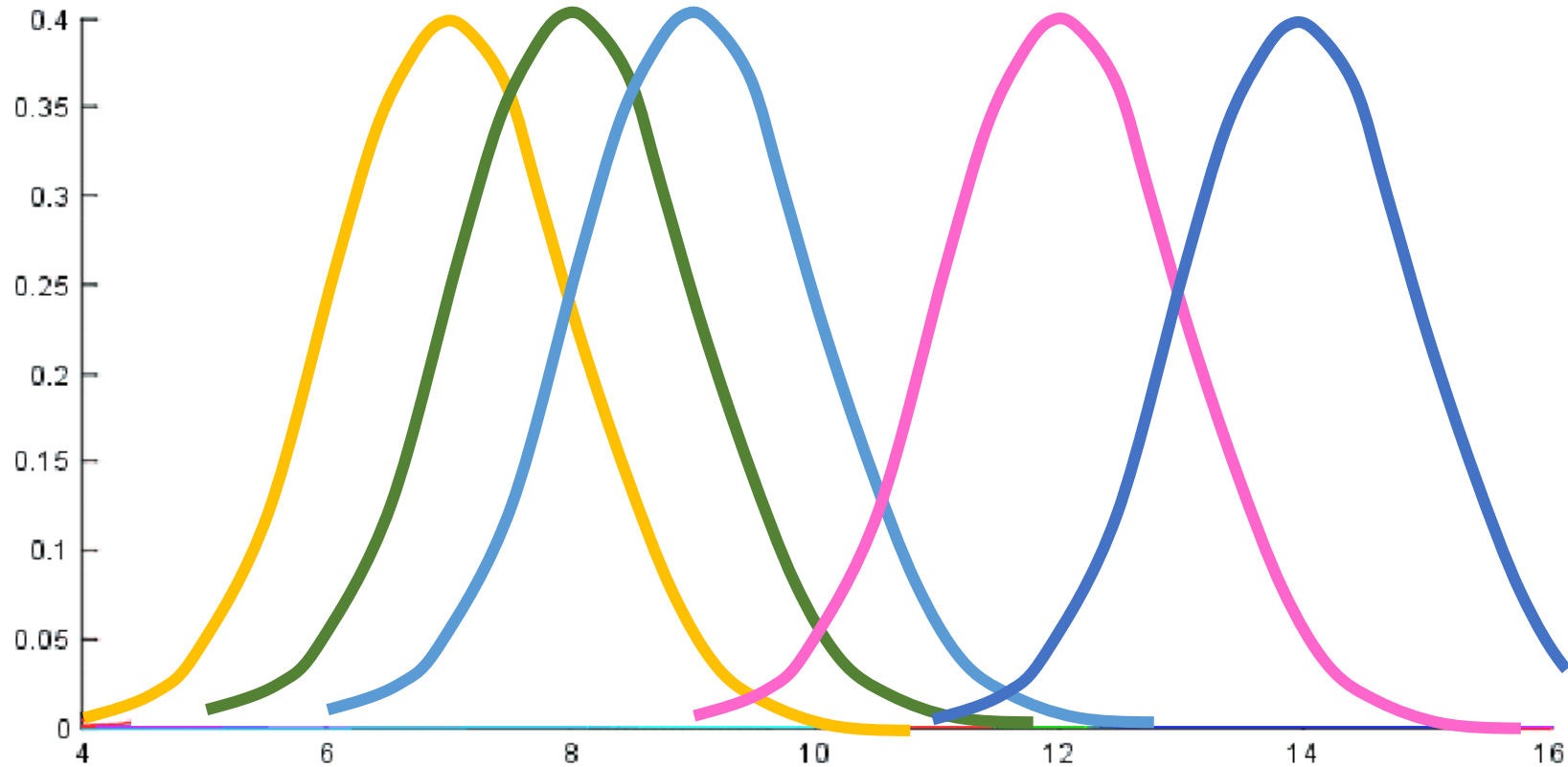


*Fonte: Smith, Goodchild e Longley, 2013*

Densidades acumuladas, suavizadas, univariadas de Kernel, baseadas na distribuição normal.



# Kernel

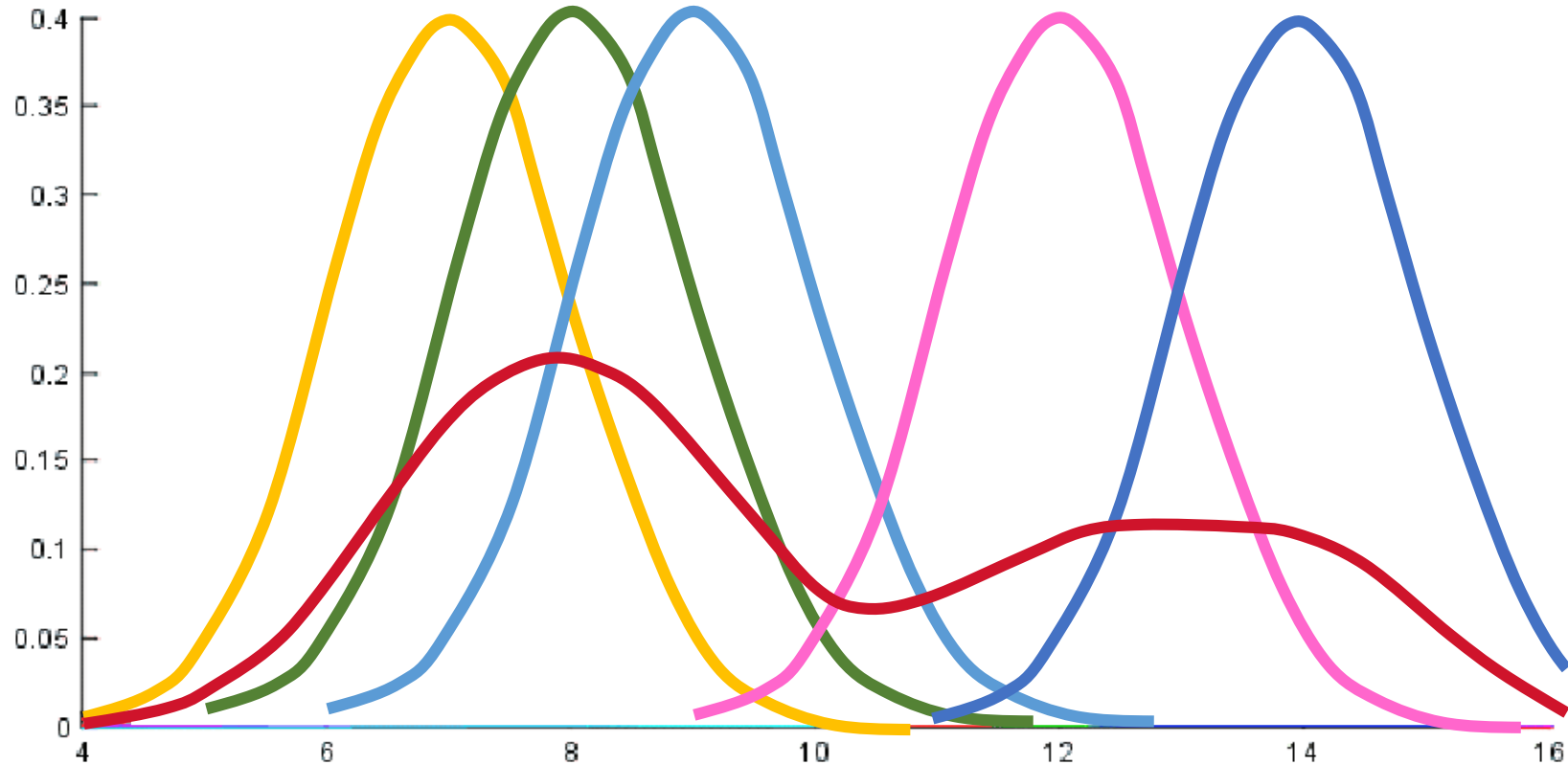


*Fonte: Smith, Goodchild e Longley, 2013*

Para cada ponto provemos uma curva de distribuição normal, com valor central (média) no ponto e um espalhamento médio para os dois lados (desvio padrão).



# Kernel

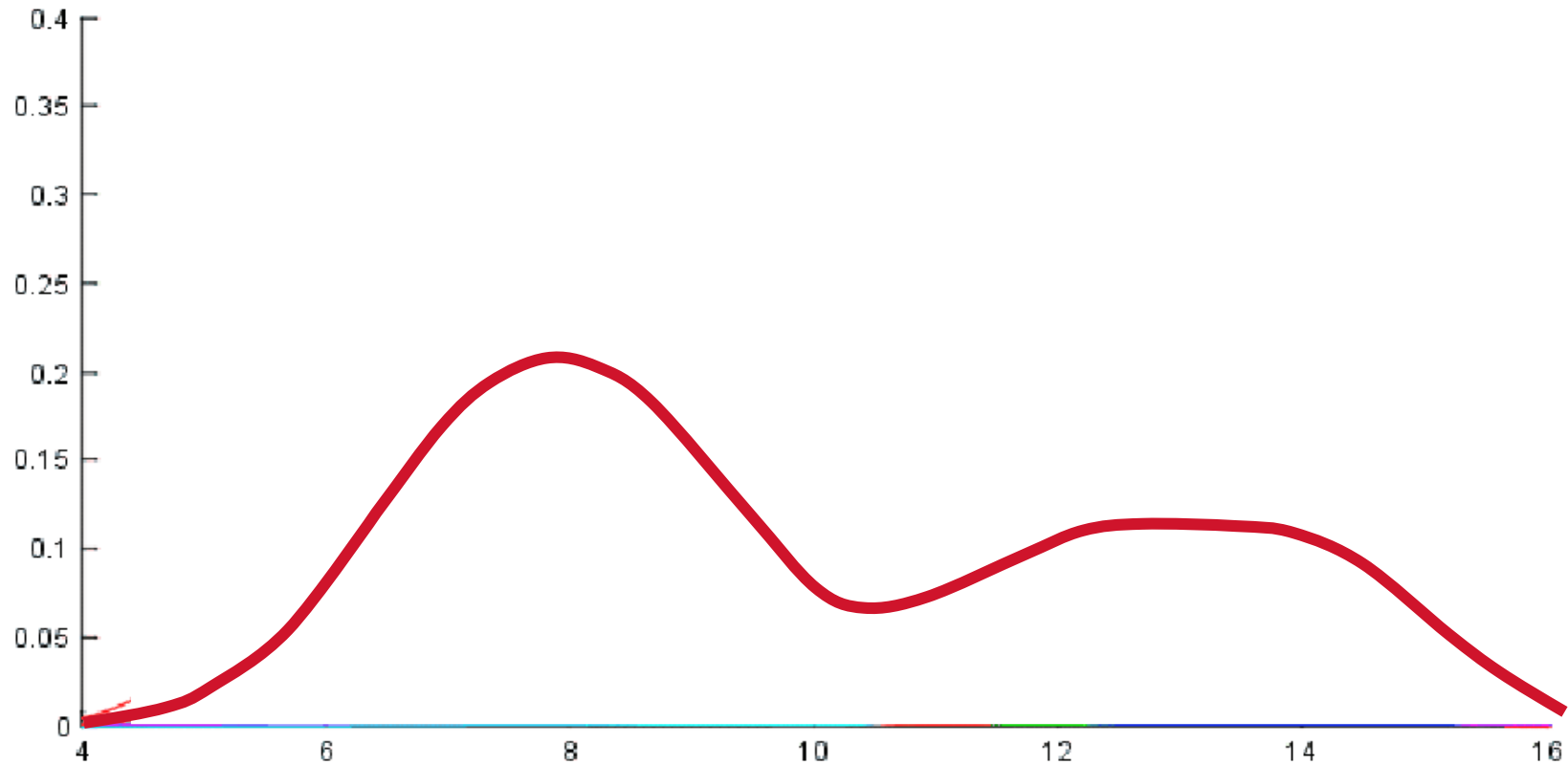


*Fonte: Smith, Goodchild e Longley, 2013*

Áreas embaixo das curvas podem ser “juntadas” para obter uma curva acumulativa com dois picos e, então, dividir esta curva por 5 para ajustar a área da curva para 1: densidades de probabilidade.



# Kernel

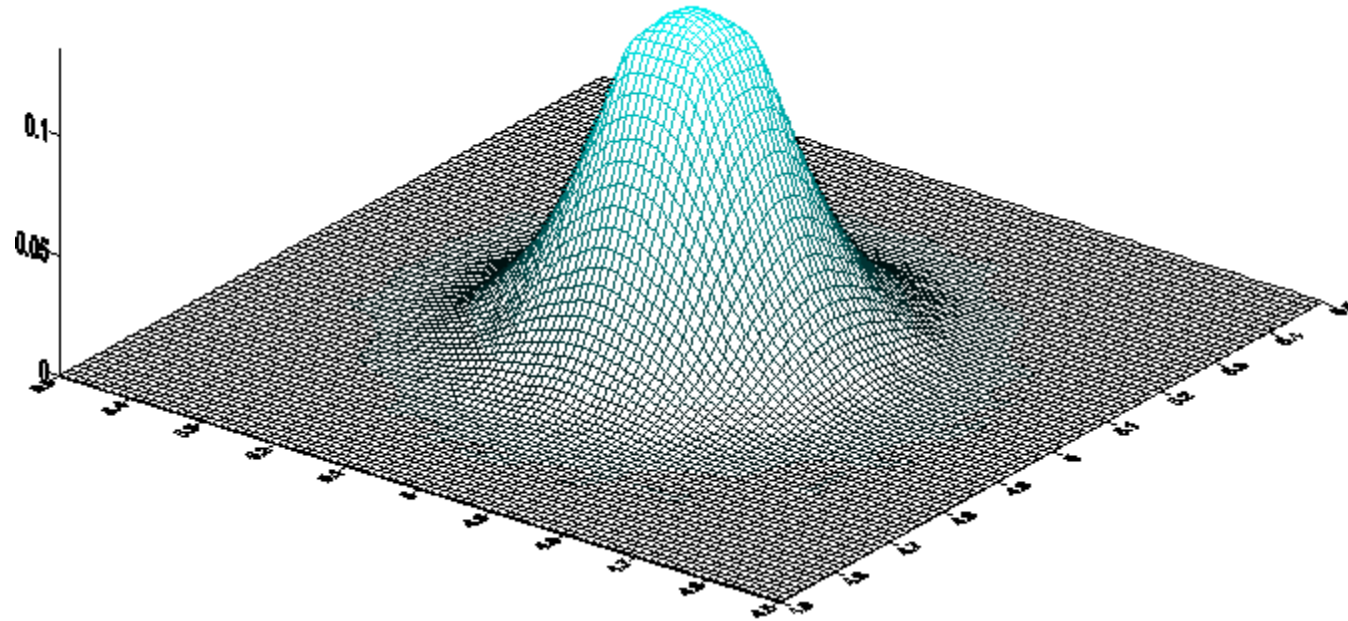


*Fonte: Smith, Goodchild e Longley, 2013*

Valores de densidade para cada posição ao longo da linha, com transições suaves entre os valores.



# Kernel

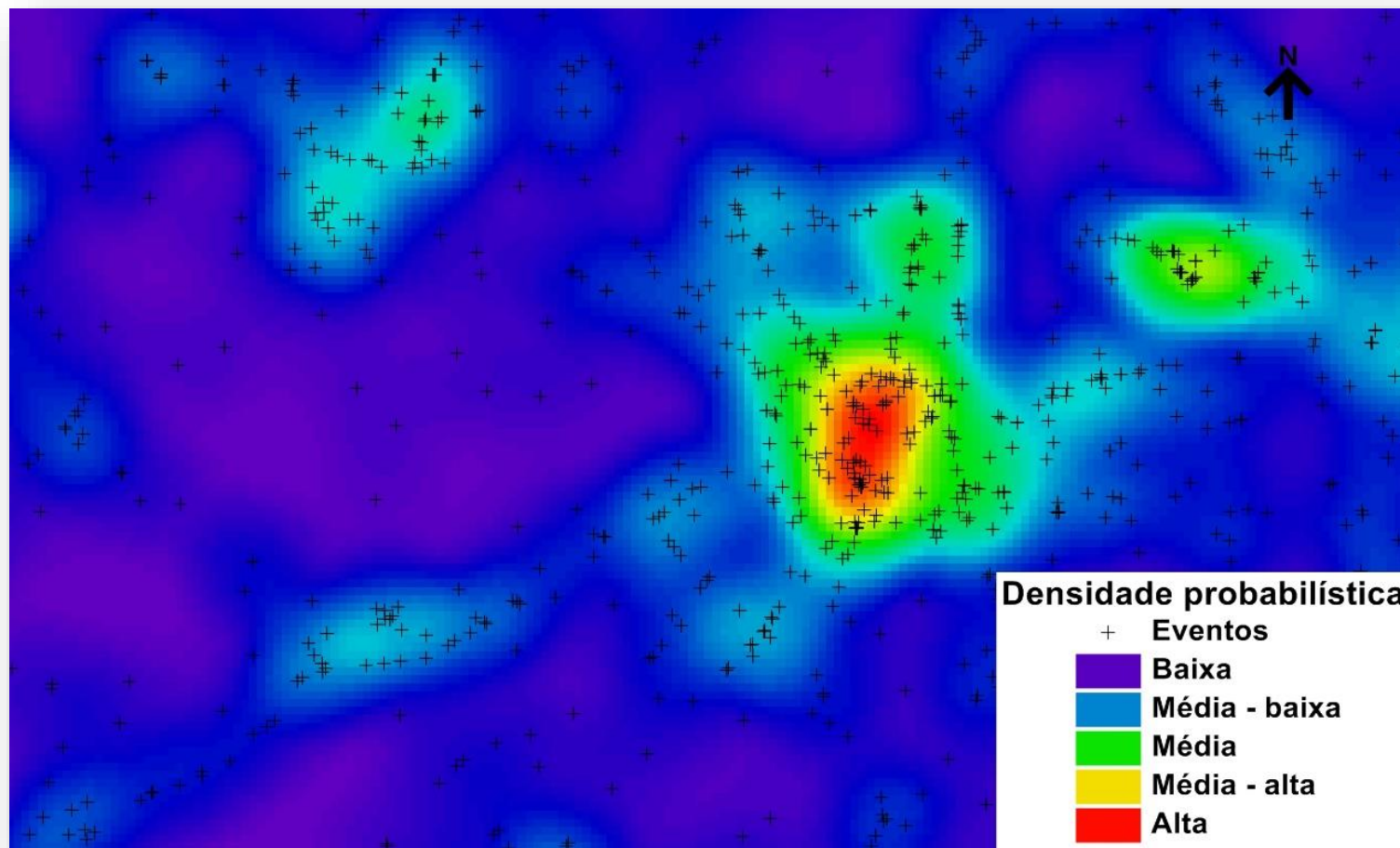


Em 2D: superfície de densidade de probabilidade.





# Kernel

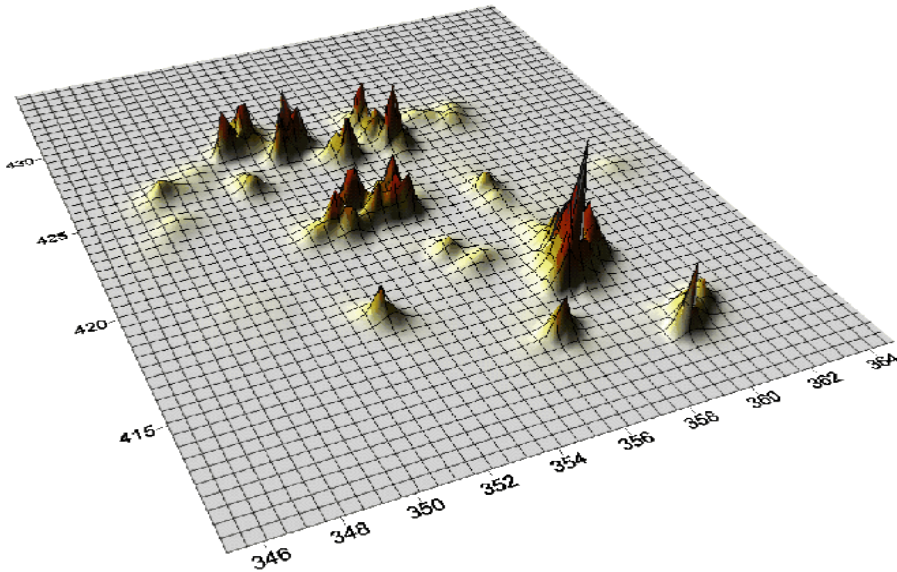




# Kernel

Os valores do grid podem ser fornecidos em:

- Densidades relativas.
  - Ajustadas ao tamanho da grade dando uma noção de número de eventos por área em  $m^2$  ... (usual em SIGs)
- Densidades absolutas.
  - Eventos por célula da grade, não ajustados ao tamanho da célula (o número total em todas as células deve ser igual ao número de eventos usados na análise).
- Probabilidades
  - Segundo caso dividido pelo total de eventos.



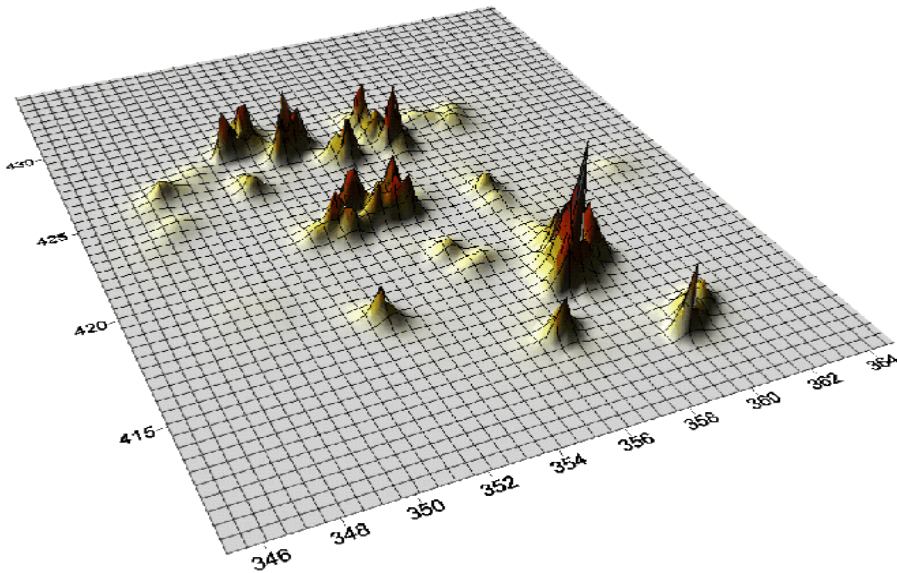
Fonte: Smith, Goodchild e Longley, 2013



# Kernel

Como o kernel para um ponto de um evento particular contribui para o valor em um ponto da grade, ele depende de:

- Tipo de função Kernel
- Parâmetro ( $k$ )
- Bandwith ( $h$ )
  - Bandas mais largas espalham a influência do evento pontual para uma distância maior, mas também há maior probabilidade de ter efeitos de borda.
- Weighted (or not)



Fonte: Smith, Goodchild e Longley, 2013



# Vizinho mais Próximo

- O método do vizinho mais próximo estima a função de distribuição cumulativa ( $\hat{G}(w)$ ) baseado nas distâncias entre eventos em uma região de análise.

- Pode ser estimado empiricamente por: 
$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n}$$

onde:

- $w$  distância de entrada
  - $w_i$  distância entre eventos
  - $n$  número de eventos
- A plotagem dos resultados de  $\hat{G}(w)$  em relação as distâncias, pode ser utilizado como um **método exploratório** para verificar se existe evidência de interação entre os eventos.

Fonte: Slide adaptado de DPI/INPE