



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÃO



# CURSO 4

Aplicação de Métodos Estatísticos Espaciais para Análise de Dados Geográficos.

AULA 1 - Análise de Padrão de Distribuição de Pontos.

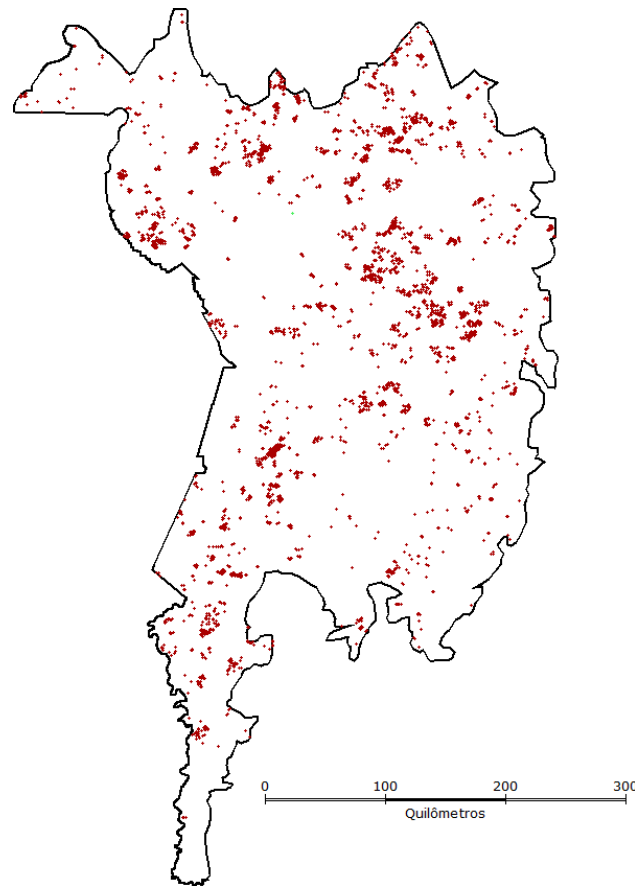
Dr. Eduardo Camargo  
eduardo.camargo@inpe.br

Dr. Carlos Felgueiras  
carlos.felgueiras@inpe.br

- Introdução
- Caracterização da Distribuição de Pontos
- Estimador de Kernel
- Método do Vizinho Mais Próximo
- Método do Vizinho Mais Próximo com Simulação
- Função L
- Função L com Simulação
- Exemplos Práticos com o Sistema TerraView 4.2.2 e SPRING 5.5.6

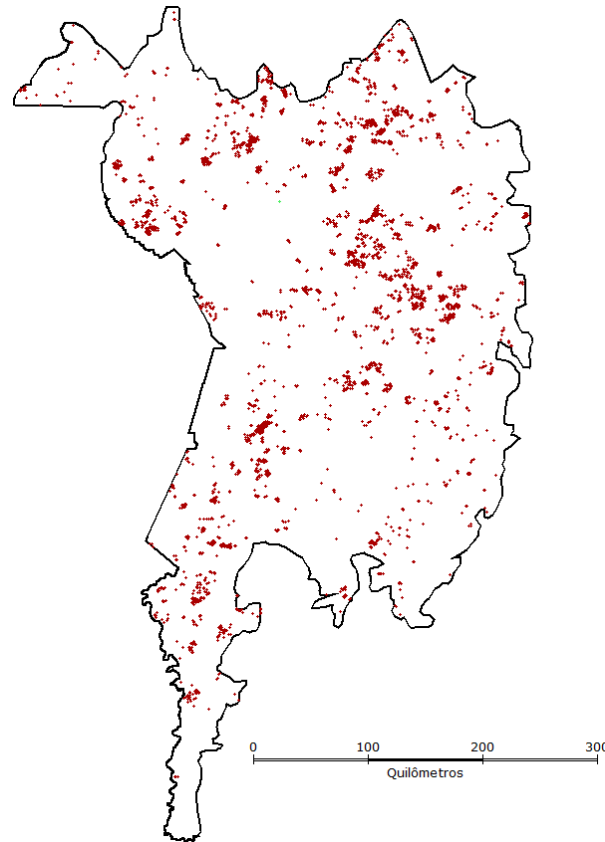
# Introdução

A distribuição de pontos é representada por um conjunto de localizações, contidos em uma área de estudo, que estão associados a eventos de interesses.

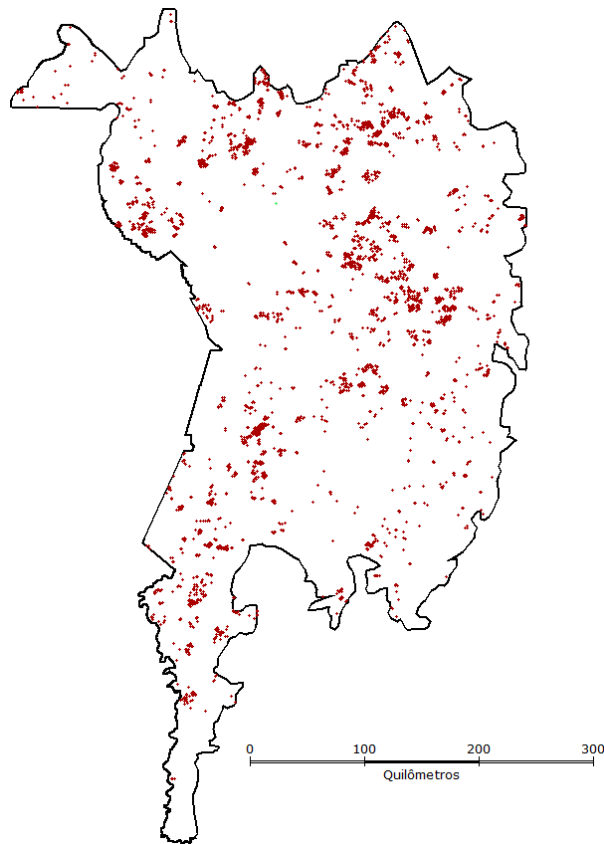


A distribuição de pontos pode corresponder a dados como:

- índices de mortalidades;
- ocorrências de doenças;
- localização de espécie vegetais, focos de queimadas, etc.



Na análise de distribuição de pontos, em geral o que se deseja é determinar se os eventos observados exibem algum padrão sistemático sobre a região de estudo.



## Exemplo:

A distribuição dos casos de uma doença formam um padrão no espaço?

Existe associação com alguma fonte de poluição?

Evidência de contágio?

## **Distribuições pontuais tem as seguintes características:**

- as localizações não estão associadas a valores, mas apenas a ocorrência dos eventos.
- dimensão das medidas é zero.
- medidas válidas: o número de ocorrências e as localizações geográficas.
- a região que contém os eventos não é uma medida válida apesar de ocupar espaço.
- entidades geográficas representadas como pontos no mapa são considerados de mesma qualidade.

## **Características básicas da distribuição de pontos:**

- Frequência
- Densidade
- Centro Geométrico
- Dispersão Espacial

**Frequência:** número de eventos que ocorrem no mapa.

*Nota: a comparação de dois padrões de pontos pode ser enganosa, se ambos estão contidos em áreas distintas. Neste caso, é aconselhável compará-los pela densidade.*

**Densidade:** Frequência / Área

**Centro Geométrico e Dispersão Espacial:** são medidas que caracterizam as propriedades geográfica de um padrão de pontos.

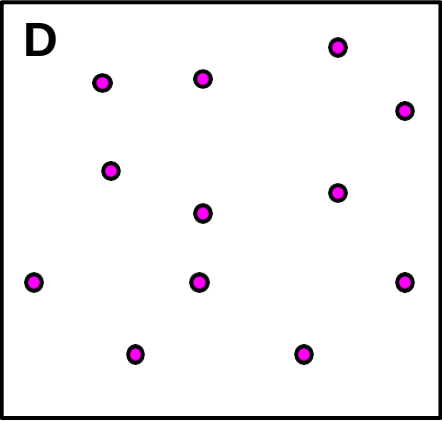
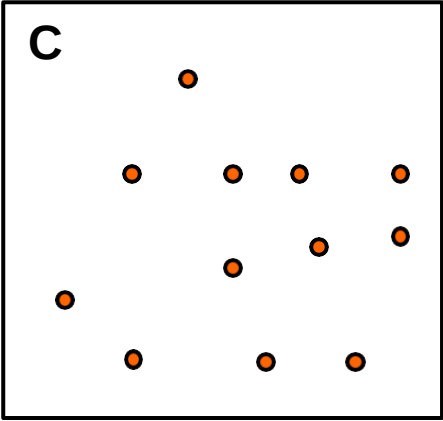
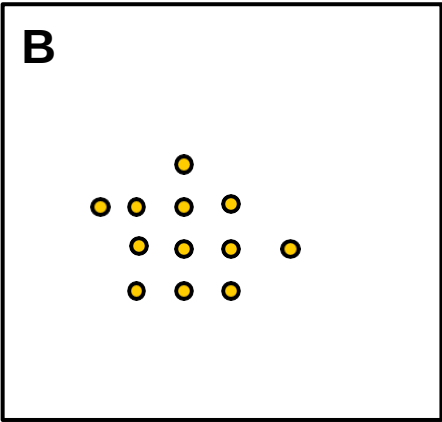
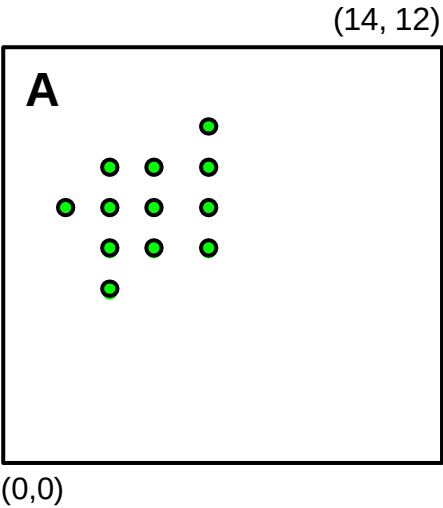
**Centro Geométrico:** média das coordenadas de localização X e Y.

**Dispersão:** desvio padrão de cada média (X e Y).



# Caracterização da Distribuição de Pontos

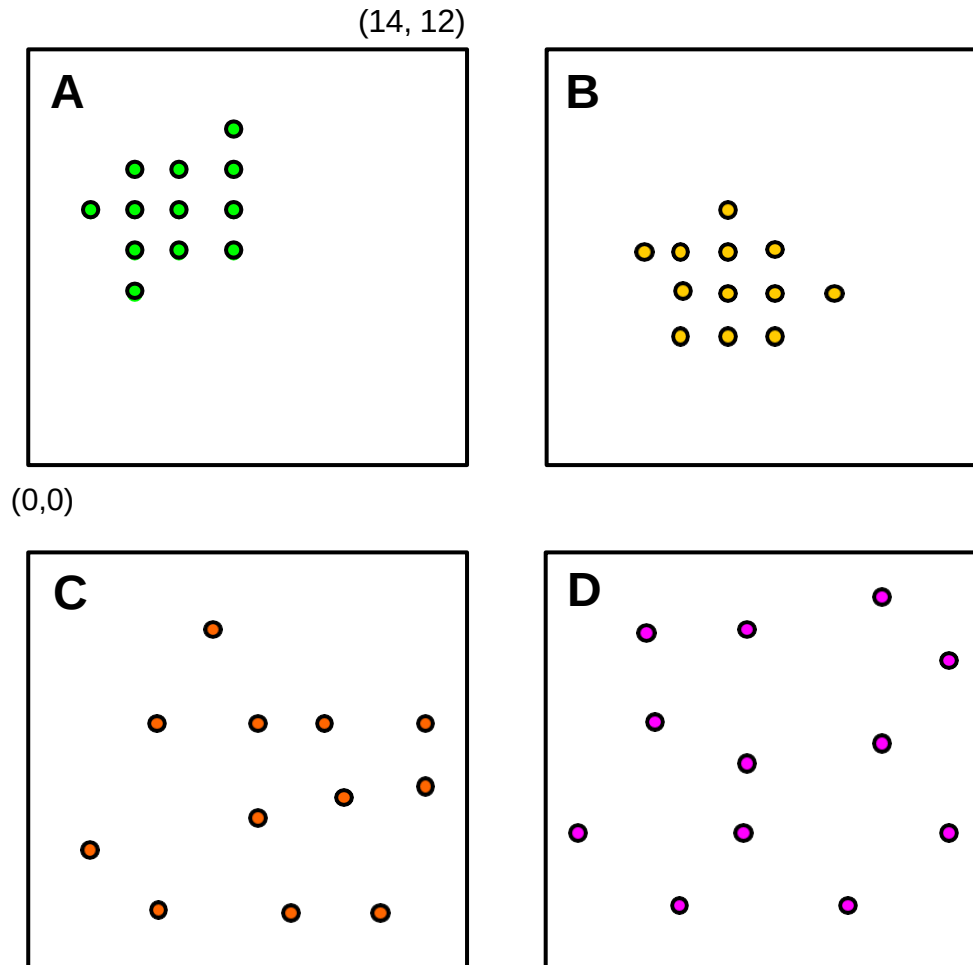
Considere os padrões de pontos **A**, **B**, **C** e **D**.



| 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) |

# Caracterização da Distribuição de Pontos

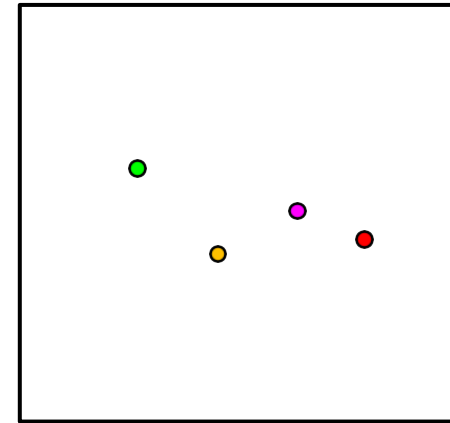
Padrões de pontos **A**, **B**, **C** e **D**.



**Frequência** = 12 em **A**, **B**, **C** e **D**

**Densidade** = 12/168 em **A**, **B**, **C** e **D**

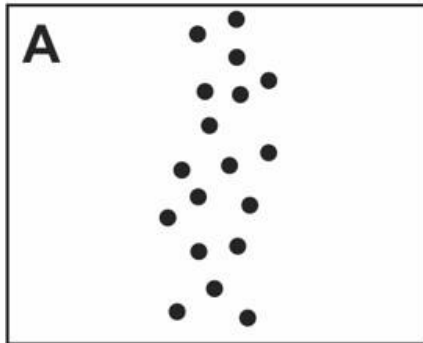
**Centro Geométrico (CG)**



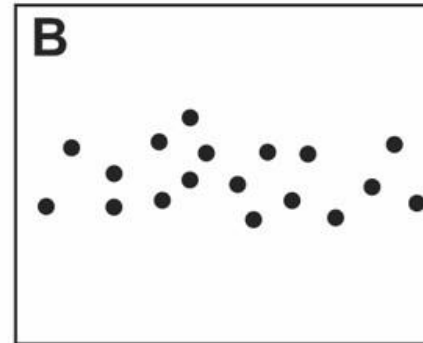
**CGA** e **CGB** representam bem a tendência central, porque ambas distribuições estão concentradas em torno dos respectivos centros geométricos.

**CGC** e **CGD** não são bons indicadores para suas respectivas distribuições.

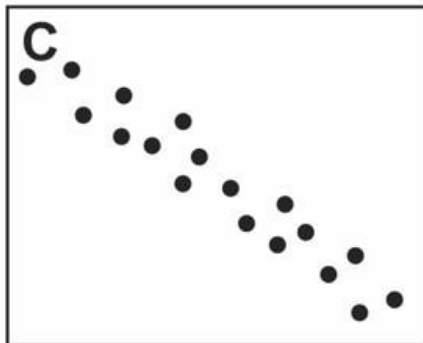
## Dispersão espacial



$$\delta_x < \delta_y$$

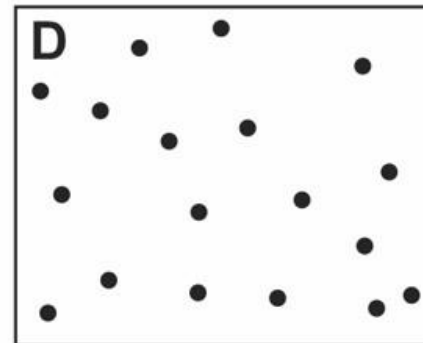


$$\delta_x > \delta_y$$



$$\delta_x \sim \delta_y$$

$\rho$  (significante)



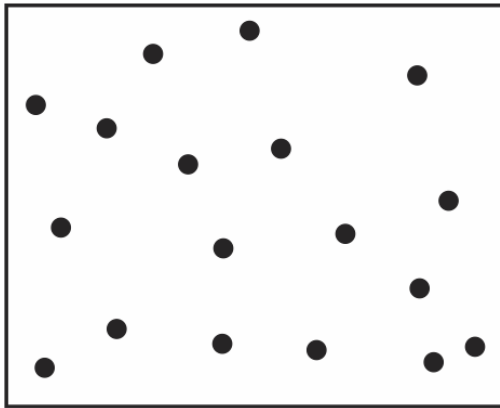
$$\delta_x \sim \delta_y$$

$\rho$  (não significante)

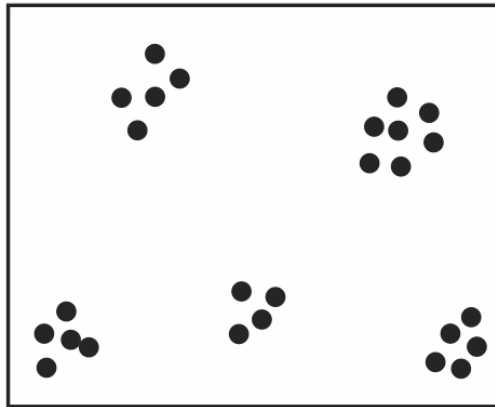
Uma característica importante de um padrão espacial de pontos é a localização dos eventos e a relação entre eles (**arranjos espaciais**).

## Objetivo:

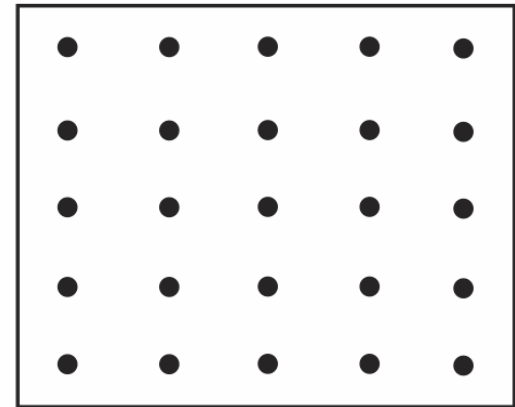
verificar se os eventos observados apresentam algum tipo de padrão sistemático, ao invés de estarem distribuídos aleatoriamente.



aleatório



agrupado



regular

Na prática o que se deseja é detectar padrões de aglomerados espaciais.

**Processos de análise de pontos podem ser descritos em termos de:**

## **Efeitos de Primeira Ordem**

- considerados globais ou de grande escala.
- correspondem a variações no valor médio do processo.
- neste caso estamos interessados na intensidade do processo ( $n^{\circ}$  eventos / unidade de área).

## **Efeitos de Segunda Ordem**

- denominados locais ou de pequena escala.
- representam a dependência espacial no processo.

**Limitação:** a maior parte das técnicas de análise de distribuição de pontos supõe **comportamento isotrópico** do fenômeno investigado.

## **Técnicas a serem abordadas**

### **Efeitos de Primeira Ordem**

Estimador de Kernel

### **Efeitos de Segunda Ordem**

Função do Vizinho mais Próximo

Visa obter uma estimativa suavizada de uma densidade de probabilidade a partir de uma amostra de dados observados.

Baseia-se na contagem de eventos dentro de um dado raio  $r$  para estimar a densidade de eventos associada a cada ponto de uma região de estudo.

Essa contagem é ponderada pela distância de cada evento ao ponto de referência através de uma função de suavização chamada função de Kernel.

Segundo (Bailey & Gatrell, 1995):

$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{1}{\tau^2} I\left(\frac{(s - s_i)}{\tau}\right)$$

em que:

$I$  : é uma função escolhida de forma adequada para construir uma superfície contínua sobre os eventos.

$\tau$  : é denominado largura de faixa (ou largura de banda). Controla o amaciamento da superfície gerada.

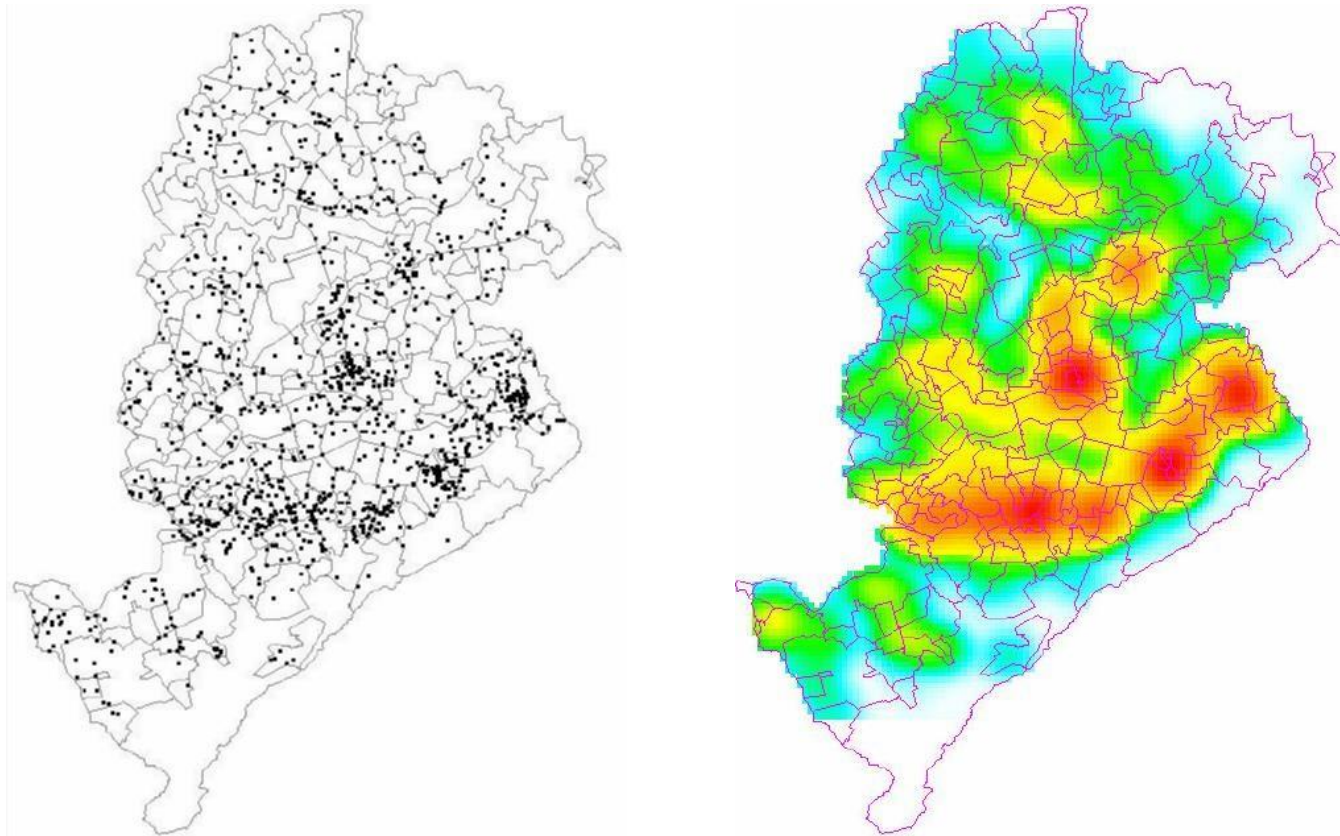
$S$  : representa uma localização qualquer na área de estudo a ser estimada e  $S_i$  são as localizações dos eventos observados.

$n$  : representa o número de eventos dentro da largura de faixa ( $\tau$ ).



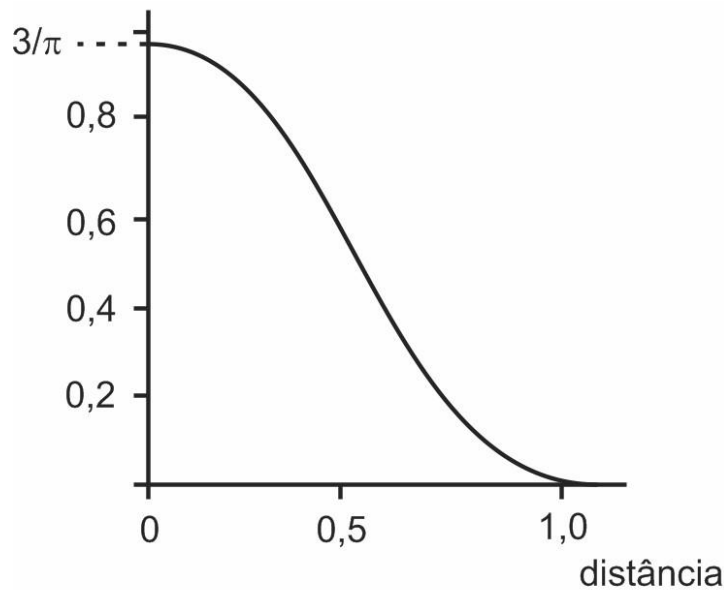
## Exemplo:

Roubo a taxi em Belo Horizonte de 1995 a 2001 (Fonte: PMMG)

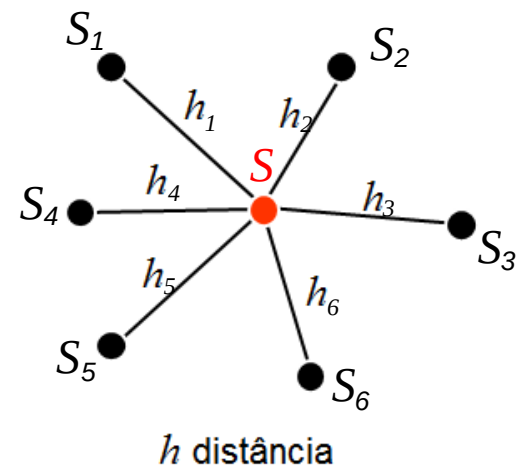


No estimador de Kernel  $\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{1}{\tau^2} I\left(\frac{(s - s_i)}{\tau}\right)$

uma função muito utilizada para  $I$  é:  $I(h) = \frac{3}{\pi} (1 - h^2)^2$



Kernel Quártico



- eventos observados
- localização a ser estimada

Substituindo  $I(h) = \frac{3}{\pi} (1 - h^2)^2$  na equação do estimador

de Kernel  $\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{1}{\tau^2} I\left(\frac{(s - s_i)}{\tau}\right)$ , tem-se:

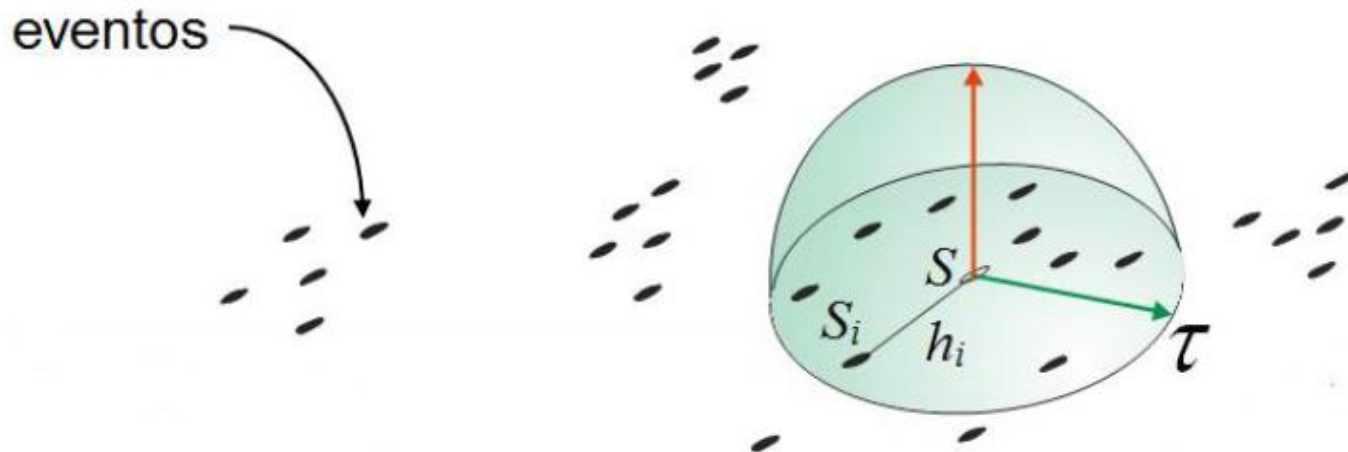
o estimador do Kernel Quártico:  $\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi \tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2$

em que:

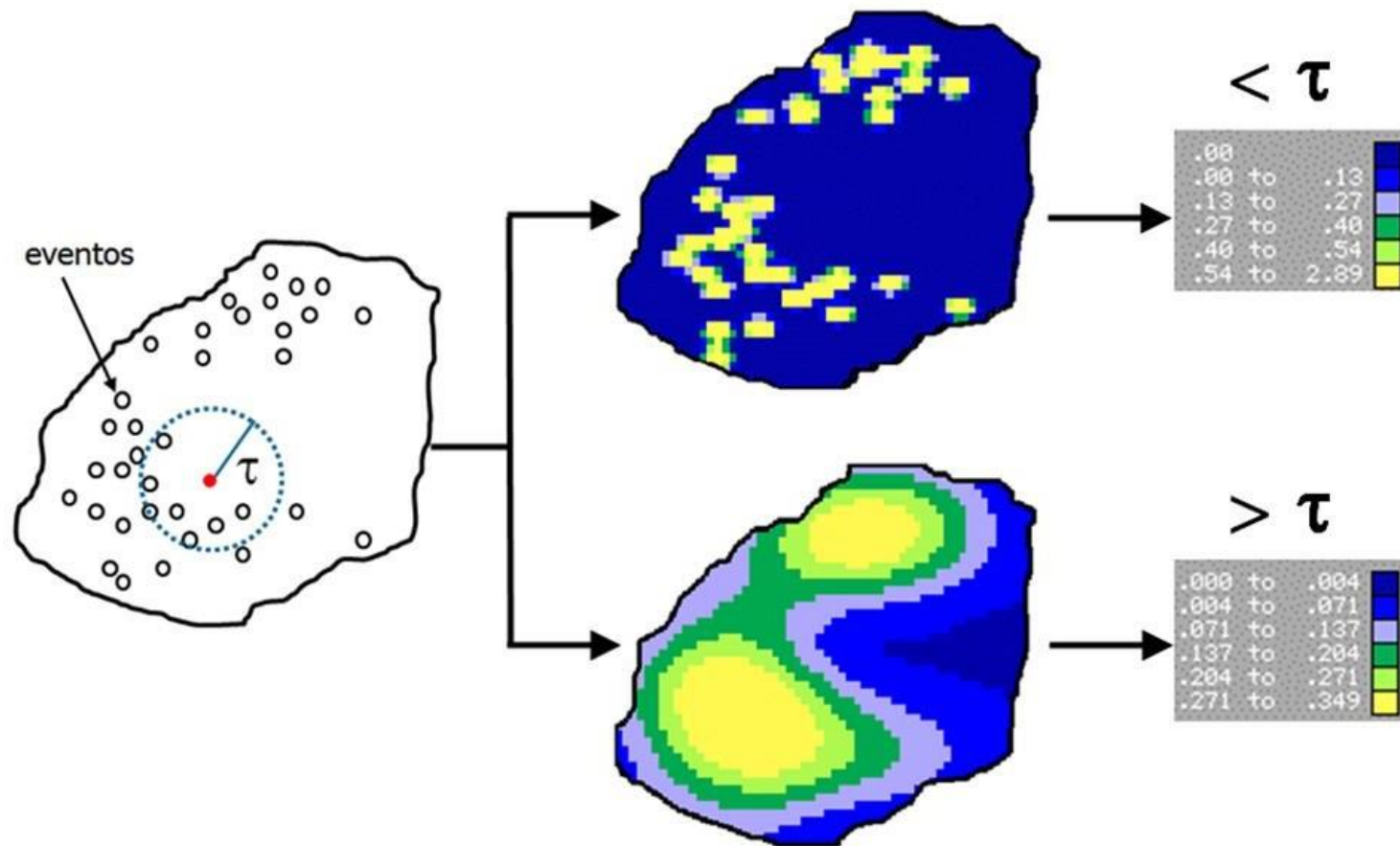
$h_i$ : é a distância entre o ponto a calcular  $s$  e o valor observado  $s_i$ .

## Idéia do estimador de Kernel

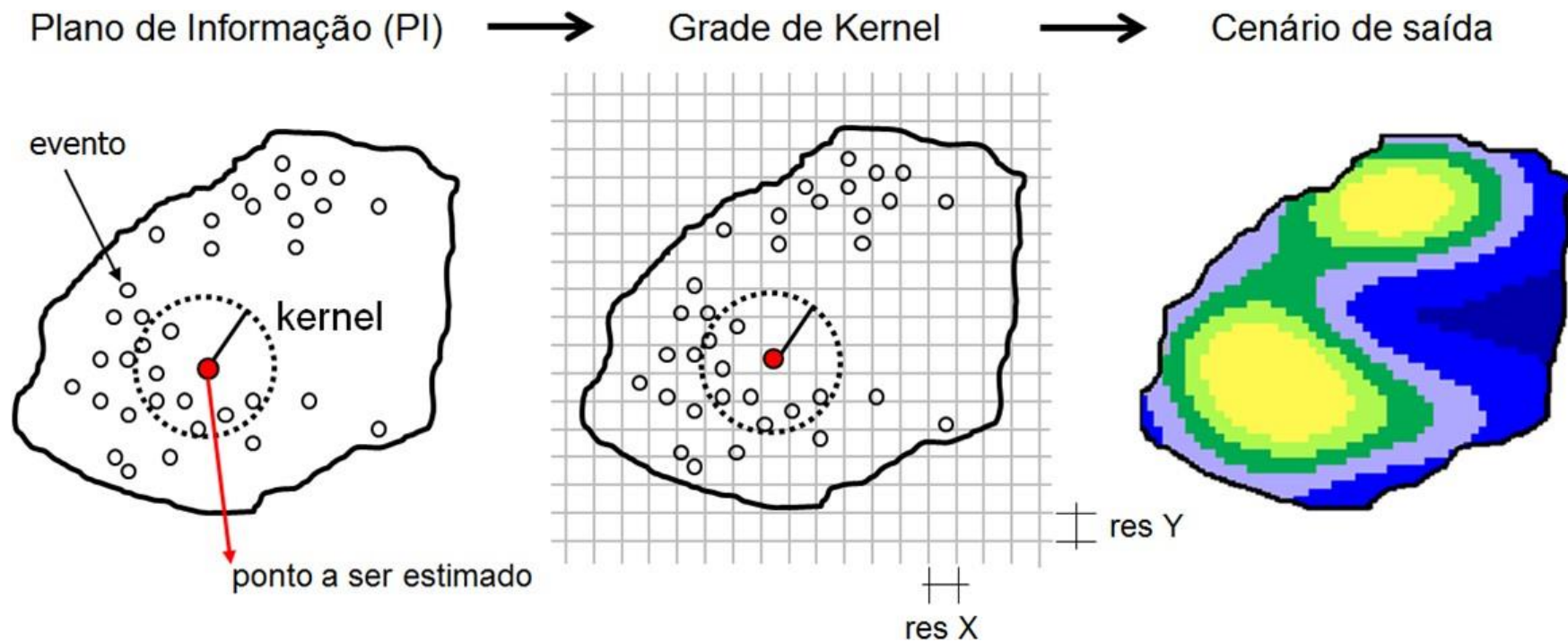
$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi \tau^2} \left( 1 - \frac{h_i^2}{\tau^2} \right)^2$$



## Efeitos da Largura de Faixa ( $\tau$ )

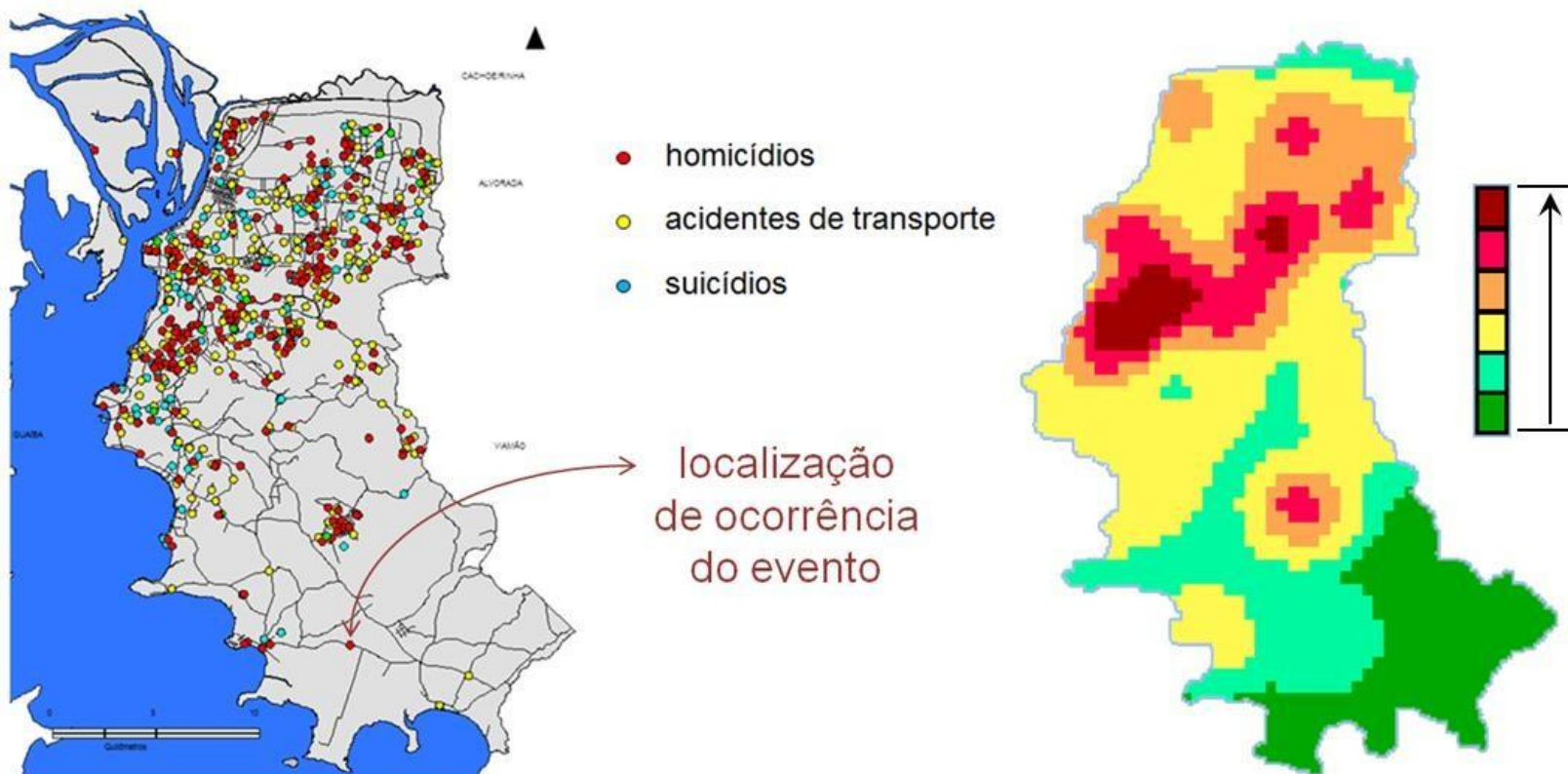


## Na prática (SIG)





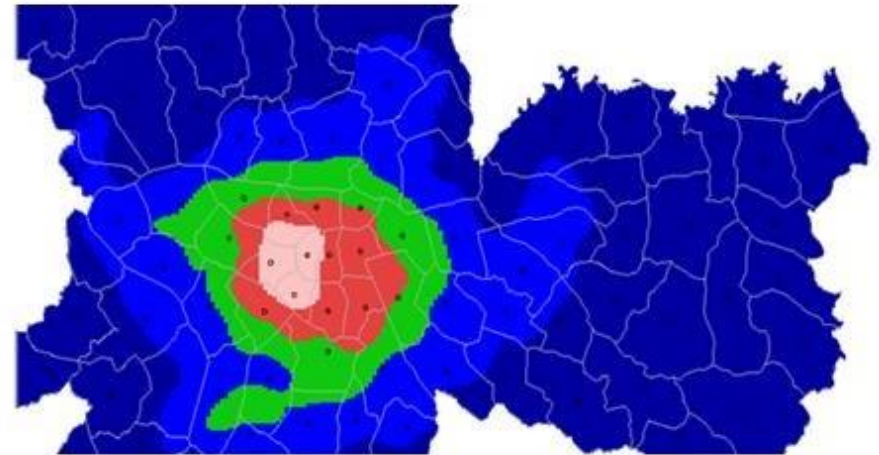
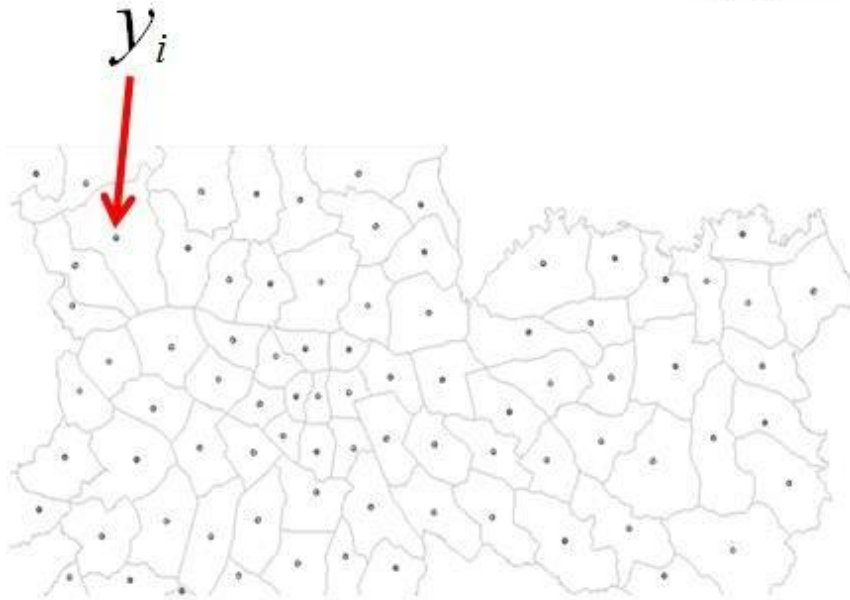
## Exemplo: mapeamento da violência em Porto Alegre - RS



Fonte: Santos, S.M. (1999).

Kernel com valor de atributo agregado:

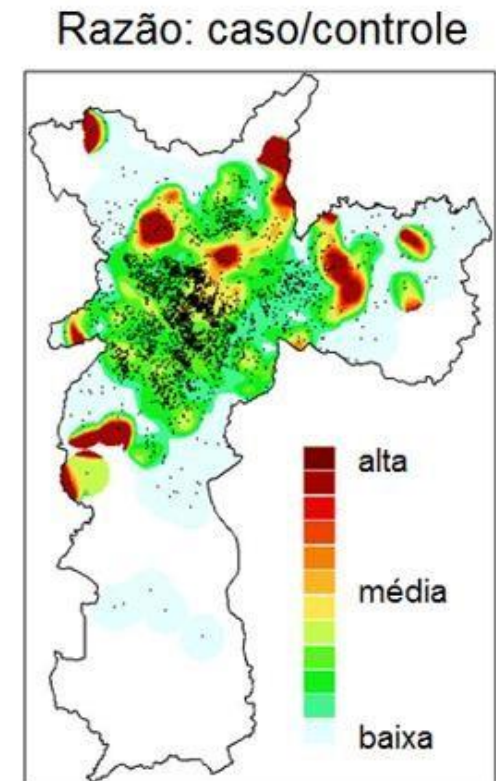
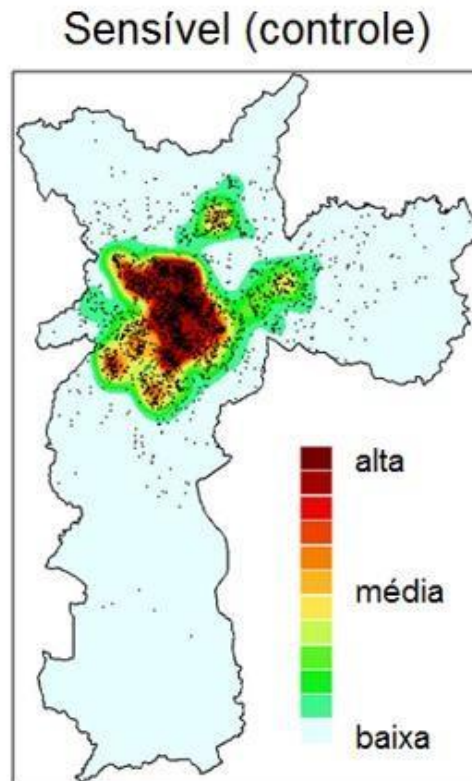
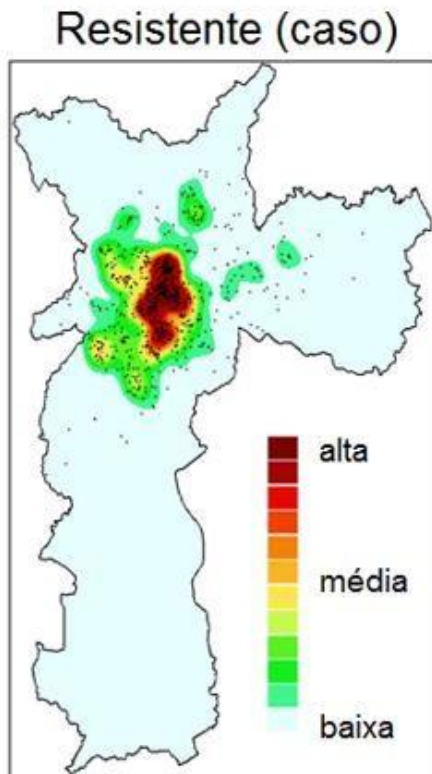
$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi \tau^2} \left( 1 - \frac{h_i^2}{\tau^2} \right)^2 \cdot y_i$$



Disparidade social: distribuição do percentual de idosos no centro de São Paulo.



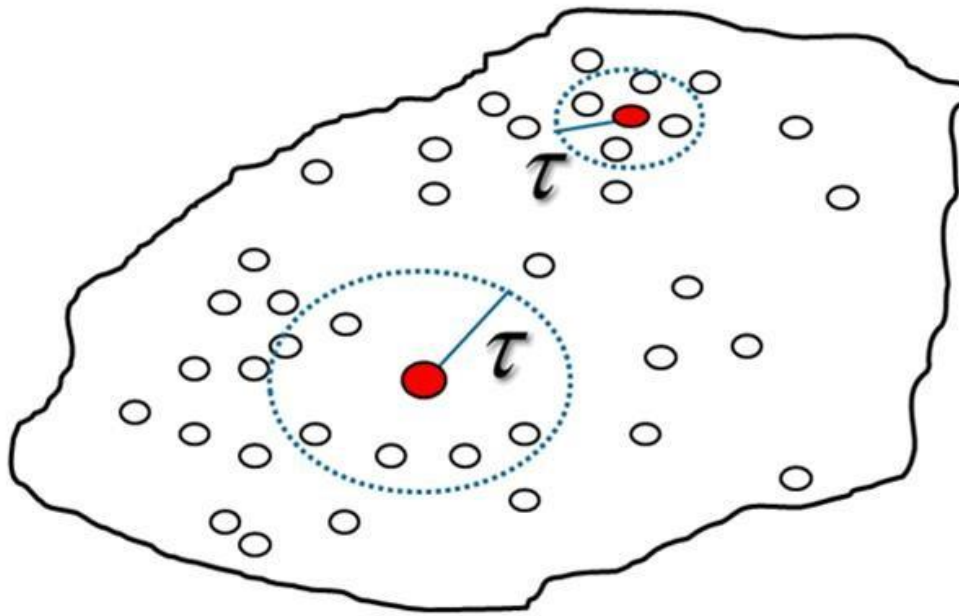
**Kernel de razão:** 
$$\hat{\lambda}_{\tau}(s) = \frac{\sum_{i=1}^n \frac{3}{\pi\tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2}{\sum_{i=1}^n \frac{3}{\pi\tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2}$$



Exploração da correlação espacial sobre o uso de antimicrobianos (ciprofloxacina) e a resistência de infecções de trato urinário na comunidade causado por *Escherichia coli*.

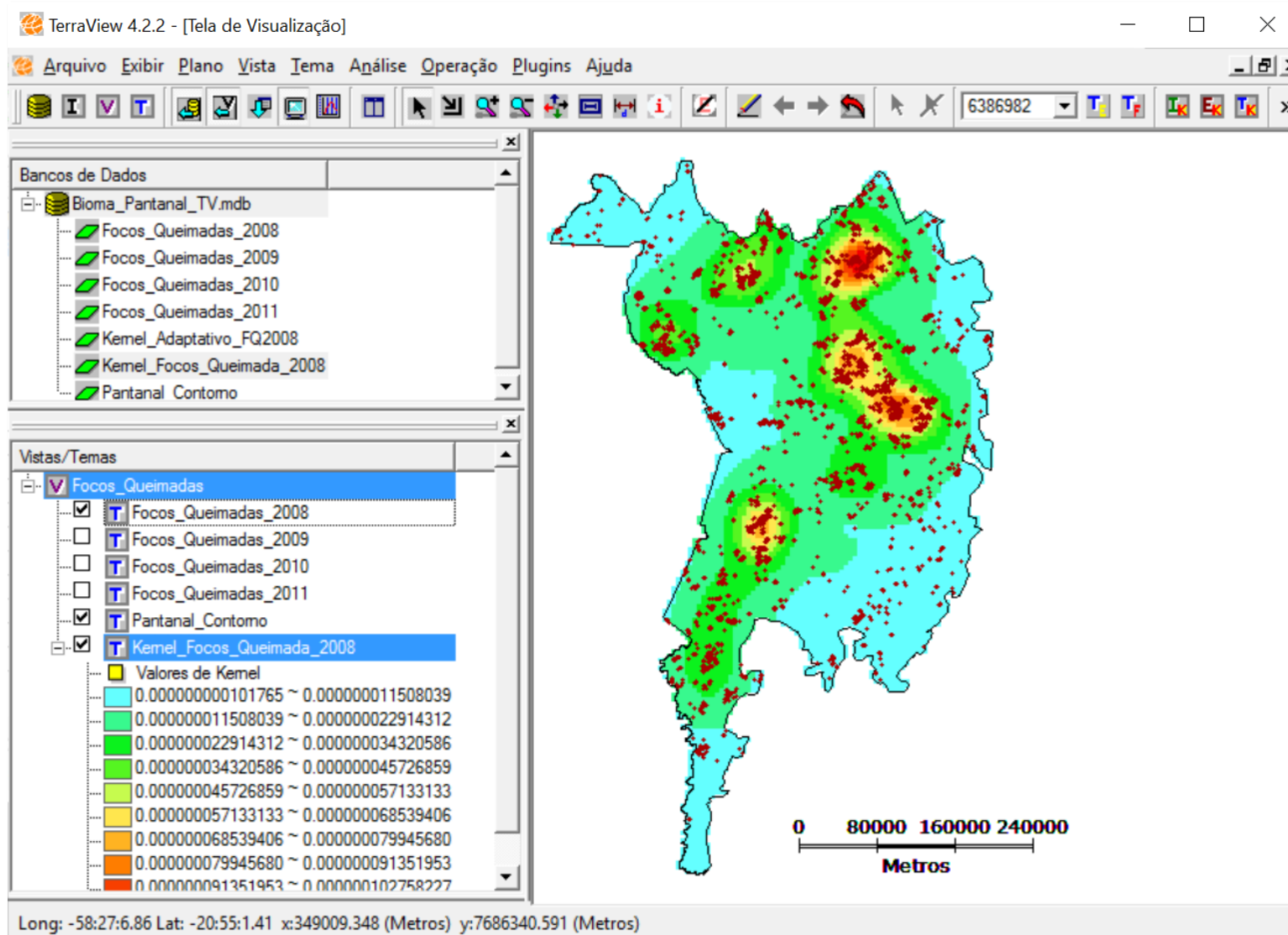
**Kernel adaptativo:**  $\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi\tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2$

↑                      ↑

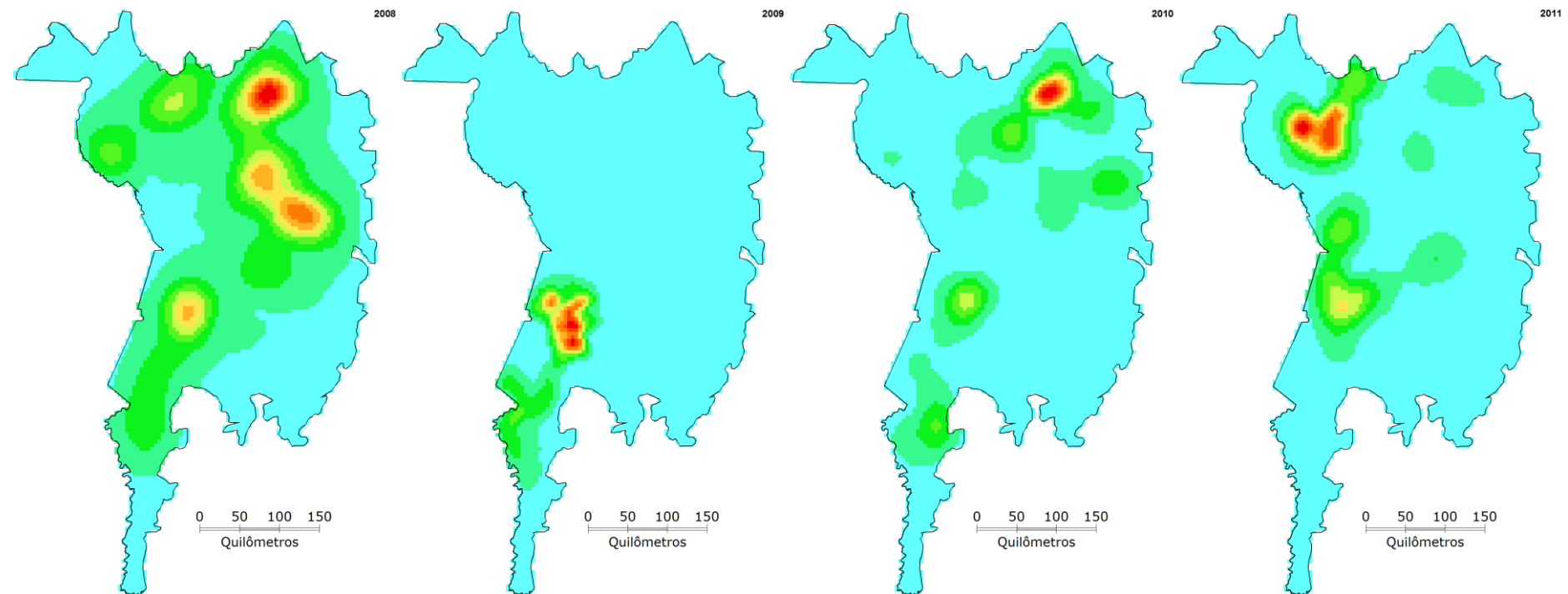


- Eventos observados
- Ponto a ser estimado

## Exemplo prático: Evolução Temporal de Focos de Queimadas no Bioma Pantanal



Resultados da Evolução Temporal de Focos de Queimadas no Bioma Pantanal, de 2008 a 2011.



Animação

▶ **modo passo** ( C:\Curso\_Analise\_Espacial\Analise\_Ponto\Animacao\Ani\_FQ2008\_2011\_passo.exe )

▶ **modo automático** ( C:\Curso\_Analise\_Espacial\Analise\_Ponto\Animacao\Ani\_FQ2008\_2011\_auto.exe )

# Executar o Roteiro Prático

AULA 1 - Análise de Distribuição de Pontos, até a página 17.

# Método do vizinho mais próximo

O método do vizinho mais próximo estima uma função de distribuição cumulativa baseado nas distâncias entre eventos em uma região de estudo.

É estimado empiricamente por:

$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n}$$

em que:

$w$  : distância de entrada

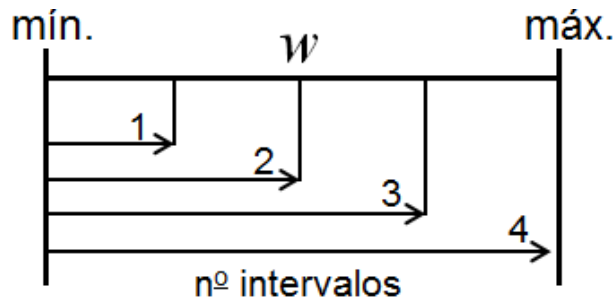
$w_i$  : distância entre eventos

$n$  : número de eventos

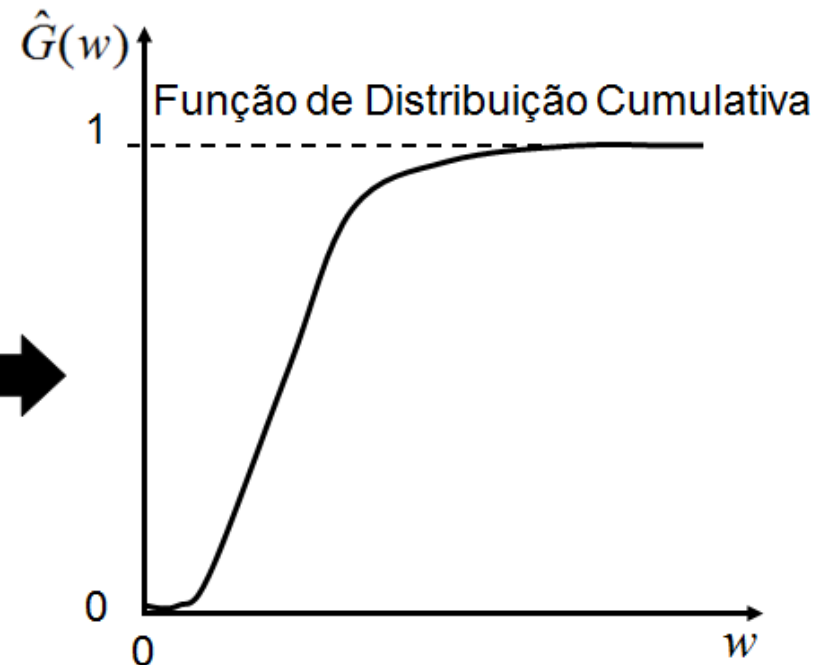
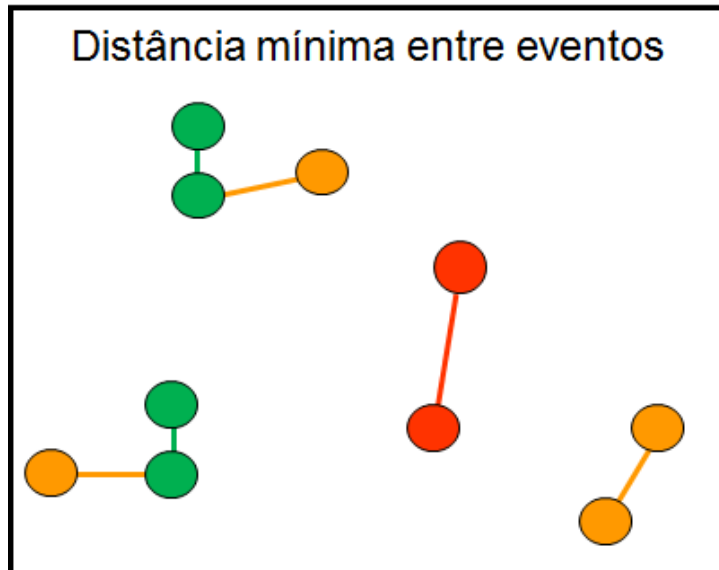
A plotagem dos resultados de  $\hat{G}(w)$  em relação as distâncias, pode ser utilizado como um método exploratório para verificar se existe evidência de interação entre os eventos.

# Método do vizinho mais próximo

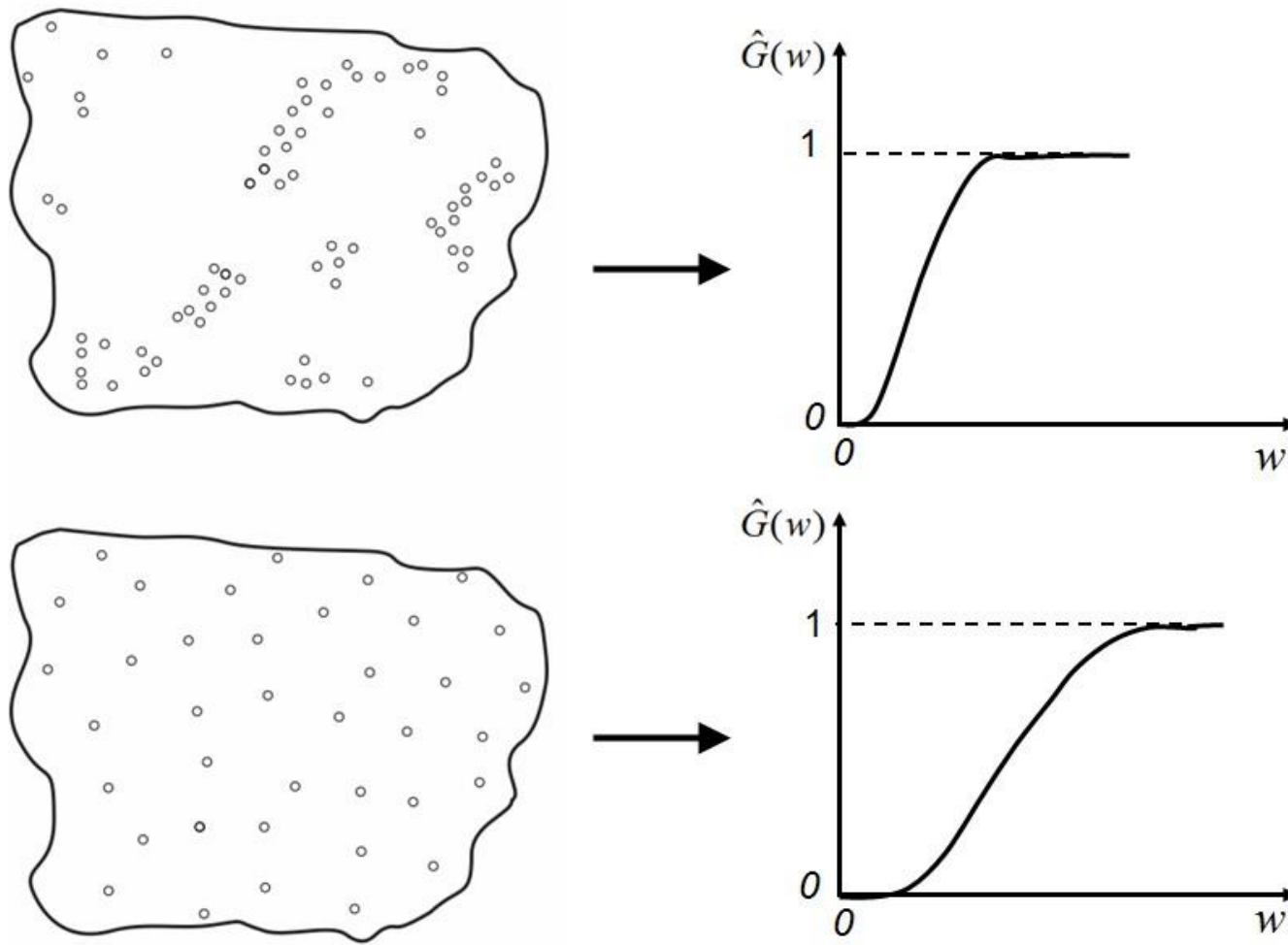
Na prática a distância de entrada  $w$  está compreendida entre um valor mínimo e máximo estabelecido pelo analista. Além disso, deve-se definir também o número de intervalos desejados.



$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n}$$



Análise exploratória de padrões de distribuição de pontos utilizando  $\hat{G}(w)$





# Método do vizinho mais próximo com simulação

A significância do resultado de  $\hat{G}(w)$  pode ser avaliada através de um modelo teórico denominado *Aleatoriedade Espacial Completa* - AEC (Bailey & Gatrell, 1995).

Na prática isto é realizado da seguinte forma:

**Primeiro:** realiza-se  $m$  simulações independentes dos  $n$  eventos assumindo-se AEC:

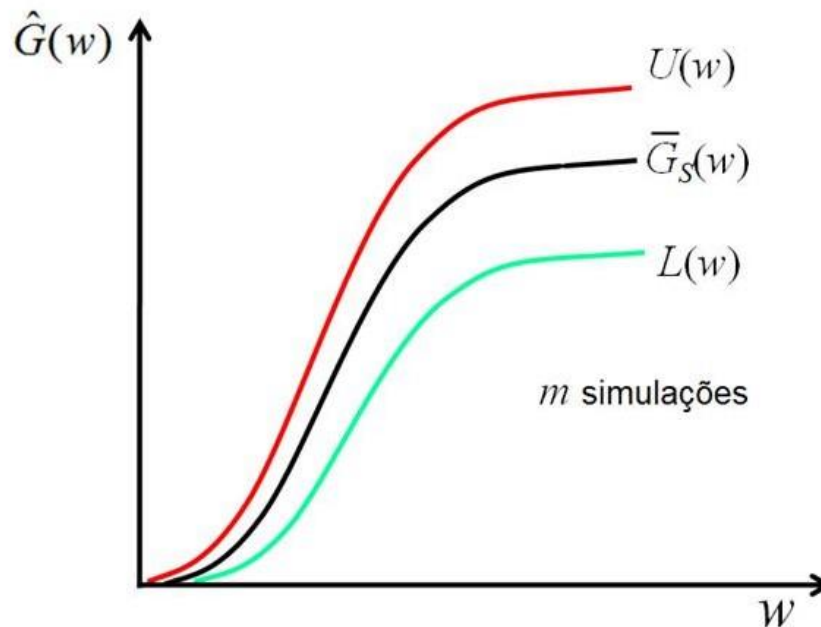
$$\hat{G}_{S_i}(w), i = 1, \dots, m$$

**Segundo:** calcula-se a média das  $m$  simulações:  $\overline{G}_S(w) = \frac{\sum_{i=1}^m \hat{G}_{S_i}(w)}{m}$

**Terceiro:** calcula-se os envelopes de simulação

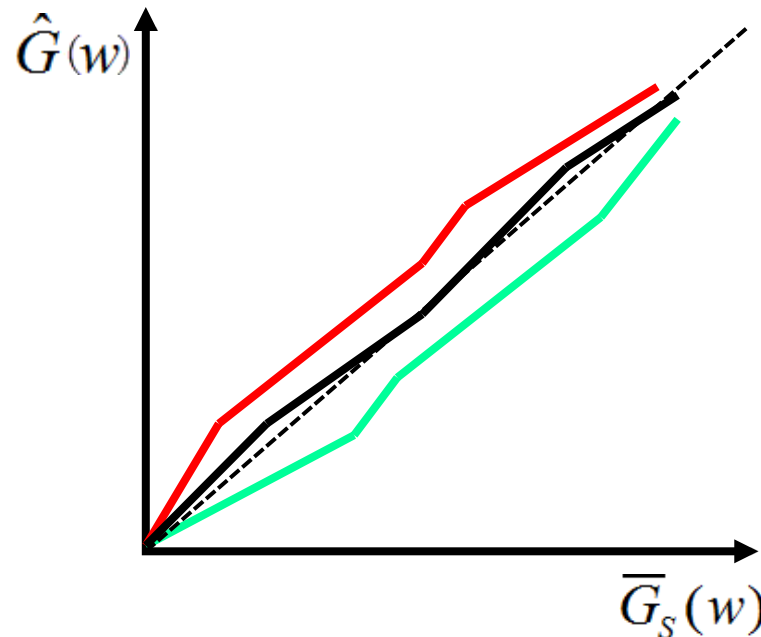
$$U(w) = \max \left\{ \hat{G}_{s_i}(w) \right\}, i = 1, \dots, m \quad \text{envelope superior}$$

$$L(w) = \min \left\{ \hat{G}_{s_i}(w) \right\}, i = 1, \dots, m \quad \text{envelope inferior}$$



# Método do vizinho mais próximo com simulação

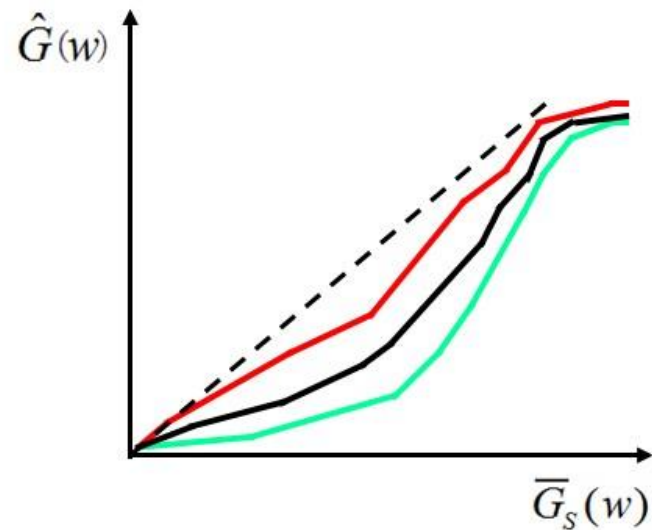
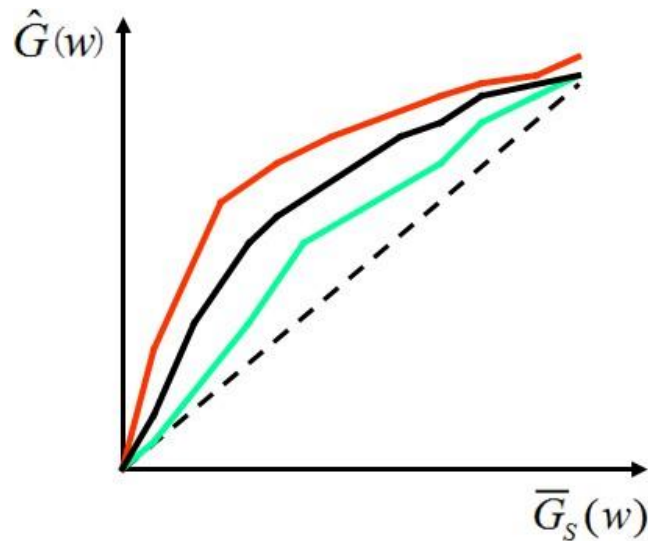
**Quarto:** plota-se  $\hat{G}(w) \times \overline{G}_S(w)$  com adição dos envelopes. Isto permite avaliar a significância do resultado de  $\hat{G}(w)$  relativo a hipótese de AEC.



Se a condição de AEC for válida a plotagem  $\hat{G}(w) \times \overline{G}_S(w)$  deve ser praticamente linear com um ângulo de 45°.

# Método do vizinho mais próximo com simulação

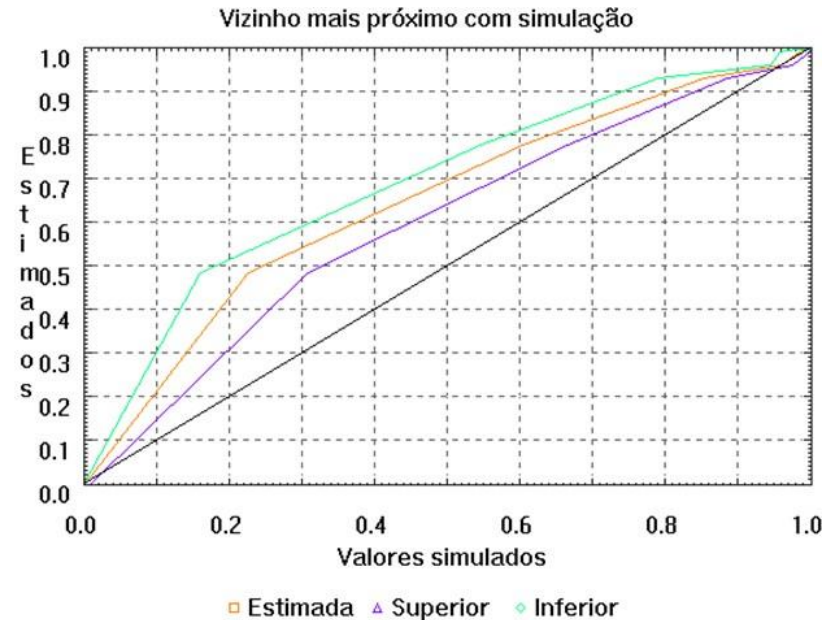
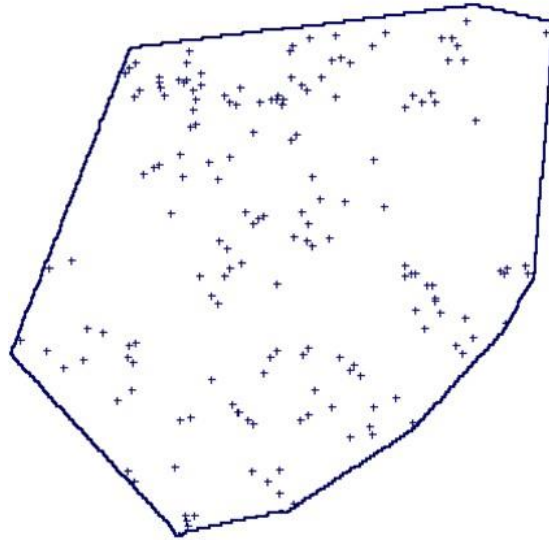
Se o dado apresenta tendências para agrupamento os traçados no gráfico estarão acima da linha de 45°.



Por outro lado, se o dado apresenta padrões de regularidade os traçados ficarão abaixo da linha de 45°.

# Método do vizinho mais próximo com simulação

**Exemplo:** Crimes juvenis na região de Cardiff, UK (Herbert, 1980).



Distâncias: mín.=0.2 máx.=16 #intervalos=10 #simulações=10

Observe que a posição dos envelopes e da estimativa estão acima da linha de 45°, o que indica evidências de agrupamento para as distâncias em análise.

# Continuar o Roteiro Prático

AULA 1 - Análise de Distribuição de Pontos, da pág. 18 a 26.

A função  $K$  considera escalas maiores para análise de padrões pontuais.

Conforme Bailey & Gatrell (1995):

$$\lambda K(h) = E [\#(\text{eventos a distância } h \text{ de um evento arbitrário})]$$

em que:

$\lambda$  : número de eventos por área

$E$  : operador esperança.

## Estimador da função $k$

1)  $\lambda K(h) = E [\#(\text{eventos a distância } h \text{ de um evento arbitrário})]$

2)  $\lambda \hat{K}(h) = \frac{1}{\lambda R} \sum_{i=1}^n \sum_{j=1}^n I_h(d_{ij})$  em que:  $d_{ij}$  é a distância entre os eventos  $i$  e  $j$ .  
 $I_h(d_{ij}) = 1$  se  $d_{ij} \leq h$ , caso contrário  $= 0$  se  $d_{ij} > h$ .  
 $\lambda R = n = \text{número de eventos em } R$ .

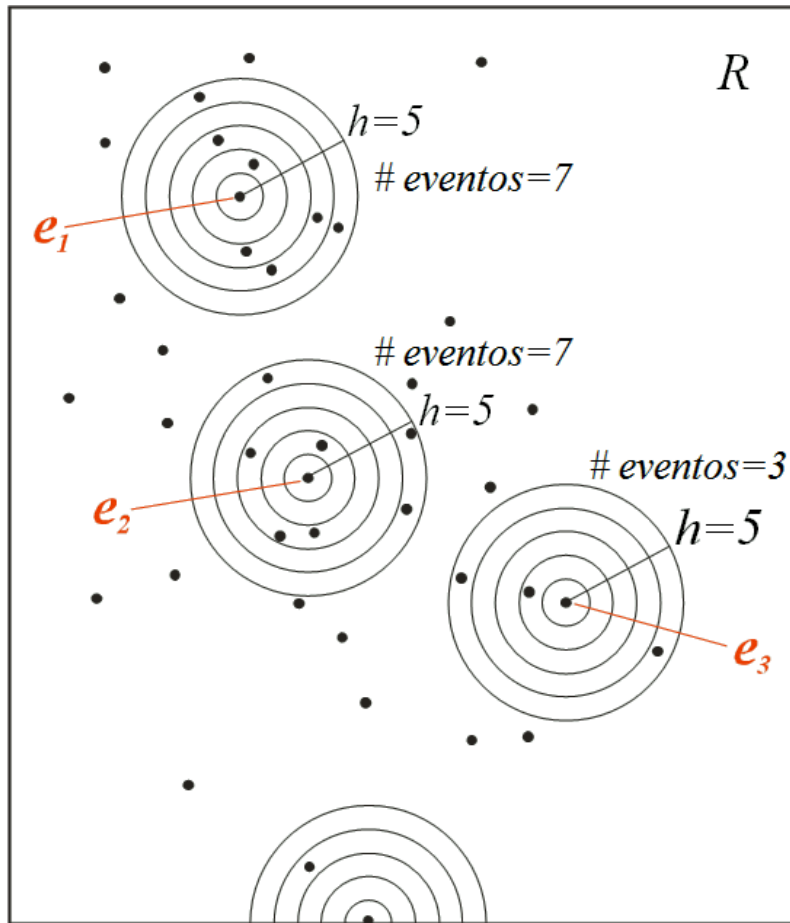
3)  $\hat{K}(h) = \frac{1}{\lambda^2 R} \sum_{i=1}^n \sum_{j=1}^n I_h(d_{ij})$

4) O estimador de lambda  $\hat{\lambda} = \frac{n}{R}$

$$\hat{K}(h) = \frac{R}{n^2} \sum_{i=1}^n \sum_{j=1}^n I_h(d_{ij})$$



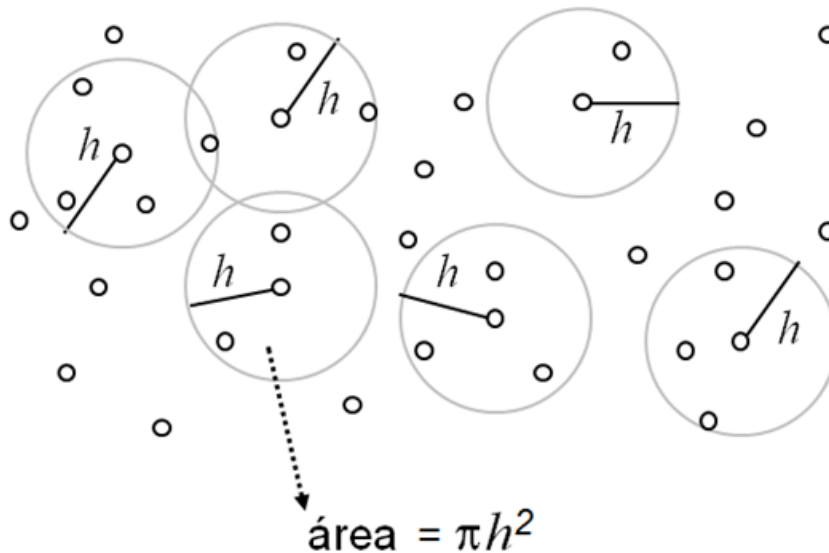
## Idéia do estimador da função $K$



$$\hat{K}(h) = \frac{R}{n^2} \sum_{i=1}^n \sum_{j=1}^n I_h(d_{ij})$$

Segundo Bailey & Gatrell (1995), para um processo aleatório o número esperado de eventos a uma distância  $h$  de um evento escolhido de forma arbitrária é:

$$\lambda k(h) = \lambda \pi h^2$$



| Processo           | $\hat{K}(h)$ |
|--------------------|--------------|
| aleatório          | $= \pi h^2$  |
| regular            | $< \pi h^2$  |
| agregação espacial | $> \pi h^2$  |

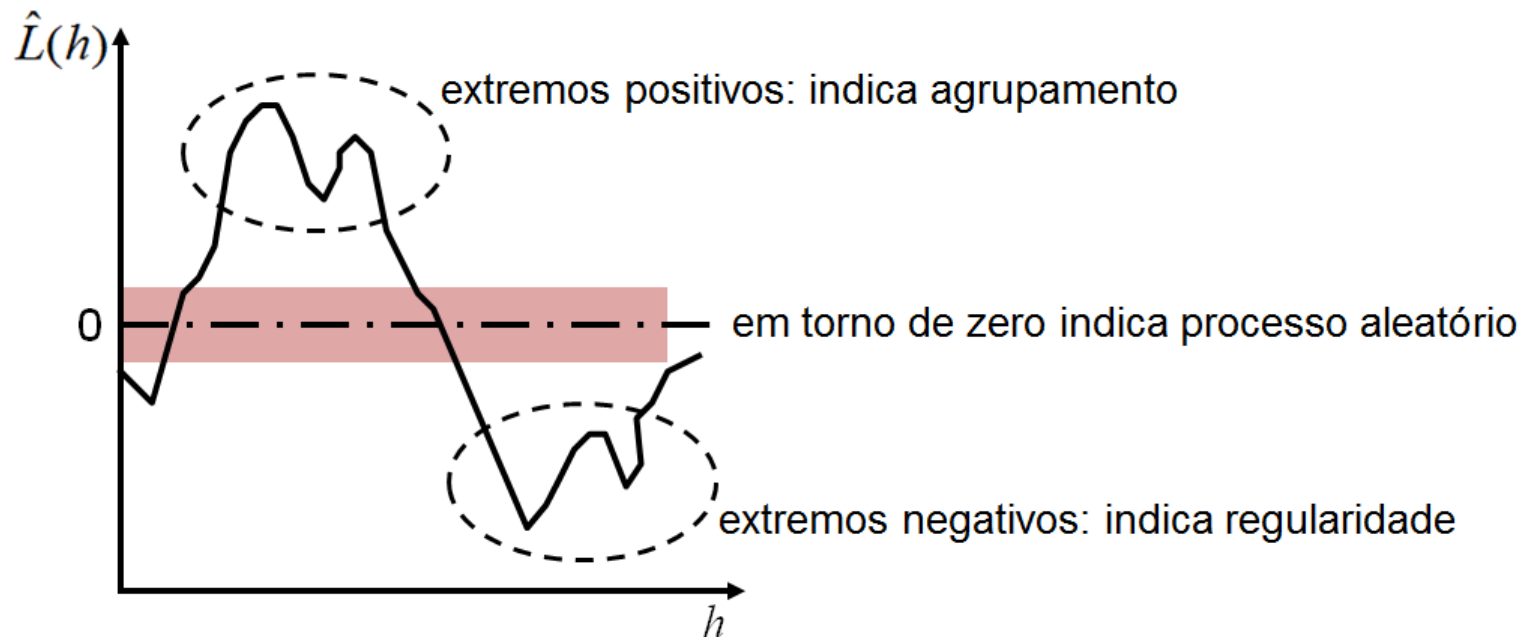
O gráfico da função  $K$  não é tão intuitivo quanto a do gráfico do vizinho mais próximo.

**Artifício:** emprega a função auxiliar  $L$  para facilitar a interpretação.

**Estimador da função  $L$**  é dado por:  $\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h$

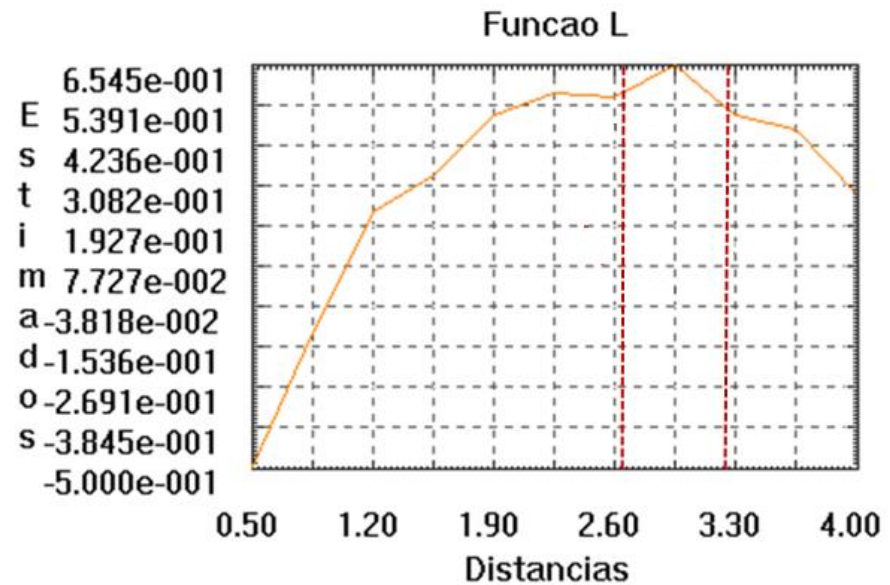
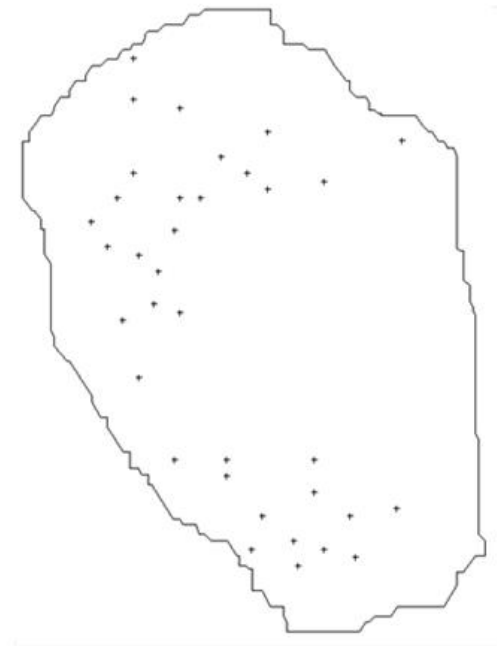
| Processo           | $\hat{L}(h)$ | $\hat{K}(h)$ |
|--------------------|--------------|--------------|
| aleatório          | $= 0$        | $= \pi h^2$  |
| regular            | $< 0$        | $< \pi h^2$  |
| agregação espacial | $> 0$        | $> \pi h^2$  |

Interpretação da plotagem de  $\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h$



## Exemplo:

Dados geomorfológicos com 36 localizações de rochas de granito, no município Bood Moor, situado no condado de Cornwall, Inglaterra.



Distância: mín.=0.5 máx.=4 #intervalos=10

Neste exemplo percebe-se que para distâncias entre aproximadamente 2.5 a 3, extremo positivo do gráfico, há evidências de agrupamento.

Uma abordagem similar a do vizinho mais próximo pode ser feita para se estimar a significância dos desvios da distribuição  $\hat{L}(h)$  em relação a  $AEC$ .

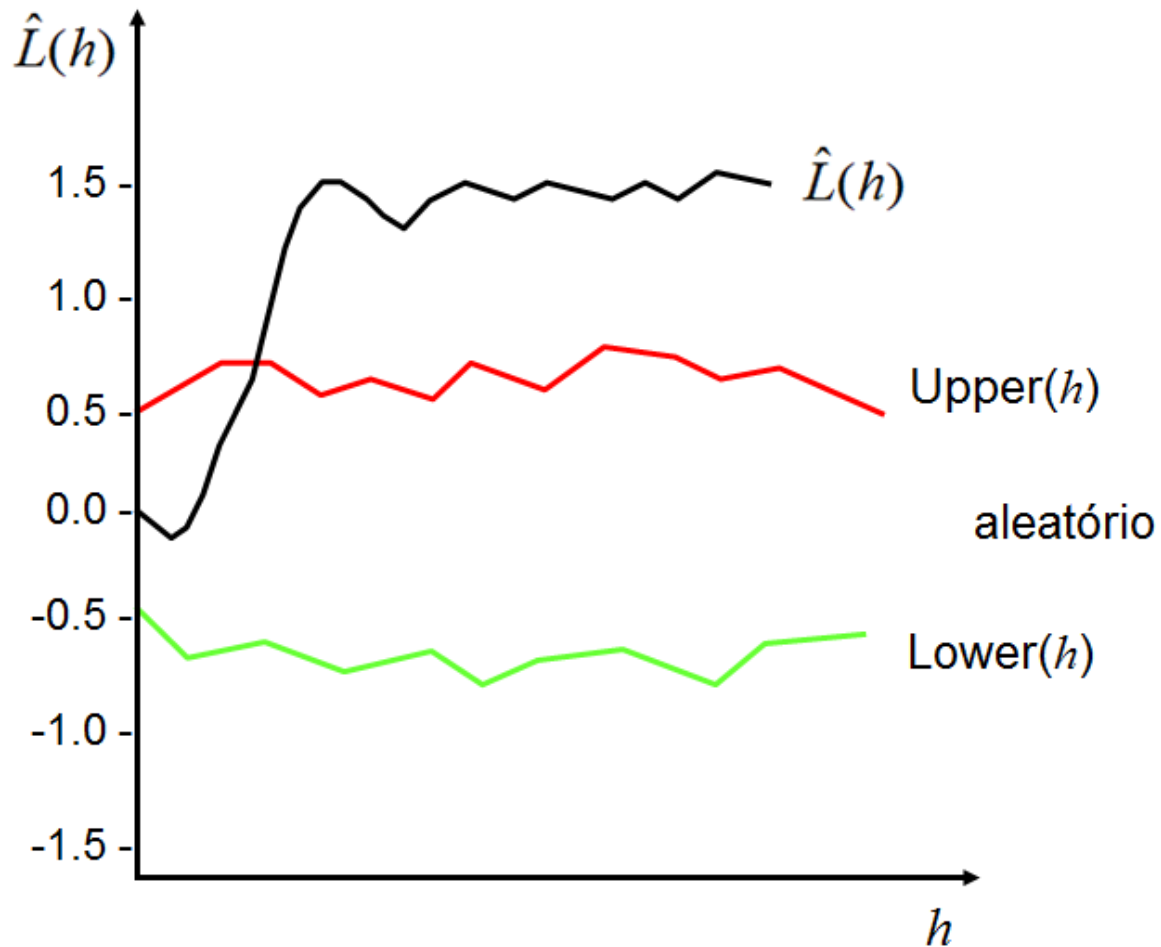
A idéia é realizar  $m$  simulações sobre a região  $R$  e computar os envelopes superior e inferior.

Os envelopes superior e inferior são definidos como (Baley & Gratel, 1995):

$$Upper(h) = \max \left\{ \hat{L}_i(h) \right\}, i = 1, \dots, m$$

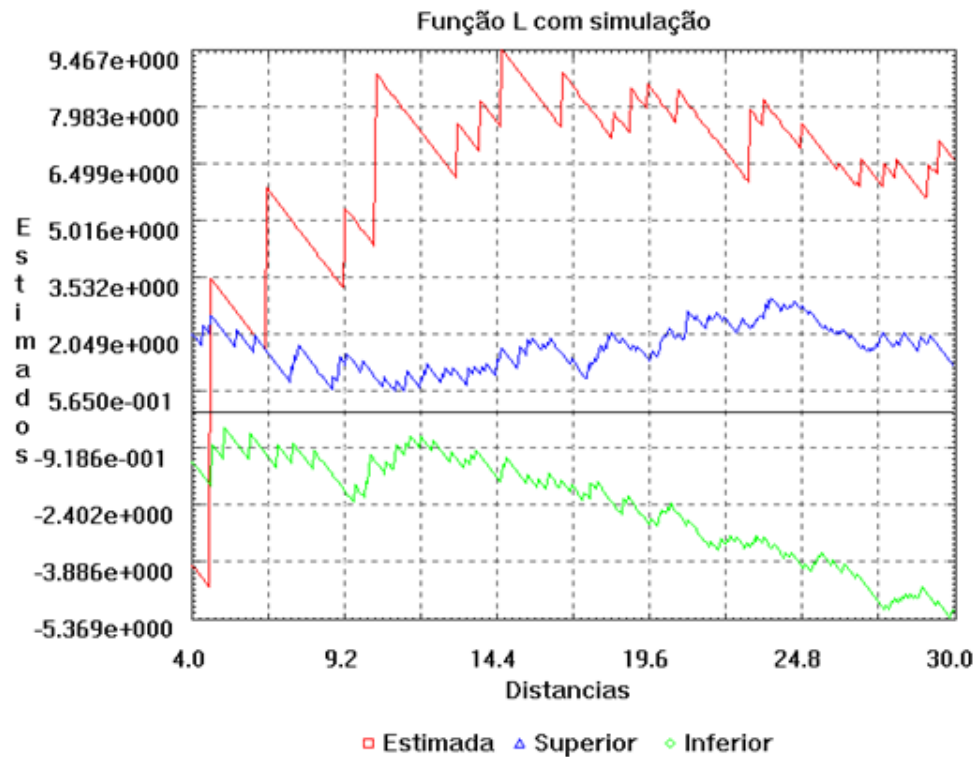
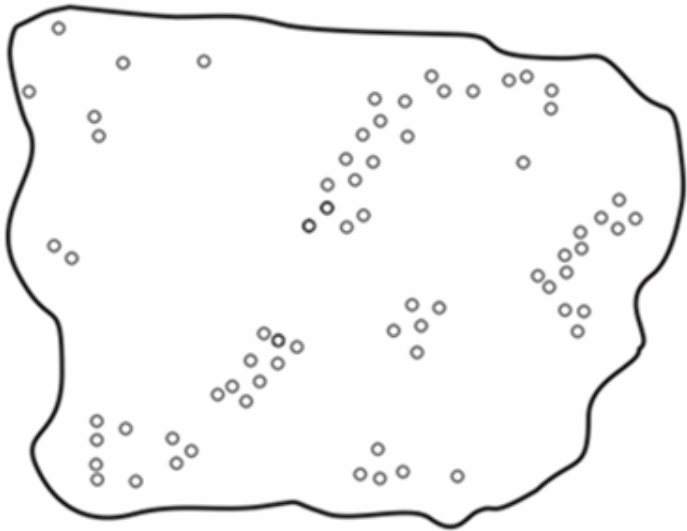
$$Lower(h) = \min \left\{ \hat{L}_i(h) \right\}, i = 1, \dots, m$$

Análise do gráfico  $\hat{L}(h)$  com os envelopes Upper( $h$ ) e Lower( $h$ ).



## Exemplo:

Dados de mudas de sequoia distribuídas numa região de 23km<sup>2</sup>.





## Encerrar o Roteiro Prático

AULA 1 - Análise de Distribuição de Pontos, da pág. 27 a 30.