

Análise Espacial de Dados Geográficos

Laboratório

Módulo: Análise de Padrões de Distribuição de Pontos

Aluno: FELIPE DE LUCIA LOBO

Mestrado SERE 2007

O objetivo deste laboratório é ilustrar as várias formas de analisar padrão de pontos, a partir de alguns conjuntos de dados. As ferramentas de análise disponíveis no SPRING são: o Interpolador Kernel, o Método do Vizinho Mais Próximo e a Função K.

Os dados disponíveis são:

- **Bairros de São Paulo**

Dados: percentual de idosos (mais que 70 anos) na cidade de São Paulo, agregados por bairros.

- **Porto Alegre - RS**

Dados: violência por homicídios, acidentes de transporte e suicídios, na cidade de Porto Alegre – RS.

1. INTERPOLADOR KERNEL PONTOS

Para a interpolação, isto é, gerar uma grade numérica de superfície a partir de dados pontuais, da intensidade de eventos de mortalidade no município de Porto Alegre foi aplicado estimativa Kernel. O estimador de Kernel é uma medida de um suavizador de pontos (**# de eventos por unidade de área**), conforme definido abaixo:

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

onde : h_i é a distância do ponto s e a localização s_i ; τ é a largura da banda; quanto menor menos suave é o efeito.

Para efeito de comparação foram realizadas 3 interpolações Kernel utilizando diferentes larguras de bandas, 1500, 2500 e 5000 metros, de modo a verificar seus efeitos e resultados.

Na figura 1 estão os resultados de cada interpolação onde se pode verificar o crescente efeito de suavização, não só das bordas entre as classes como da intensidade de eventos de mortalidade, com o aumento da largura de banda de 1500 a 5000 metros. Isso ocorre porque para o calculo da intensidade em cada ponto da superfície a largura da banda ao quadrado entra na equação como um fator de divisão.

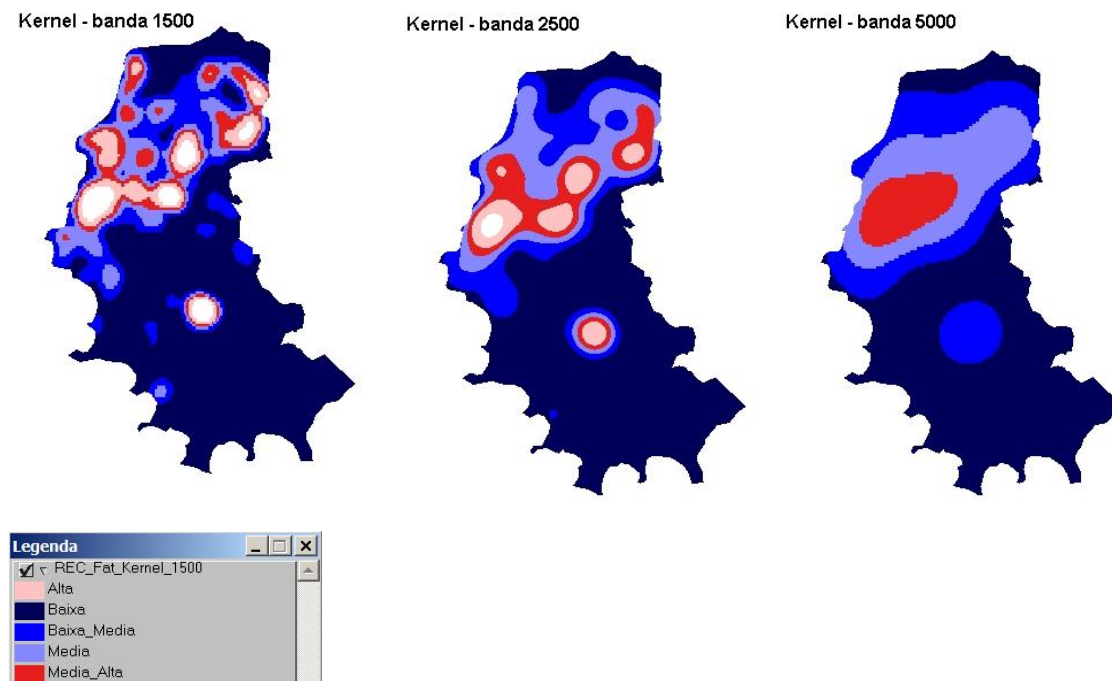


Figura 1: Resultados da interpolação do estimador Kernel – com bandas de 1550, 2500 e 5000 metros - sobre o s eventos de mortalidade em Porto Alegre.

2. APLICAÇÃO DO KERNEL CONSIDERANDO O VALOR DO ATRIBUTO

Neste segundo exercício do Kernel, utilizaremos o banco de dados Bairros_SP. Os dados contidos neste banco referem-se somente aos bairros da região central e leste da cidade de São Paulo, conforme Figura abaixo. O atributo a ser analisado é o percentual de idosos (pessoas com mais que 70 anos).

Diferente do exemplo anterior, agora o Kernel refere-se à medida da quantidade total do atributo por unidade de área, conforme expresso:

$$\lambda'_\tau(s) = \sum_{j=1}^m \frac{1}{\tau^2} k\left(\frac{(s - s'_j)}{\tau}\right) y_j$$

Para realizar o cálculo acima, o estimador assume que cada observação de área Y_i deve ser associado a um evento S_j , como por exemplo, o centróide dessa área. Assim, a estimativa para um novo ponto S considera o número de S_i presentes conforme a largura da banda. O mesmo raciocínio de suavização para bandas mais largas é válido também nesse exercício. Pode-se perceber, na figura 2, que a estimativa Kernel é mais suave quando se utiliza bandas mais largas.

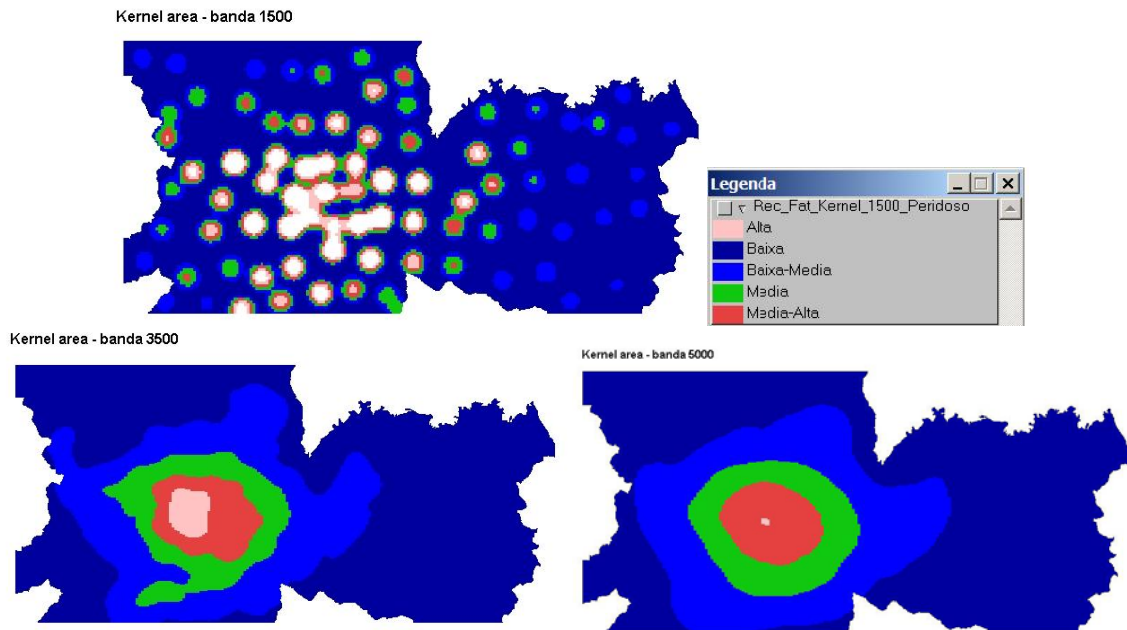


Figura 2: Resultado do estimador Kernel áreas por atributo – porcentagens de idosos – utilizando as seguintes larguras de bandas: 1500, 3500 e 5000 m.

3. ANÁLISE DO VIZINHO MAIS PRÓXIMO – *BANCO DADOS PORTO ALEGRE*

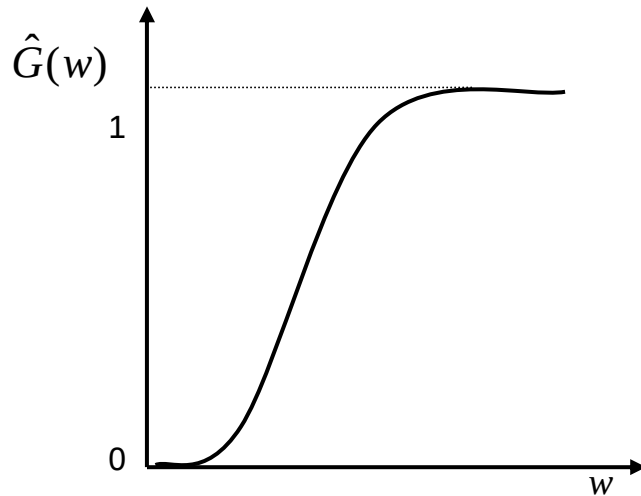
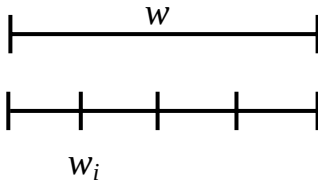
A técnica de análise pelo método do vizinho mais próximo, nada mais é do que um gráfico da frequência acumulada da distância de cada ponto (evento) ao seu vizinho mais próximo, conforme a equação e a Figura abaixo:

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

n : número de eventos;

w_i : faixas de distâncias e

w : distância de análise.



A interpretação do gráfico acima é conforme segue:

- Se o gráfico mostra uma função com brusca elevação no início, pode sugerir um agrupamento na escala considerada.
- Por outro lado, se a elevação ocorrer a distâncias maiores sugere-se regularidade entre os eventos.
- Caso essa elevação seja constante, com inclinação próxima a 45° , sugere-se aleatoriedade.

Esse método investiga sobre as propriedades de segunda ordem, processos isotrópicos, utilizando distancias entre os eventos na área de estudo (Bailey, 2000). É preciso ressaltar que esse método se aplica a pequenas distâncias entre os eventos, uma vez que busca identificar os processos de segunda ordem entre-eventos.

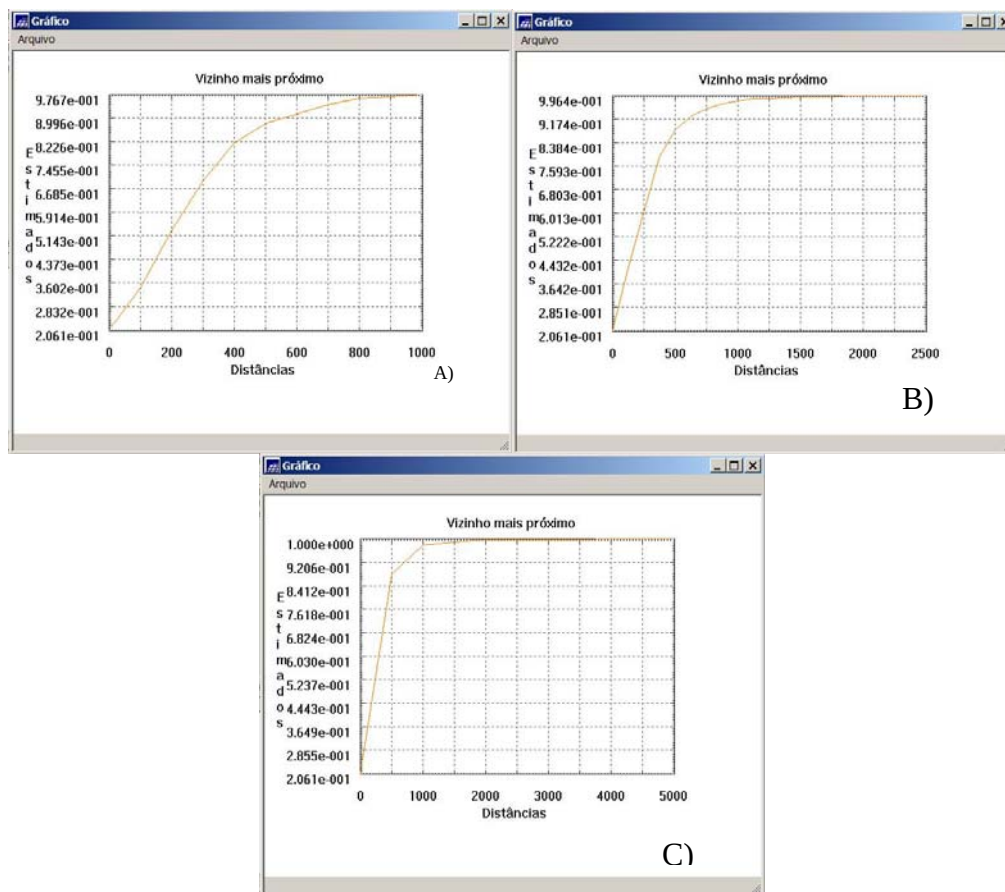


Figura 3: Resultado do método vizinho mais próximo em 3 tratamentos: a) largura de banda 1000 m; b) 2500; e c) 5000m.

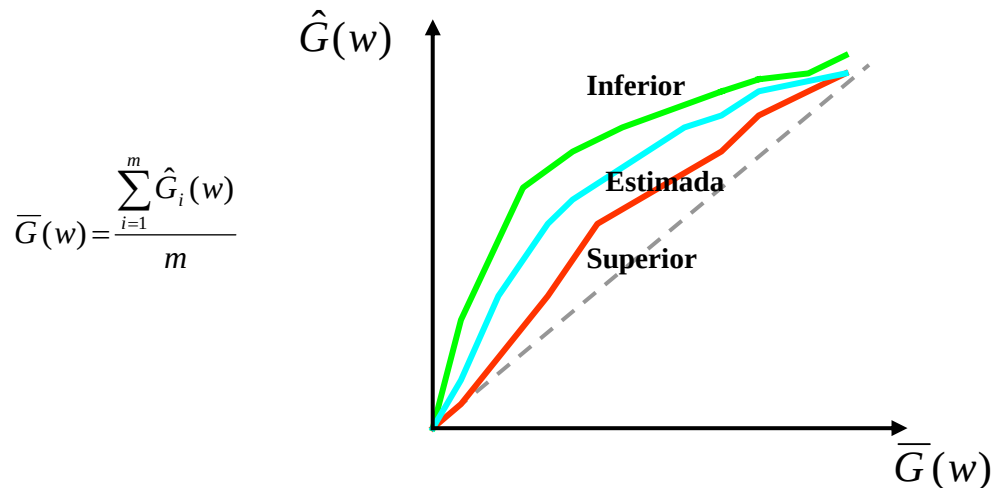
Nota-se que para a distância de 1000 m a distribuição dos pontos indica aleatoriedade dos pontos, quando a distância é 2500, já se percebe uma indicação de agrupamento, confirmada na estimativa com distância de 5000m. A definição da distância a ser adotada deve ser tomada de acordo com o objetivo e geometria dos dados, além, da escala de trabalho.

4. ANÁLISE DO VIZINHO MAIS PRÓXIMO COM SIMULAÇÃO

Esse método permite a comparação da função acumulada das amostras com os de dados gerados aleatoriamente (envelopes de simulação superior e inferior).

Envelope Superior: $U(w) = \max\{\hat{G}_i(w)\}$, $i = 1, 2, 3, \dots, m$ (m é número de simulações).

Envelope Inferior: $I(w) = \min\{\hat{G}_i(w)\}$, $i = 1, 2, 3, \dots, m$



A interpretação do gráfico acima é conforme segue:

- Se as amostras estão distribuídas aleatoriamente, o gráfico da distribuição estimada pela simulada deve ser linear, próxima de 45° ;
- Existindo agrupamento, a função observada deverá estar acima da reta de 45° ;
- Na presença de regularidade ficará abaixo da reta de 45° .

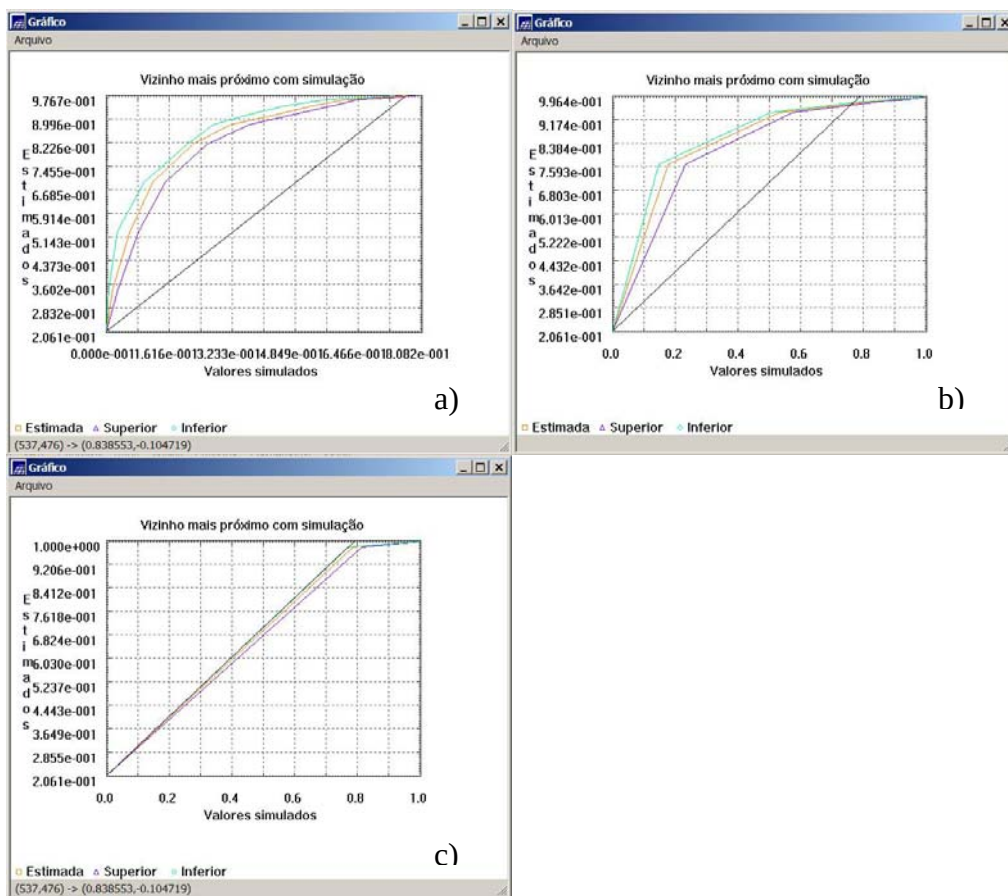


Figura 4: Resultado de vizinho mais próximo com simulação variando a distância da banda: a) banda de 10000m; b) 3500; c) 1000m.

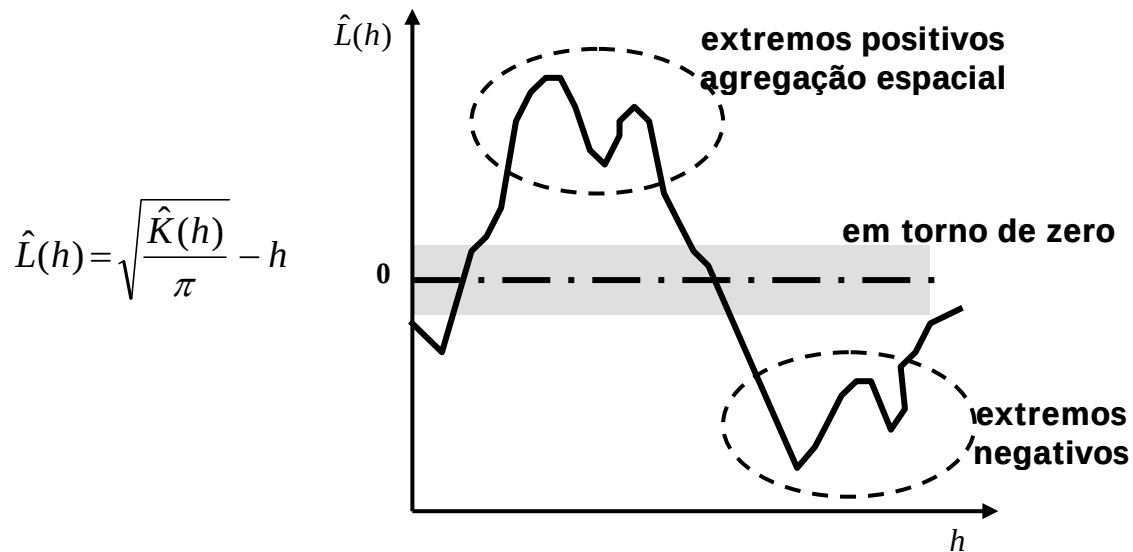
A análise com simulação reforça as indicações obtidas na seção 3. Com banda de 1000m percebe-se indicação de aleatoriedade na distribuição dos pontos, com distância de banda de 3500 m já é possível sugerir agrupamento, e para 10000m a sugestão de agrupamento é nítida.

5. ANÁLISE DA FUNÇÃO L (DERIVADA DA FUNÇÃO K)

Ao contrário do método do vizinho mais próximo, a função K considera escalas maiores para análise de padrões pontuais.

O gráfico da função K não é tão intuitivo quanto a do gráfico do vizinho mais próximo. Portanto utiliza-se uma função auxiliar L, para facilitar a interpretação.

O estimador da função L é:



A interpretação do gráfico acima é como segue:

- 1- **Extremos positivos:** correspondem às distâncias (h) onde há evidências de agregação na distribuição espacial dos eventos;
- 2- **Em torno de zero:** diz-se que os eventos apresentam um tipo de distribuição espacial aleatória;
- 3- **Extremos negativos:** há evidências de ordenação regular na distribuição espacial dos eventos.

Os resultados sugerem haver agrupamento na distribuição dos pontos analisados nas distâncias de 10000 e 12000, que é maior que a distância onde se observou agrupamento no método por vizinho mais próximo (2500m). Dessa forma, verifica-se que o método da função k é mais adequado para escalas geográficas menores.

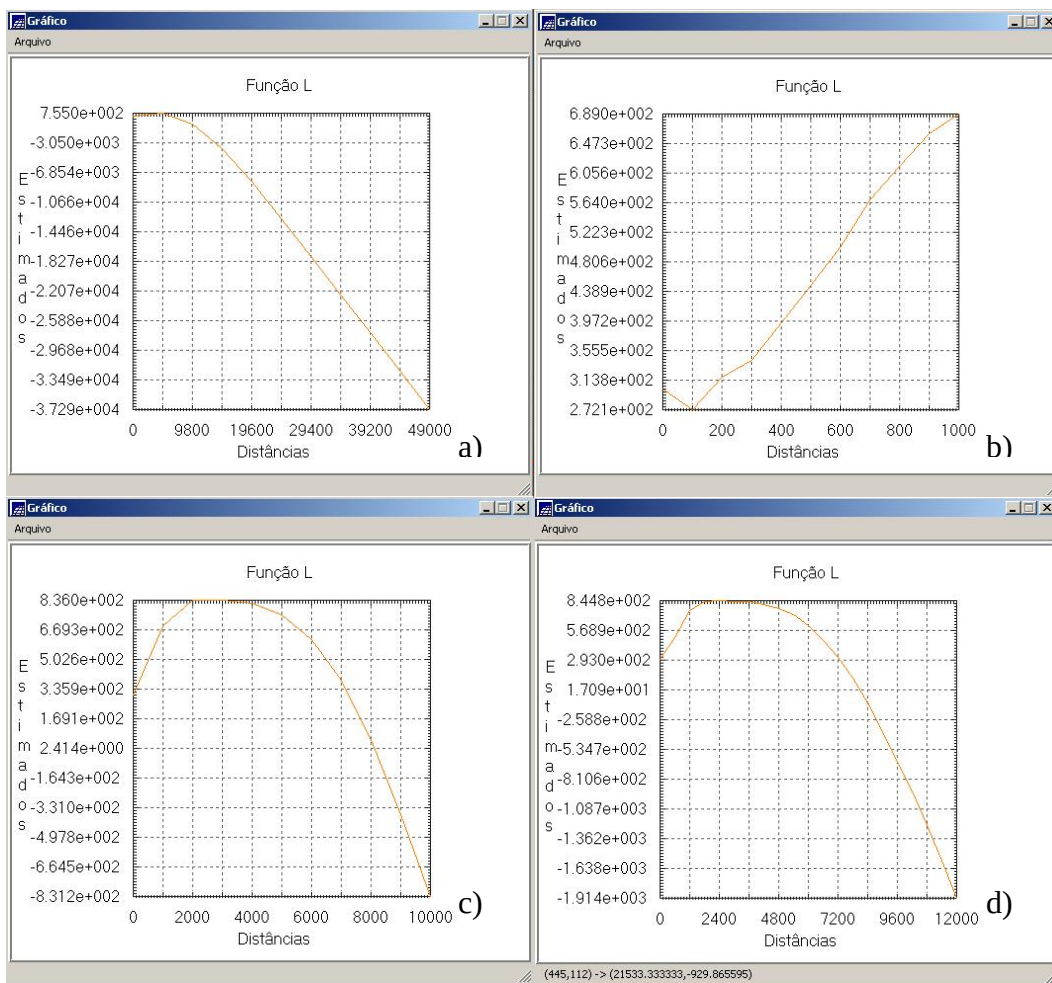
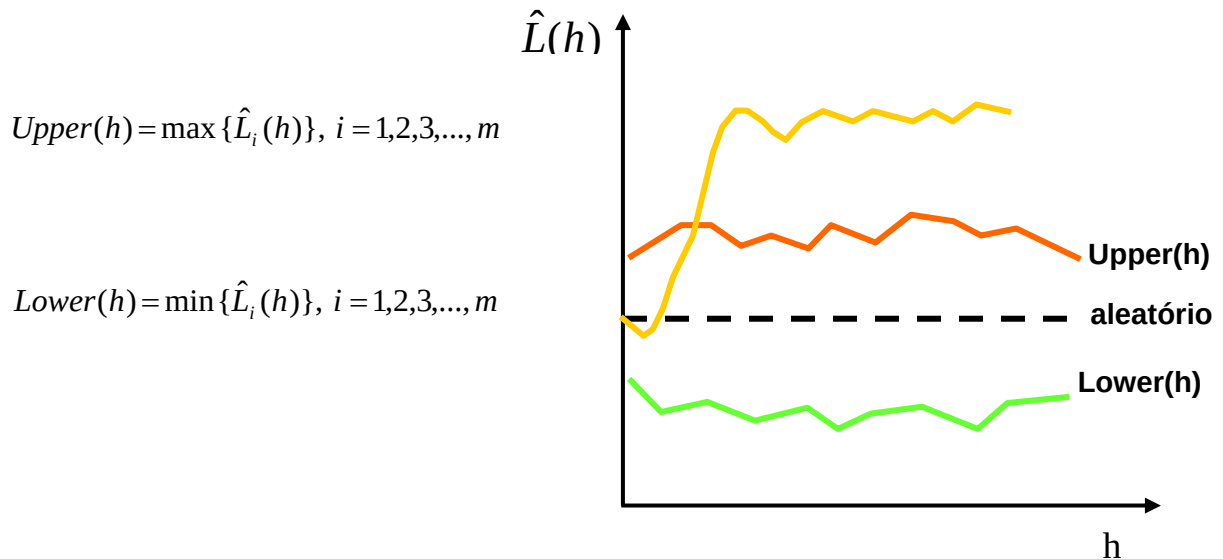


Figura 5: Resultados da Função L aplicada sobre diferentes distâncias de bandas: a) 49000m; b) 1000 m; c) 10000 m; d) 12000m.

6. ANÁLISE DA FUNÇÃO L COM SIMULAÇÃO

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 aleatoriedade (*CSR-Complete Spatial Randomness*).

A Idéia é realizar simulações CSR sobre a região R e computar os envelopes superior e inferior, definidos conforme segue:



A interpretação do gráfico acima é como segue:

Verifica-se valores positivos para $\hat{L}(h)$, estando os mesmos acima dos envelopes, o que caracteriza agrupamento nesta escala de distância. O agrupamento é mais forte para as distâncias que correspondem aos extremos da curva de $\hat{L}(h)$.

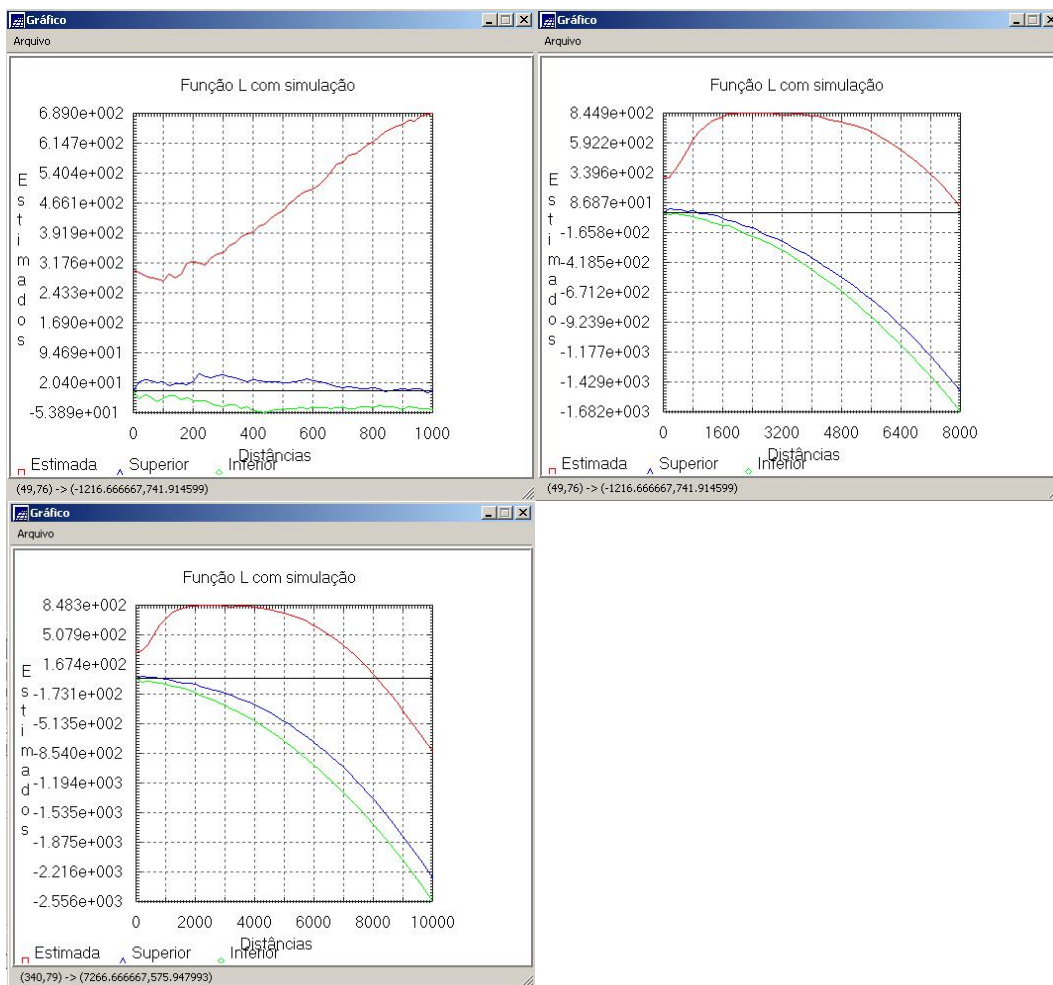


Figura 6: Resultados das funções L com simulações: a) distância de 1000m; b) 8000; c) 10000m.

Para todas distâncias testadas há indicação de agrupamento entre as amostras. Isto é, os valores estimados foram superiores ao envelope de aleatoriedade da distribuição dos pontos.