

Métodos Computacionais para Analisar Padrões de Pontos Espaciais

João A. C. Paiva

*Instituto Nacional de Pesquisas Espaciais (INPE)
Departamento de Processamento de Imagens - DPI
Av. dos Astronautas, 1758 - Jardim da Granja
São José dos Campos - SP CEP:12201-970
Telefone: (012) 345-6474 Fax: (012) 345-6468
e-mail: miro@dpi.inpe.br*

Andrea Rodríguez

*National Center for Geographic Information and Analysis
University of Maine - Boardman Hall
Orono, Me 04469 - U.S.A.
Telephone: (011) (207) 581-2128 Fax: (011) (207) 581-2206
e-mail: andrea@spatial.maine.edu*

Virgínia R. M. Correia

*Instituto Nacional de Pesquisas Espaciais (INPE)
Departamento de Processamento de Imagens - DPI
Av. dos Astronautas, 1758 - Jardim da Granja
São José dos Campos - SP CEP:12201-970
Telefone: (012) 345-6507 Fax: (012) 345-6468
e-mail: virginia@dpi.inpe.br*

RESUMO

O objetivo básico no processo de análise de padrões de pontos espaciais consiste em verificar se um conjunto de pontos ou eventos exibe algum padrão sistemático, tipo regularidade, agrupamento, ou aleatoriedade. Este trabalho objetiva o desenvolvimento de ferramentas computacionais para análise univariada de padrões pontuais. Essas ferramentas exploratórias se concentram, em particular, na análise das propriedades de segunda ordem dos dados pontuais que sugerem a existência ou não de dependência espacial entre os pontos. Os métodos de análise desenvolvidos se baseiam nas distâncias aos vizinhos mais próximos, e na função K que corresponde à expectativa do número de eventos para uma certa distância de um evento arbitrário. Um módulo computacional foi desenvolvido e adicionado ao software SPRING contendo ferramentas para explorar a dependência espacial entre os dados pontuais.

ABSTRACT

The basic objective of point pattern analysis is to check whether the observed events in a region exhibit any systematic pattern, such as clustering, regularity, or randomness. The goal of this work is to develop software components to support the univariate point pattern analysis. In particular, the software concentrates on exploratory tools for analyzing second order properties of the data. The second order or spatial dependence of a spatial point pattern comprises the analysis of the relationship between events in a region. The methods included in this work are the nearest neighbor distances and the K-function which is the estimation of the number of events within a certain distance of an arbitrary event. These exploratory tools have been integrated into the software SPRING.

1. Introdução

O processo de análise de dados espaciais envolve a descrição dos dados relacionados a um processo espacial, assim como a exploração de padrões e relacionamentos entre estes dados, e a busca de explicações para estes padrões e relacionamentos. Fenômenos espaciais geralmente são representados por

primitivas geométricas tipo pontos, linhas e polígonos. O tipo de representação de um objeto espacial pode variar dependendo da escala de trabalho. A representação pontual pode corresponder a dados do tipo população, crimes em certas localidades, ocorrências de doenças, característica do solo em um ponto, e outros.

O objetivo na análise de padrões de pontos espaciais é de verificar se existe uma tendência nos eventos observados de exibirem um padrão sistemático, como regularidade, agrupamentos ou aleatoriedade. No contexto desse trabalho atributos não espaciais associados aos pontos não são considerados. A análise é feita baseando-se nas coordenadas dos eventos na região de estudo.

Como em qualquer processo estocástico, o processo de análise de pontos pode ser descrito em termos dos efeitos de primeira e segunda ordem. Os efeitos de primeira ordem, considerados globais ou de grande escala, correspondem a variações no valor médio do processo no espaço. Efeitos de segunda ordem, denominados locais ou de pequena escala, representam a dependência espacial no processo proveniente da estrutura de correlação espacial. O processo de análise de dados espaciais contém métodos de visualização, métodos exploratórios para investigar algum padrão nos dados e métodos que auxiliem a escolha de um modelo estatístico e a estimação dos parâmetros desse modelo.

Este trabalho descreve algumas ferramentas exploratórias para análise espacial de pontos univariados, que foram introduzidas recentemente no software SPRING (SPRING, 1996). Essas ferramentas se concentram, em particular, com a análise das propriedades de segunda ordem do dado espacial. Os métodos de análise desenvolvidos baseiam-se nas distâncias ao vizinho mais próximo, e na função K que corresponde a estimativa do número de eventos para uma certa distância de um evento arbitrário. Para esses dois métodos, o programa apresenta ferramentas para explorar e visualizar a característica dos dados.

As seções seguintes deste trabalho estão organizadas da seguinte forma: a seção 2 apresenta os métodos para investigar dependência espacial entre pontos espaciais contidos em uma região de trabalho. A seção 3 descreve como estas ferramentas de análise foram incorporadas ao software SPRING. A seção 4 descreve exemplos de aplicações utilizando-se alguns dados pontuais, e a seção 5 conclui este trabalho.

2. Análise de segunda ordem

Segundo Bailey e Gatrell (1995), as propriedades de segunda ordem ou dependência espacial em um processo espacial de pontos envolvem o relacionamento entre os eventos localizados em uma região. Uma abordagem prática é se procurar relações de distâncias entre eventos e propriedades de segunda ordem. Os métodos de distância ao vizinho mais próximo e a função K são exemplos de procedimentos que consideram a distância entre eventos. Estes métodos assumem estacionaridade sobre pequenas escalas na região de trabalho. Este conceito de estacionaridade significa que a média e a variância do processo são independentes da localização e constante na região.

Os métodos da distância ao vizinho mais próximo e da função K constituem uma base para o desenvolvimento de métodos formais que checam a significância dos resultados exploratórios quando comparados com um modelo teórico conhecido. Neste caso, o modelo teórico é conhecido como aleatoriedade espacial completa ("complete spatial randomness - CSR"). Inicialmente verifica-se o desvio da aleatoriedade, e caso isto ocorra, identifica-se a natureza do desvio. Assumindo-se a hipótese de aleatoriedade CSR, os eventos seguem o processo homogêneo de Poisson. Dado um processo espacial pontual $Y(A)$ com A pertencendo a uma região de estudo R , $Y(A_i)$ e $Y(A_j)$ são independentes para qualquer escolha de A_i and A_j . No processo de Poisson, um evento tem a mesma probabilidade de ocorrer em qualquer posição da região R , e a posição de um evento é independente da posição de outro.

É possível se obter n eventos aleatórios em uma região compreendida por um retângulo envolvente $\{(x,y) : x_1 \leq x \leq x_2, y_1 \leq y \leq y_2\}$. Para isso, os eventos são gerados a partir de coordenadas x , obtidas de uma distribuição uniforme em (x_1, x_2) e de coordenadas y , obtidas de uma distribuição uniforme em (y_1, y_2) . Pontos que caem fora da região são rejeitados. Este processo é repetido até que n eventos tenham sido obtidos na região. Descreve-se a seguir o processo de geração deste pontos aleatórios dentro de uma região (Algoritmo 1):

```

Seja uma região R
Repita até que n eventos tenham sido adicionados a lista de saída
Início
    Gere um evento aleatório dentro do retângulo envolvente (MBR) de R
    Adiciona-se este evento na lista de saída se o mesmo cair dentro de R
Fim

```

Algoritmo 1: Geração de pontos aleatórios dentro de uma região.

Os conceitos acima formam uma base teórica para a descrição que se segue sobre os métodos do vizinho mais próximo e da função K.

2.1 Vizinho mais próximo

O método do vizinho mais próximo calcula a função de distribuição cumulativa ($G(w)$) baseado nas distâncias entre eventos em uma região de análise. Na análise de pontos univariados, esta função de distribuição é definida da seguinte forma:



onde o valor normalizado acumulado G para uma distância de entrada w corresponde à soma dos vizinhos mais próximo de cada evento cuja distância é menor ou igual a w , dividido pelo número de eventos na região. A plotagem dos resultados desta função de distribuição cumulativa $G(w)$ em relação as distâncias w , pode ser usado como um método exploratório para se verificar se existe evidência de interação entre os eventos. Se esta plotagem apresentar um crescimento rápido para pequenos valores de distância, esta situação aponta para interação entre os eventos caracterizando agrupamentos nestas escalas. Por outro lado se esta plotagem apresentar valores pequenos no seu início, e só crescer rapidamente para valores maiores de distância, esta situação aponta para uma distribuição mais regular. A Figura 1.a mostra as localizações vulcânicas da região Oeste de Uganda (Tinkler, 1971), e a Figura 1.b mostra um exemplo de plotagem para a análise dos vizinhos mais próximo. Estes dados foram importados do programa INFO-MAP (Bailey e Gatrell, 1995). Percebe-se um rápido crescimento da curva para valores pequenos de distância, o que caracteriza agrupamentos nesta escala de análise.

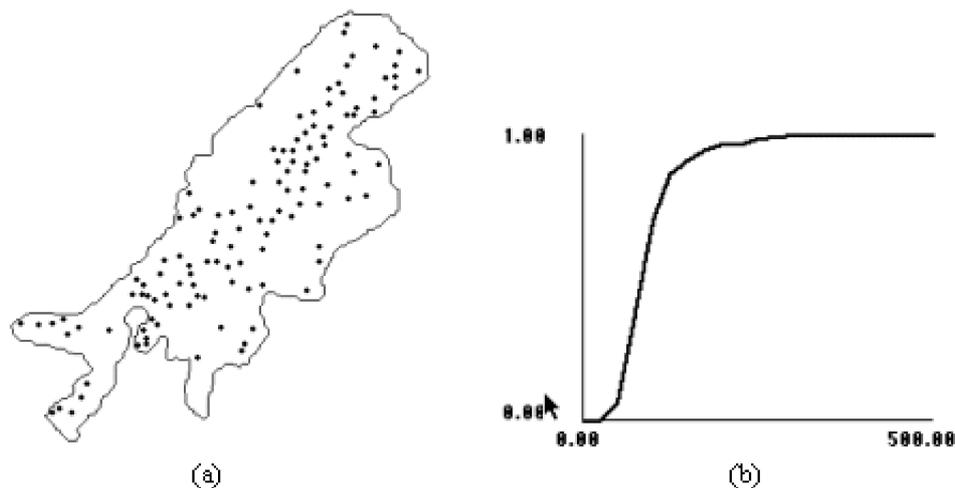


Figura 1: (a) Localização das crateras de vulcões em Uganda; (b) Distribuição baseada no vizinho mais próximo.

O procedimento abaixo (Algoritmo 2) mostra como se calcular o valor da distribuição G para uma dada distância w , usando a análise do vizinho mais próximo:

```

Dada uma distância  $w$ , Faça  $total = 1$ 
Para cada evento  $e$ 
  Início
    Calcule a distância  $dist$  ao vizinho mais próximo de  $e$ 
    Se  $dist \leq w$  incremente  $total$  de 1
  Fim
Divida  $total$  pelo número de eventos  $n$ 
Faça  $G(w) = total$ 

```

Algoritmo 2: Cálculo da distribuição G baseado nos vizinhos mais próximo dos eventos.

A análise de vizinhança pode ser usada como método formal para se comparar estatisticamente a distribuição dos eventos observados com o que se esperaria na hipótese da aleatoriedade espacial completa (CSR). Esta metodologia consiste em se criar envelopes de simulação para a distribuição CSR, a fim de se acessar a significância dos desvios. A função simulada para a distribuição $G(w)$ assumindo-se CSR é calculada como:

$$\bar{G}(w) = \sum G_i(w)/m$$

onde:

$G_i(w)$, $i = 1, 2, \dots, m$ = funções de distribuição simuladas desconsiderando os efeitos da borda da região.
 m = número de simulações.

Os envelopes de simulação superior e inferior são definidos como se segue:

$$U(w) = \max_{i=1, \dots, m} \{G_i(w)\} \quad L(w) = \min_{i=1, \dots, m} \{G_i(w)\}$$

A plotagem da distribuição estimada  versus a distribuição simulada , com a adição dos envelopes inferior e superior, permite se acessar a significância dos desvios relativo a aleatoriedade. Assumindo-se CSR a plotagem de  versus  deve ser praticamente linear com uma ângulo de 45 graus. Se o dado apresenta tendências para agrupamentos, os traçados no gráfico estarão acima da linha de 45 graus, ao passo que para padrões de regularidade os traçados ficarão abaixo da linha de 45 graus. A Figura 2 mostra um exemplo de gráfico mostrando o posicionamento da distribuição e dos envelopes com relação a linha de 45 graus, para os dados referentes a crimes juvenis na região de Cardiff (Herbert, 1980). Neste caso percebe-se a posição dos envelopes e da distribuição acima da linha de 45 graus, o que caracteriza agrupamento para as distâncias em análise.

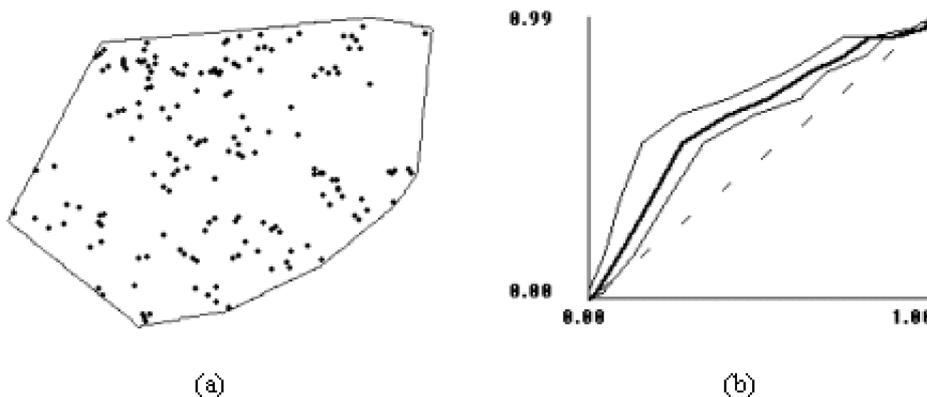


Figura 2: a) Localização de crimes cometidos por juvenis em Cardiff; b) Vizinho mais próximo com os envelopes.

O procedimento para a obtenção da distribuição simulada assim como dos envelopes usando-se a análise do vizinho mais próximo está descrito a seguir (Algoritmo 3):

Para 1 até n simulações

Início

Obtenha um conjunto s de m eventos aleatórios na região R

Calcule a *distribuição acumulada* (h) do conjunto s para as distâncias h

Incremente a *distribuição simulada* (h) com a *distribuição acumulada* de s

Se a *distribuição acumulada* (h) < *envelope inferior* (h)
então *envelope inferior* (h) = *distribuição acumulada* (h)

Se a *distribuição acumulada* (h) > *envelope superior* (h)
então *envelope superior* (h) = *distribuição acumulada* (h)

Fim

Divida *distribuição simulada* por n

Calcule *distribuição acumulada* a partir do conjunto original de eventos

Plote *distribuição acumulada*, *envelope inferior*, *envelope superior* versus *distribuição simulada*.

Algoritmo 3: Cálculo da distribuição simulada e dos envelopes usando o método do vizinho mais próximo.

Embora este método do vizinho mais próximo forneça uma boa indicação da distribuição espacial, ele considera apenas escalas pequenas. Para se ter informação mais efetiva para o padrão espacial em escalas maiores, o melhor método a ser utilizado é o da função K .

2.2 Função K :

A função K , também denominada medida de momento de segunda ordem reduzido, é definida para o processo univariado como:

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

onde $\#$ está associado ao número de eventos, $E()$ é o operador de estimativa, e λ é a intensidade ou número médio de eventos por unidade de área, assumida constante na região. Uma estimativa de $K(h)$ é:

$$\bar{K}(h) = \frac{R}{n^2} \sum_{i \neq j} \sum \frac{i_k(d_{ij})}{w_{ij}}$$

onde R é a área da região, n é o $\#$ de eventos observados, $i_k(d_{ij})$ é uma função indicatriz cujo valor é 1 se $(d_{ij}) \leq h$ e 0 em caso contrário, e w_{ij} é a proporção da circunferência do círculo centrado no evento i que está dentro da região. Portanto, w_{ij} é o que se denomina correção devido ao efeito de borda na região.

A função K é usada como ferramenta exploratória na comparação da estimativa empírica de $K(h)$ para os eventos e de $K(h)$ resultante de um processo de padrão de pontos espacial aleatório. Para um processo aleatório $K(h)$ seria ph^2 . Portanto, uma forma de comparar a estimativa $K(h)$ de um conjunto de dados observados com ph^2 seria plotar a função $L(h)$ versus h onde $L(h)$ é definido como:

$$\bar{L}(h) = \sqrt{\frac{\bar{K}(h)}{\pi}} - h$$

A plotagem de $L(h)$ versus a distância h indica atração espacial entre eventos ou agrupamentos para valores positivos, sendo o agrupamento mais forte em picos positivos, e indica repulsão espacial ou regularidade em pontos de valores negativos. O procedimento para se estimar $L(h)$ é o seguinte:

```

Para todo evento  $i$  em uma região  $R$  dada uma distância  $h$ 
Início
  Para todo evento  $j \neq i$  em  $R$ 
  Início
    Encontre distância entre  $j$  e  $i$ 
    Se (distância  $\leq h$ )
    Início
      Se não há correção de bordas então incremente de 1 o contador
      Senão
      Início
        Gere círculo centrado em  $i$  com raio igual a distância (sómente uma vez)
        Calcule a porcentagem da área do círculo que pertence a região  $R$ 
        Incremente o contador com o inverso da porcentagem
      Fim
    Fim
  Fim
Fim
Multiplique contador pela área da região e divida pelo quadrado do número de eventos
Aplique a raiz quadrada do contador dividida por  $\pi$  e subtraia  $h$  para se obter  $L(h)$ 

```

Algoritmo 4: Cálculo da função L .

Uma abordagem similar a do vizinho mais próximo pode ser feita para se acessar a significância dos desvios da distribuição $L(h)$ em relação a aleatoriedade (CSR). Os envelopes inferior e superior são construídos a partir de m simulações independentes de n eventos na região R . Na análise do gráfico com a distribuição e os envelopes, picos positivos na função estimada $L(h)$ que estão acima do envelope superior evidenciam ocorrência de agrupamento na escala considerada, portanto, se toda função $L(h)$ estiver acima do envelope superior e com valores positivos, teremos agrupamentos em todas as escalas. Depressões negativas na função estimada $L(h)$ que estiverem abaixo do envelope inferior, evidenciam regularidade nessa escala, portanto, se toda função $L(h)$ estiver abaixo do envelope inferior e com valores negativos, tem-se regularidade em todas as escalas. O procedimento (Algoritmo 5) para o cálculo da função de simulação de $L(h)$ e dos envelopes é o seguinte:

```

De 1 até  $n$  simulações
Início
  Obtenha  $m$  eventos aleatórios contidos na região  $R$ 
  Calcule  $L(h)$  usando os eventos  $m$  e as distâncias  $h$ 
  If the  $L(h) < lower\ envelope$  then  $lower\ envelope(h) = L(h)$ 
  If the  $L(h) > upper\ envelope$  then  $upper\ envelope(h) = L(h)$ 
End
Calcule  $L(h)$  para o conjunto original de eventos
Plote  $L(h)$ , e os envelopes inferior e superior versus as distâncias  $h$ 

```

Algoritmo 5: Cálculo dos envelopes de simulação baseado na função K .

A Figura 3 mostra a plotagem da função L e dos envelopes para o dado de Cardiff (Figura 2.a). Verifica-se valores positivos para a função L , estando os mesmos acima dos envelopes, o que caracteriza agrupamento nesta escalas de distância. O agrupamento é mais forte para as distâncias que correspondem aos extremos da curva de L .

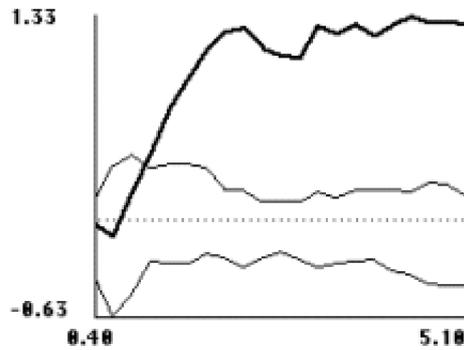


Figura 3: Função L e envelopes para dados da Figura 2.a.

3. Integração ao Software SPRING:

A ferramenta computacional para avaliar efeitos de segunda ordem em dados pontuais está implementada usando-se a linguagem de programação C++ e foi acrescentada ao software SPRING (INPE, 1996). O SPRING (Sistema de Processamento de Informações Georeferenciadas) é um software desenvolvido pelo Divisão de Processamento de Imagens do INPE, e que está em domínio público. O modelo de dados do SPRING está representado na Figura 4 e está dividido em 4 níveis de abstração:

- ☐ Nível do mundo real, que possui as categorias de dados a serem modeladas;
- ☐ Nível matemático ou conceitual, que contém as definições formais para as entidades. Este nível está baseado nos conceitos de campos e objetos geográficos .
- ☐ Nível de representação, que mapeia as entidades dos dados em representações geométricas que podem variar em função da escala da informação;

Nível de implementação, que possui as estruturas de dados e algoritmos para manipular os dados.

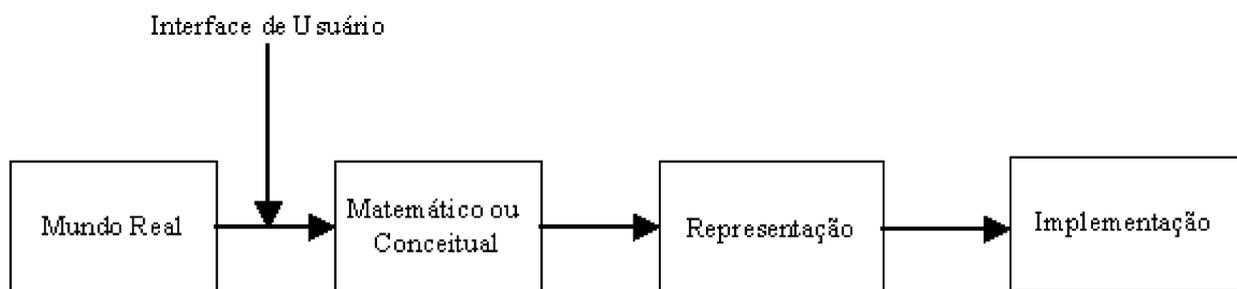
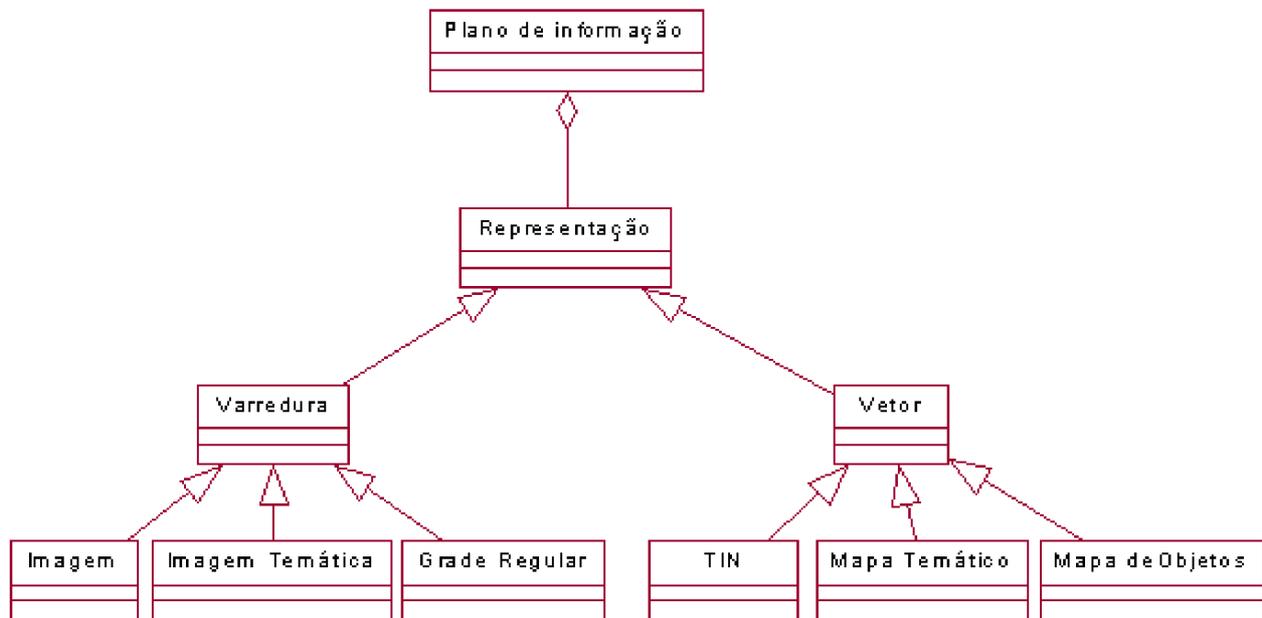
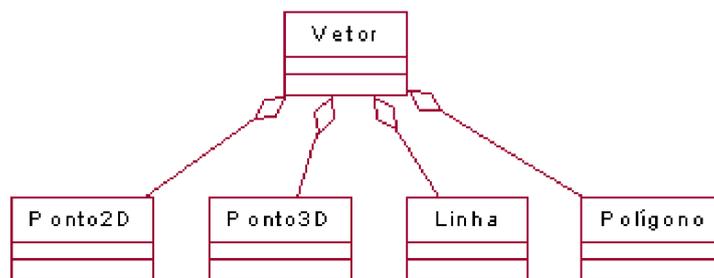


Figura 4: Níveis de abstração do SPRING.

Um plano de informação pode ter diferentes formas de representação para os seus dados. A Figura 5.a mostra a estrutura geral das representações no sistema SPRING. Estas representações podem ser do tipo varredura (correspondem a imagens e grades), ou do tipo vetor (rede triangular para modelo numérico de terreno, mapa de classes, mapa de objetos). A representação vetorial é composta pelas primitivas geométricas básicas ponto, linha e polígono. A Figura 5.b mostra alguns dos elementos que fazem parte da representação vetorial, tipo pontos bi-dimensionais, pontos tri-dimensionais, linhas, e polígonos. Por exemplo, um plano de informação referente a Divisão Política pode ter uma representação vetorial composta por linhas e polígonos que delimitam cada Estado ou região, assim como pode conter uma lista de pontos bi-dimensionais que representam determinada característica do dado (por exemplo, localização das capitais). No procedimento de análise do padrão de pontos no SPRING, a região é obtida a partir do polígono que a delimita e os pontos que representam os eventos, são obtidos da estrutura de pontos bi-dimensionais.



(a)



(b)

Figure 5: (a) Esquema de representações no SPRING; (b) Elementos de uma representação vetorial.

3.1 Interface de Programação

As funções de análise de padrão de pontos desenvolvidas neste trabalho usam a estrutura de representação de pontos bi-dimensionais existente na representação vetorial do SPRING. Os métodos do vizinho mais próximo e da função K incluindo as simulações, estão associados com a lista de pontos extraídas da estrutura de pontos bi-dimensionais do SPRING. As interfaces de programação destes métodos são as seguintes:

a) Vizinho mais próximo:

```
float* Point2dList::CumulativeDistribution (double distmin,double distmax,short intervalos) ;
```

Parâmetros de Entrada:

distmin: distância mínima.

distmax: distância máxima.

intervalos: número de intervalos entre distâncias mínima e máxima.

Parâmetro de Saída:

retorna ponteiro para um vetor de valores reais contendo a distribuição cumulativa para cada intervalo de distância.

b) Vizinho mais próximo com simulação:

float* Point2dList::NearestNeighborCSR (double distmin,double distmax,short intervalos, Polygon &poly,short simulações) ;

Parâmetros de Entrada:

distmin: distância mínima.

distmax: distância máxima.

intervalos: número de intervalos entre distâncias mínima e máxima.

poly: polígono que delimita a área da região em análise (pontos aleatórios gerados devem cair dentro desta região).

Parâmetro de Saída:

retorna ponteiro para um vetor de valores reais contendo a distribuição cumulativa estimada e os envelopes inferior e superior.

c) Função K:

float* Point2dList::LFunction (double distmin,double distmax,short intervalos,Polygon &poly,short borda) ;

Parâmetros de Entrada:

distmin: distância mínima.

distmax: distância máxima.

intervalos: número de intervalos entre distâncias mínima e máxima.

poly: polígono que delimita a área da região em análise (pontos aleatórios gerados devem cair dentro desta região).

borda: indica se borda é considerada na análise ou não.

Parâmetro de Saída:

retorna ponteiro para um vetor de valores reais contendo o resultado para a função L para os intervalos de entrada.

d) Função K com simulação:

float* Point2dList::LFunctionCSR (double distmin,double distmax,short intervalos,Polygon &poly, short simulações,short borda) ;

Parâmetros de Entrada:

distmin: distância mínima.

distmax: distância máxima.

intervalos: número de intervalos entre distâncias mínima e máxima.

poly: polígono que delimita a área da região em análise (pontos aleatórios gerados devem cair dentro desta região).

simulações: número de simulações.

borda: indica se borda é considerada na análise ou não.

Parâmetro de Saída:

retorna ponteiro para um vetor de valores reais contendo o resultado para a função L e os envelopes inferior e superior considerando os intervalos de entrada.

3.2 Interface com o usuário:

A figura 6 mostra os componentes básicos da interface de usuário do SPRING. Ao fundo existe um plano de informação já selecionado do modelo temático, com a visualização do contorno da área e dos pontos que o plano possui. A opção do menu de Análise de Pontos fica disponível para ser usada.

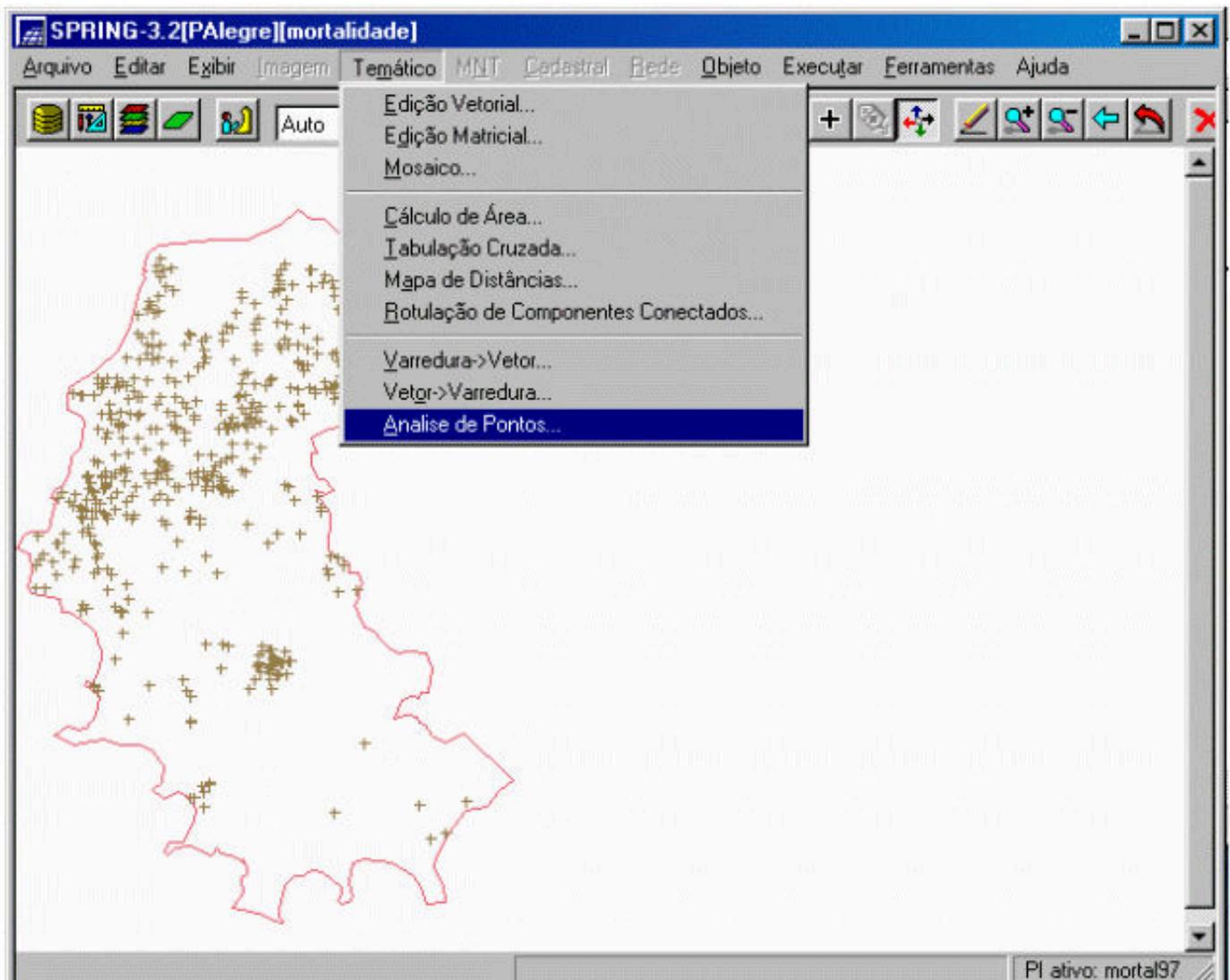


Figura 6: Menu principal do Spring e entre as opções disponíveis no caso do modelo temático a opção Análise de Pontos.

Ao se acionar a opção Análise de Pontos, uma nova janela (Figura 7.a) é apresentada, e as opções de análise pelo vizinho mais próximo ou pela função K, com ou sem simulação ficam disponíveis (Figura 7.b). Os parâmetros a serem fornecidos pelo usuário correspondem ao intervalo de distância a ser considerado, fornecendo-se as distâncias mínima e máxima, ao número de intervalos em que a faixa de distância será dividida, e no caso de modelagem incluindo simulações, deve-se fornecer também o número de simulações desejadas.

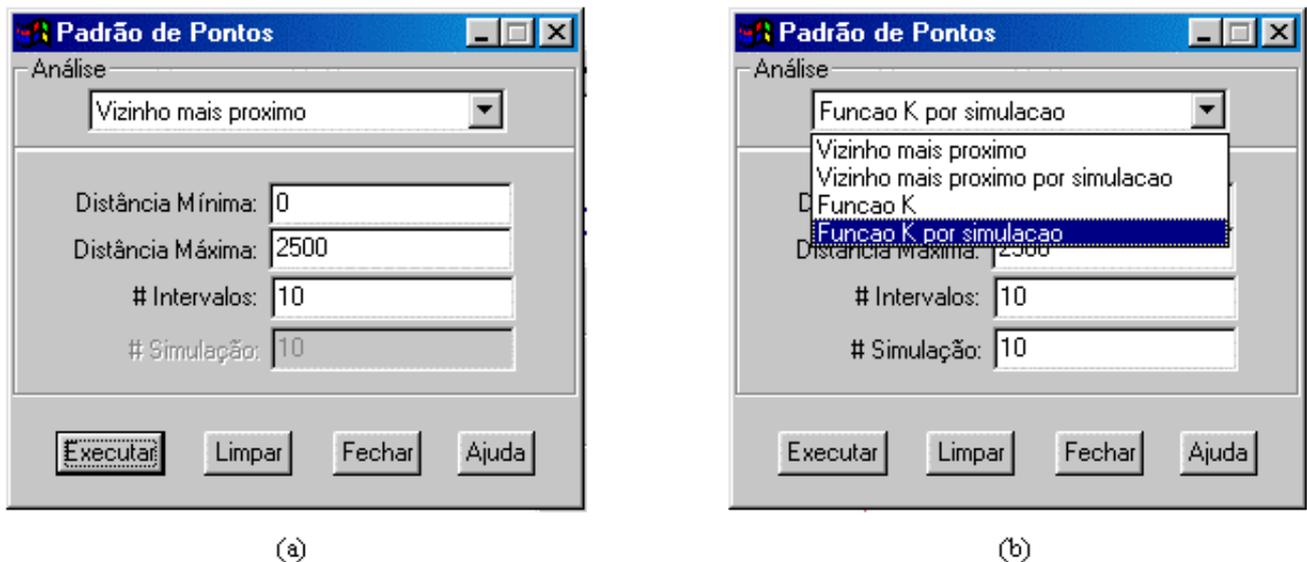


Figura 7: a) interface da análise de pontos no SPRING; b) opções de análise de pontos no SPRING.

4. Exemplos de Aplicações

Nos próximos sub-itens apresentam-se alguns exemplos de conjunto de padrões de dados para explorarmos os métodos apresentados. É importante deixar claro que esses dados são utilizados apenas para demonstração da potencialidade dos métodos, não estando os mesmos vinculados diretamente a qualquer projeto específico de análise espacial.

4.1. Mortalidade em 1997 na cidade de Porto Alegre:

Os dados de mortalidade ocorridas em 1997 foram gerados pela PROCEMPA e colocados a nossa disposição pela Secretaria Municipal de Saúde (SMS) de Porto Alegre. A Figura 8 mostra as localizações de 558 óbitos registrados no ano de 1997 de pessoas nascidas em 1996 na cidade de Porto Alegre, Rio Grande do Sul. A análise de padrões neste tipo de dado pode ser utilizada como uma forma de identificação de possíveis áreas com maior concentração de mortes infantis. A identificação de agrupamento nestes dados pode levar a uma análise mais detalhada de quais outros fatores possam estar relacionados com as mortes infantis nos respectivos locais.

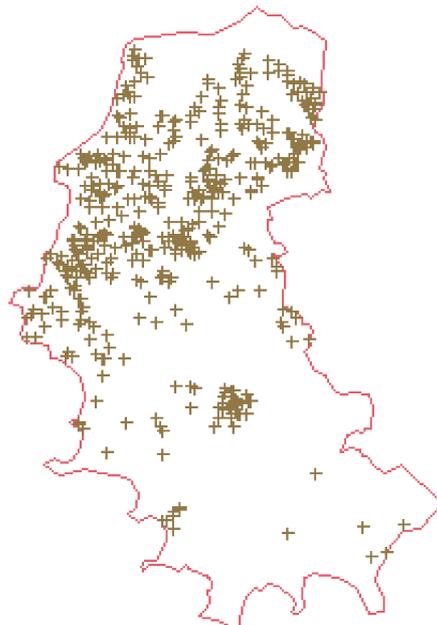
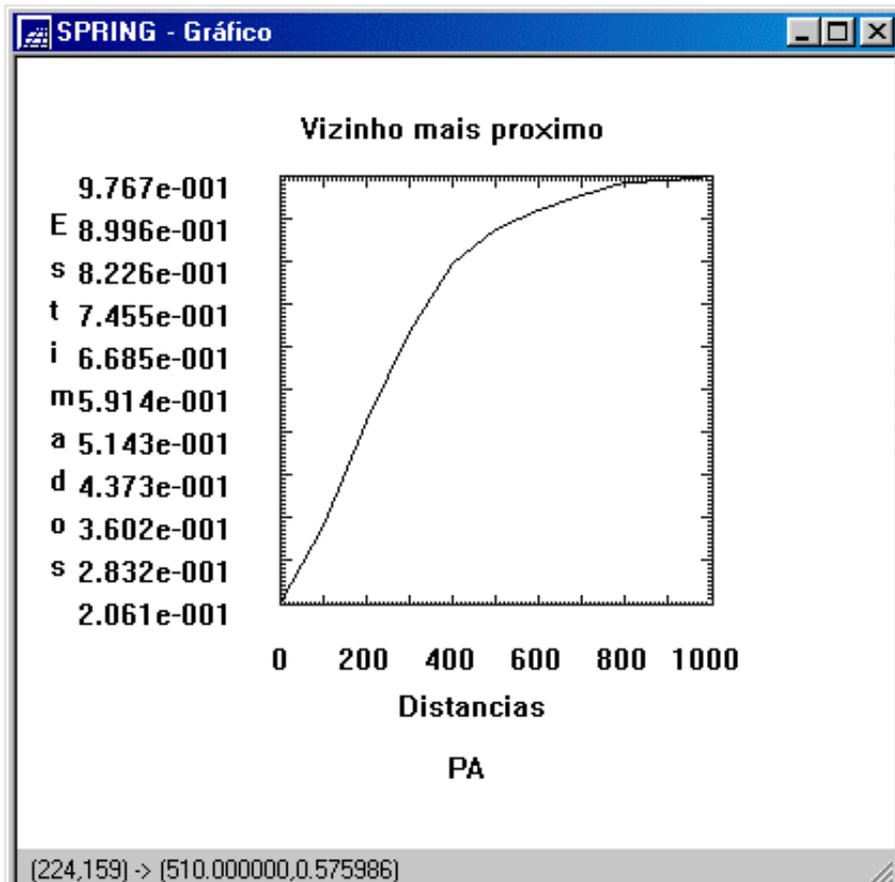


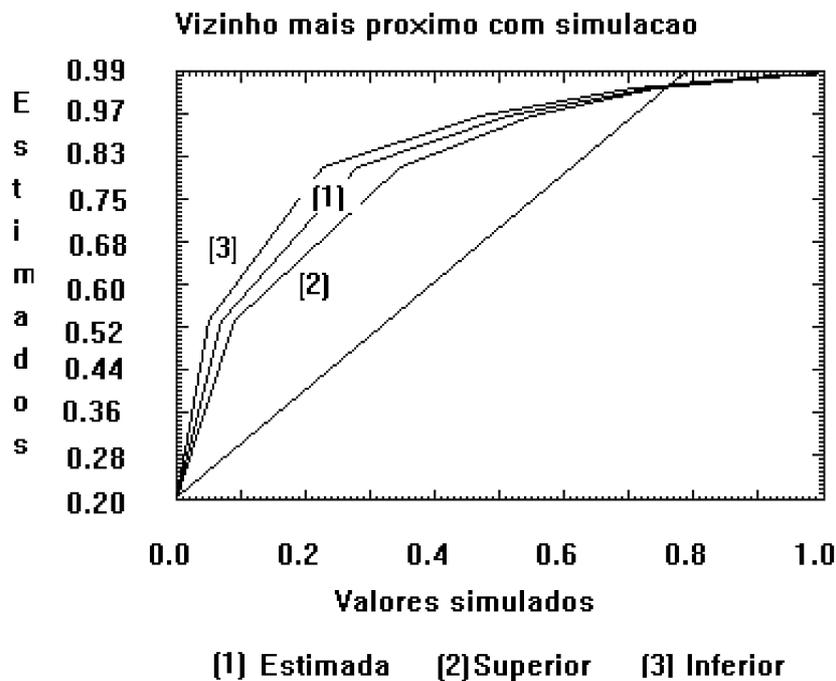
Figura 8: Mortalidade de 1997 em Porto Alegre (nascidos em 1996).

Uma análise exploratória nesses dados utilizando o método do vizinho mais próximo com distância mínima de 0 km e distância máxima de 1 km resulta na distribuição acumulada da Figura 9.a e permite que se façam algumas considerações. Verifica-se que a curva mostra um crescimento acentuado para distâncias até 500 m

para depois se estabilizar, o que caracteriza agrupamento nesta faixa de distâncias. A Figura 9.b é uma evidência da não aleatoriedade deste dado já que a curva de distribuição e seus envelopes estão acima da reta de 45 graus, o que caracteriza o agrupamento. O resultado da função K (Figura 9.c) também evidencia o agrupamento, já que os valores da distribuição da função L apresentam valores positivos, e estão acima dos envelopes simulados.



(a)



(b)

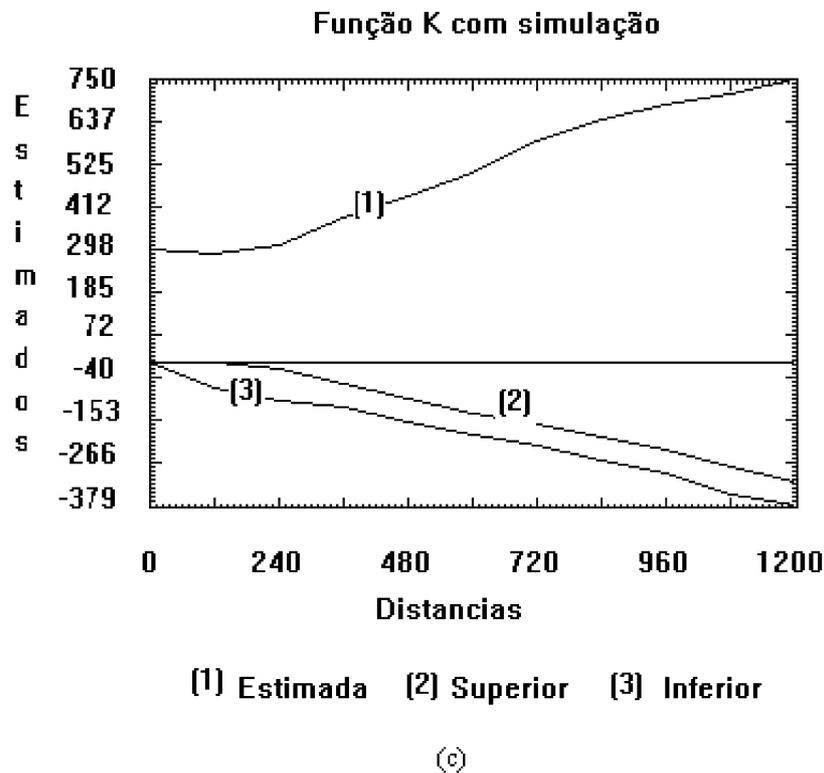


Figura 9: Mortalidade: (a) Vizinho mais próximo; (b) Vizinho mais próximo com simulação; (c) Função K com simulação.

4.2 Cidades de São Paulo

Este dado corresponde à localização de 180 cidades do Estado de São Paulo que possuem mais de 20000 habitantes (Figura 10). Uma análise dos padrões de localização baseada na variável população, pode levar a identificação de agrupamentos em áreas que tenham sido influenciadas por outros fatores tipo atividades industriais, agrícolas, ou algum outro fator.

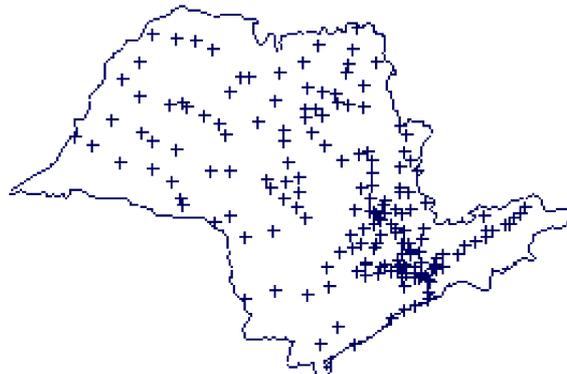


Figura 10: Municípios do Estado de São Paulo com mais de 20000 habitantes.

Uma análise exploratória nesses dados utilizando o método do vizinho mais próximo, com distância mínima de aproximadamente 0 km e máxima de 58 km, intervalo 10 e 20 simulações, mostra que existe um agrupamento mais significativo na faixa que vai até 40 km (Figura 11.a). Ou seja, este resultado caracteriza a interação espacial entre as localidades dentro desta faixa. O gráfico da função K com simulação (Figura 11.b) confirma a hipótese de agrupamento uma vez que a distribuição da função L apresenta valores positivos e acima dos envelopes simulados.

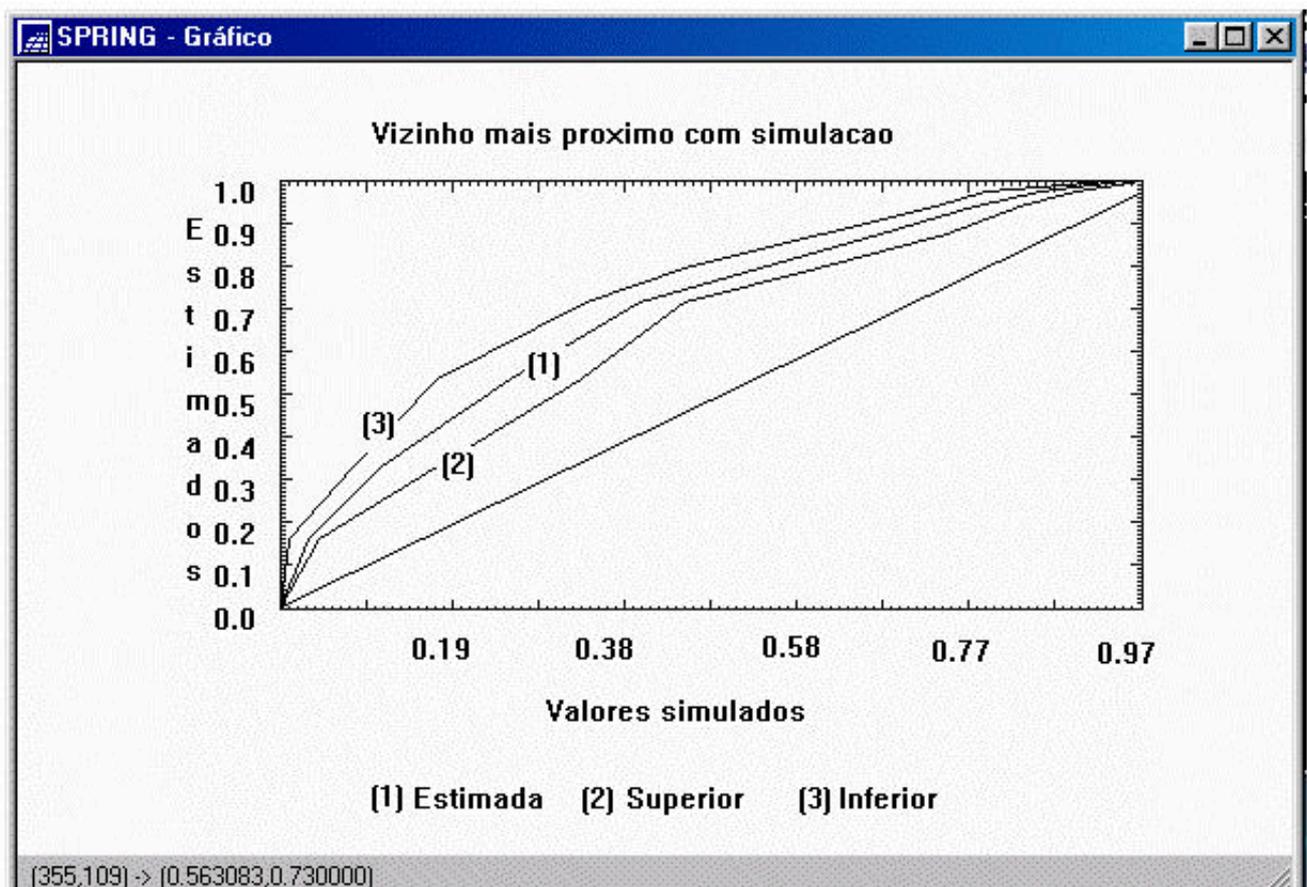
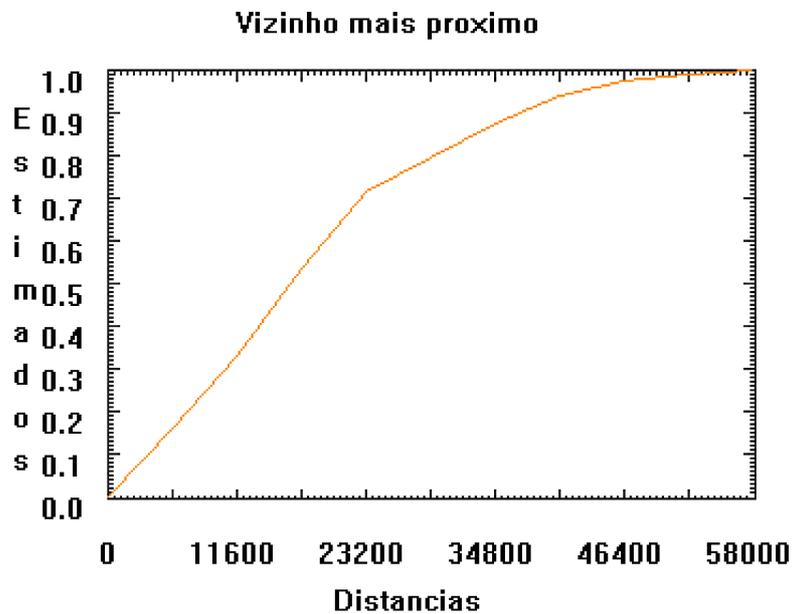


Figura 11: População no Estado de São Paulo: (a) vizinho mais próximo; (b) vizinho mais próximo com simulação.

4.3 Fazenda Canchim

Esses dados correspondem a 85 localizações que apresentam a presença de argila, adquiridas na Fazenda Canchim, em São Carlos, SP. A Figura 12 apresenta a distribuição das observações nesta área de estudo. A identificação de agrupamentos de minerais em uma região pode levar a análise futuras de outras variáveis que possam estar relacionadas com este elemento em estudo.

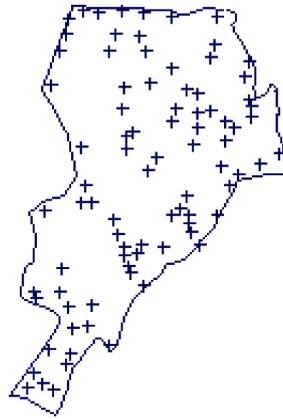


Figura 12: Fazenda Canchim (São Carlos - SP) com as localizações de argila.

Uma análise exploratória utilizando o método do vizinho mais próximo com distância mínima 0 e máxima de 900m, considerando 10 intervalos, mostra a tendência a agrupamentos com o aumento da distância (Figura 13). A relação entre a distribuição estimada e a distribuição simulada apresenta característica aleatória no começo da faixa de distâncias (curva (1) Figura 13), passando a ter uma característica de agrupamento com o aumento das distâncias. Neste caso os envelopes e a curva do relacionamento entre a distribuição estimada e a simulada, ficam acima da linha de 45 graus.

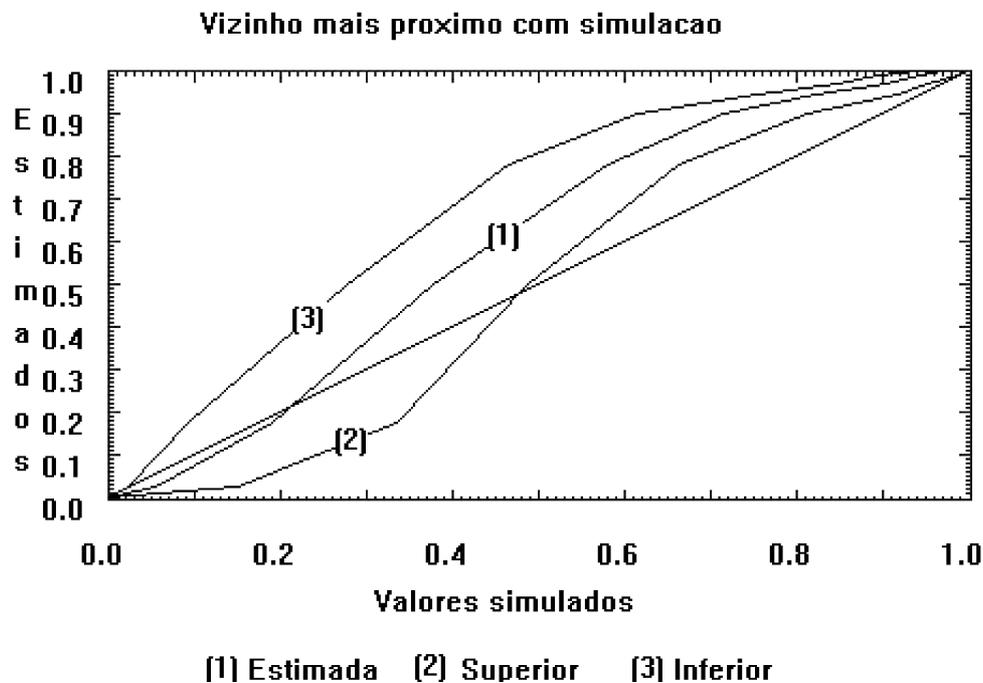


Figura 13: Fazenda Canchim: vizinho mais próximo com simulação.

5. Conclusão

Um conjunto de ferramentas computacionais para analisar efeitos de segunda ordem ou dependência espacial em dados pontuais univariados, foi implementado usando a linguagem de programação C++, e integrado ao software SPRING. Os componentes deste programa incluem ferramentas exploratórias baseadas na análise do vizinho mais próximo e da função K. Simulações podem ser realizadas com os dados a fim de se verificar a significância dos desvios relativo a aleatoriedade. A implementação está baseada no modelo orientado a objetos do SPRING que facilita o entendimento das operações assim como futuras expansões. Este trabalho se concentrou na análise de pontos univariados, ou seja se considerou apenas eventos de um tipo. Uma futura extensão é a inclusão de outro tipo de variável nos processos de análise a fim de se verificar a dependência espacial entre dois tipos de eventos diferentes. Por exemplo, pode-se comparar dados de doenças com os dados do nível de qualidade da água em uma região para se analisar se as condições da água possam estar relacionadas com estas doenças. A implementação desenvolvida suporta a análise considerando-se os efeitos da borda da região nos resultados finais, entretanto testes mais detalhados precisam ser feitos. O custo computacional incluindo este efeito de bordas cresce muito e as vezes não contribui no resultado final principalmente no caso de pequenas variações de escalas de distância. A inclusão destas ferramentas exploratórias no SPRING é certamente uma contribuição a mais ao conjunto de ferramentas de análise já

existentes.

6. Agradecimentos

Agradecimentos especiais vão para Rui Flores da Secretaria Municipal de Saúde de Porto Alegre pela disponibilização dos dados de mortalidade de 1997.

7. Referências

Bailey, C. Trevor; Gatrell, Anthony C. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical, John Wiley & Sons Inc, New York.

Goodchild, M. (1992). *Geographical Information Science*. *International Journal of Geographical Information and Analysis* 6(1): 31-45.

Herbert, D. T. (1980). The British experience. In Georges-Abeyie, D. E. e Harries, K. D. (Eds.), *Crime: a Spatial Perspective*, Colúmbia University Press.

SPRING. (1996) - Sistema de Processamento de Informações Georeferenciadas, Instituto Nacional de Pesquisas Espaciais, INPE, <http://www.inpe.br/spring>.

Tinkler, K. J. (1971). Statistical analysis of tectonic patterns in areal volcanism: the Bunyaruguru volcanic field in west Uganda, *Mathematical Geology* 3:335-355.

| [Home](#) | [Página Principal](#) | [Módulo Mix](#) | [Módulo Usuários](#) |