

ANÁLISE DE TÉCNICAS DE DISTRIBUIÇÃO ESPACIAL COM  
PADRÕES PONTUAIS E APLICAÇÃO A DADOS DE ACIDENTES  
DE TRÂNSITO E A DADOS DE DENGUE DE RIO CLARO-SP

Marcia Tiemi Kawamoto

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU  
São Paulo - Brasil  
Fevereiro – 2012

ANÁLISE DE TÉCNICAS DE DISTRIBUIÇÃO ESPACIAL COM  
PADRÕES PONTUAIS E APLICAÇÃO A DADOS DE ACIDENTES  
DE TRÂNSITO E A DADOS DE DENGUE DE RIO CLARO-SP

Marcia Tiemi Kawamoto

Orientador: Prof. Dr. **José Silvio Govone**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU  
São Paulo - Brasil  
Fevereiro – 2012

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO DE AQUIS. E TRAT. DA INFORMAÇÃO  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP  
BIBLIOTECÁRIA RESPONSÁVEL: *SULAMITA SELMA CLEMENTE COLNAGO*

Kawamoto, Marcia Tiemi.

Análise de técnicas de distribuição espacial com padrões pontuais e aplicação a dados de acidentes de trânsito e a dados de dengue de Rio Claro-SP / Marcia Tiemi Kawamoto. – Botucatu : [s.n.], 2012

Dissertação (mestrado) – Universidade Estadual Paulista, Instituto de Biociências de Botucatu

Orientador: José Silvio Govone


Capes: 90194000

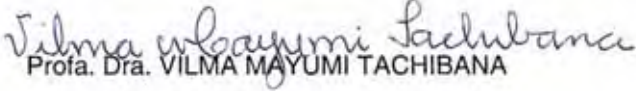
1. Dengue. 2. Acidentes de trânsito. 3. Análise espacial (Estatística). 4. Kernel, Funções de.

Palavras-chave: Análise estatística espacial; Análise de padrões pontuais; Estimador Kernel; Superfície de tendência.

**ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE Mestrado de MÁRCIA TIEMI KAWAMOTO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO(A) INSTITUTO DE BIOCÊNCIAS DE BOTUCATU.**

Aos 24 dias do mês de fevereiro do ano de 2012, às 14:00 horas, no(a) Depto. de Bioestatística, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. JOSE SILVIO GOVONE do(a) Departamento de Est Mat Apl Comp / Instituto de Geociências e Ciências Exatas, Profa. Dra. VILMA MAYUMI TACHIBANA do(a) Departamento de Matemática, Estatística e Computação / Faculdade de Ciências e Tecnologia de Presidente Prudente, Prof. Dr. PAULO MILTON BARBOSA LANDIM do(a) Depto. Bioestatística / IBB-UNESP, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE Mestrado de MÁRCIA TIEMI KAWAMOTO, intitulada "Análise de Técnicas de Distribuição Espacial com Padrões Pontuais e Aplicação à Dados de Acidentes de Trânsito e à Dados de Dengue de Rio Claro - SP". Após a exposição, a discente foi argüida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: Aprovado ----- . Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

  
Prof. Dr. JOSE SILVIO GOVONE

  
Profa. Dra. VILMA MAYUMI TACHIBANA

  
Prof. Dr. PAULO MILTON BARBOSA LANDIM

## Dedicatória

À meus pais pelo apoio incondicional.

## Agradecimentos

Agradeço, primeiramente, ao meu professor e orientador, o Prof. Dr. José Silvio Govone, por auxiliar na realização deste trabalho e por seu incentivo. Também, ao técnico do laboratório do Demac, Jorge Gustavo Falcão, pelo auxílio na parte computacional deste trabalho.

Agradeço à família Lima pelo convívio agradável em Rio Claro e também pelo tratamento carinhoso. Mais, ainda ao Filipe, por sempre estar ao meu lado, animando e acalmando nas horas de desânimo e confusão.

Agradeço aos meus familiares, que mesmo sem estarem presentes, mantiveram o apoio e confiança.

Agradeço à Secretaria de Mobilidade Urbana e Sistema Viário e à Secretaria de Segurança Pública do Município de Rio Claro-SP, por terem cedidos, gentilmente, seus dados para serem aplicados neste estudo.

Agradeço à Capes, pelo apoio financeiro.

Sou grata, também, a todos que contribuíram para realização deste trabalho ou apoiaram para realização deste.

# Sumário

	Página
<b>LISTA DE FIGURAS</b>	<b>5</b>
<b>LISTA DE TABELAS</b>	<b>8</b>
<b>RESUMO</b>	<b>9</b>
<b>SUMMARY</b>	<b>11</b>
<b>1 INTRODUÇÃO E REVISÃO DE LITERATURA</b>	<b>1</b>
1.1 Alguns conceitos estatísticos importantes . . . . .	4
1.2 Análise de Padrões Pontuais . . . . .	6
1.3 Método dos Quadrats . . . . .	9
1.4 Estudo da Dependência Espacial . . . . .	9
1.4.1 Método do Vizinho mais Próximo . . . . .	10
1.4.2 Função $K$ . . . . .	12
1.5 Estimação da Intensidade: Kernel . . . . .	16
1.6 Superfícies de Tendência . . . . .	21
<b>2 MATERIAL E MÉTODOS</b>	<b>23</b>
2.1 Conjuntos de dados . . . . .	23
<b>3 RESULTADOS E DISCUSSÃO</b>	<b>28</b>
<b>4 CONCLUSÕES</b>	<b>48</b>
4.1 Considerações finais . . . . .	49

	4
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>50</b>
<b>APÊNDICES</b>	<b>52</b>



## Lista de Figuras

	Página
1	Mapa de Londres com casos de cólera e poços. . . . . 3
2	Tipologia da distribuição de padrões pontuais. A) Aleatório. B) Aglomerado. C) Regular. (Diggle, 1983) . . . . . 6
3	Estimação da função K. Fonte:(Bailey & Gatrell, 1995) . . . . . 15
4	Passos para o cálculo de intensidade de pontos segundo a técnica de Kernel. Extraído de Brasil. Ministério da Saúde (2007) . . . . . 18
5	Distribuição de acidentes de trânsito em Rio Claro. Fonte: Secretaria de Mobilidade Urbana- P. M. Rio Claro-SP. . . . . 24
6	Distribuição de acidentes de trânsito em Rio Claro, com vítima ou não. Fonte: Secretaria de Mobilidade Urbana- P. M. Rio Claro-SP. . . . . 25
7	Distribuição de casos positivos de dengue em Rio Claro. Fonte: Secretaria de Segurança Pública - P.M. Rio Claro- SP. . . . . 26
8	Malha de pontos dos valores hipotéticos . . . . . 27
9	Estimação kernel quártico aplicado aos dados de acidentes de trânsito. A) raio de influência = 250m. B) raio de influência = 375m . . . . . 28
10	Estimação kernel quártico aplicado aos dados de acidentes de trânsito. A) raio de influência = 500m. B) raio de influência = 1000m . . . . . 29
11	Estimação kernel quártico aplicado aos dados de acidentes de trânsito, destacando a Avenida Rio Claro . . . . . 30
12	Comparação de funções kernel com raio de influência igual a 250m aplicado aos dados de acidentes de trânsito. A)Kernel Normal. B)Kernel Quártico . . . . . 31

13	Comparação de funções kernel com raio de influência igual a 375m aplicado aos dados de acidentes de trânsito. A)Kernel Normal. B)Kernel Quártico . . . . .	32
14	Estimação kernel normal aplicado aos dados de acidentes de trânsito. A) raio de influência = 100m. B) raio de influência = 125m. C)raio de influência = 150m. . . . .	33
15	Mapa de Estimação de kernel quártico com raio de influência igual a 250m aplicado a dados de acidentes de trânsito: A) sem vítimas; B) com vítimas.	33
16	Estimação kernel quártico aplicado aos dados de dengue. A) raio de influência = 250m. B) raio de influência = 375m . . . . .	34
17	Estimação kernel quártico aplicado aos dados de dengue. A) raio de influência = 500m. B) raio de influência = 625m . . . . .	34
18	Estimação kernel quártico aplicado aos dados de dengue ao raio de influência de 500m, destaque para algumas aglomerações. . . . .	35
19	Comparação de funções kernel com raio de influência igual a 375m aplicado aos dados de dengue. A)Kernel Normal. B)Kernel Quártico . . . . .	36
20	Comparação de funções kernel com raio de influência igual a 500m aplicado aos dados de dengue. A)Kernel Normal. B)Kernel Quártico . . . . .	36
21	Estimação kernel normal aplicado aos dados de dengue. A) raio de influência = 100m. B) raio de influência = 150m. C)raio de influência = 200m. . . . .	37
22	A)Mapa de estimação kernel quártico aplicado ao valor 1, com raio 2. B)Superfície de tendência linear aplicado ao valor1. . . . .	38
23	A)Mapa de estimação kernel quártico aplicado ao valor 2, com raio 2 . B)Superfície de tendência linear aplicado ao valor2. . . . .	38
24	A)Mapa de estimação kernel quártico aplicado ao valor 3, com raio 2 . B)Superfície de tendência quadrático aplicado ao valor3. . . . .	39

25	A) Mapa de estimação kernel quártico aplicado ao valor 4, com raio 2 . B) Superfície de tendência quártico aplicado ao valor 4. C) Superfície de tendência linear aplicado ao valor 4. . . . .	39
26	A) Mapa de resíduos aplicado ao valor 1. B) Mapa de resíduos aplicado ao valor 2. C) Mapa de resíduos aplicado ao valor 3. . . . .	40
27	A) Mapa de resíduos aplicado ao valor 4, supondo superfície quártica. B) Mapa de resíduos aplicado ao valor 4, supondo superfície linear. . . .	41
28	Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência linear considerando valor 1 . . . . .	42
29	Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência linear considerando valor 2 . . . . .	42
30	Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência quártica considerando valor 3 . . . . .	43
31	Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência linear considerando valor 4 . . . . .	43
32	Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência quártica considerando valor 4 . . . . .	44
33	Gráfico dos Resíduos do Kernel vs. superfície de tendência para os dados do valor 1 . . . . .	45
34	Gráfico dos Resíduos do Kernel vs. superfície de tendência para os dados do valor 2 . . . . .	45
35	Gráfico dos Resíduos do Kernel vs. superfície de tendência para os dados do valor 3 . . . . .	46
36	Gráfico dos Resíduos do Kernel vs. superfície de tendência linear para os dados do valor 4 . . . . .	46
37	Gráfico dos Resíduos do Kernel vs. superfície de tendência quártica para os dados do valor 4 . . . . .	47

## Lista de Tabelas

	Página
1	Comparação dos resíduos, quanto ao desvio padrão. . . . . 41
2	Comparação de resíduo do valor4, quanto ao desvio padrão. . . . . 41

**ANÁLISE DE TÉCNICAS DE DISTRIBUIÇÃO ESPACIAL COM  
PADRÕES PONTUAIS E APLICAÇÃO A DADOS DE ACIDENTES  
DE TRÂNSITO E A DADOS DE DENGUE DE RIO CLARO-SP**

Autora: MARCIA TIEMI KAWAMOTO

Orientador: Prof. Dr. JOSÉ SILVIO GOVONE

**RESUMO**

As análises estatísticas aplicadas a quaisquer fenômenos espaciais procuram compreender como ocorrem a distribuição espacial destes, além de tentar interpretar seus padrões.

A análise de padrões pontuais, foco deste trabalho, vai mostrar algumas técnicas que exploram o padrão pontual, em particular no estudo do estimador Kernel.

O estimador de intensidade kernel é um dos métodos mais utilizados para uma rápida análise do padrão de distribuição espacial de uma variável, concentrando-se na localização de possíveis aglomerados. Para tanto, aplicou-se esse método a dois conjuntos de dados referentes a cidade de Rio Claro-SP: um sobre

os acidentes de trânsito que ocorreram na cidade e o outro sobre casos positivos de dengue.

Uma comparação do desempenho do estimador com o ajuste de superfícies de tendências, para pequenos conjuntos de dados hipotéticos foi feita. Esta comparação não definiu exatamente uma melhor técnica, mas sugeriu que mais estudo sejam feitos.

A técnica estimador Kernel foi satisfatória na obtenção de uma visão tridimensional da variação dos fenômenos, fornecendo aglomerados e focos mais precisos das ocorrências. Com isso, tem-se uma melhor visualização de locais para ações de prevenção, no caso de acidentes de trânsito e possíveis eliminações de criadouros do mosquito transmissor de dengue, no caso dos positivos para dengue.

**ANALYSIS OF TECHNIQUES FOR SPATIAL DISTRIBUTION  
WITH POINT PATTERNS AND APPLICATION TO TRAFFIC  
ACCIDENTS DATA AND TO CASE OF DENGUE DATA FROM RIO  
CLARO-SP**

Author: MARCIA TIEMI KAWAMOTO  
Adviser: Prof. Dr. JOSÉ SILVIO GOVONE

**SUMMARY**

The statistical analysis applied to any spatial phenomena try to understand how the spatial distribution of these occur, and still try to interpret its patterns.

The point pattern analysis, the focus of this work will show some techniques that explore the point pattern, particularly, in the study of the Kernel estimator.

The kernel intensity estimator is one of the most commonly used methods for quick analysis of spatial distribution pattern of a variable, focusing on

the location of possible clusters. To this end, we applied the technique to data from the city of Rio Claro-SP: a first set is about traffic accidents that occurred in the city and the second is a set of data is about positive cases of dengue.

A comparison of the performance of the estimator, adjusting for trend surfaces for small sets of hypothetical data was made. This comparison did not exactly established a better technique, but suggested that further study be made.

The Kernel estimator technique was satisfactory in getting a view three dimensional of the variation of the phenomena, providing more accurate clusters and outbreaks of cases. Thus, we have a better view of local prevention, in the case of traffic accidents and possible elimination of mosquito breeding sites of dengue transmitter in the case of positive for dengue.



# 1 INTRODUÇÃO E REVISÃO DE LITERATURA

Obtido um conjunto de dados espacialmente distribuídos, procura-se estimar parâmetros e ajustar modelos de probabilidade, com o objetivo de conhecer o comportamento probabilístico da variável na área de estudo e fazer previsões.

A análise de fenômenos espacialmente distribuídos ressalta características diretamente ligadas a localização espacial dos dados e produz resultados mais significativos que uma análise realizada ignorando-se a dimensão espacial. Mais precisamente, segundo Bailey & Gatrell (1995), os dados em questão estão disponíveis em algum processo que está operando no espaço e essa análise procura explicar ou descrever o comportamento desse processo e suas possíveis relações com outros fenômenos espaciais. Os fenômenos espaciais assim descritos podem ocorrer em diferentes áreas do conhecimento, tais como: epidemiologia, geologia, ecologia e agronomia.

O maior interesse nessas ocorrências é a compreensão e interpretação dos padrões dessas distribuições espaciais. Segundo Câmara et al. (2004), além da percepção visual do fenômeno, é muito útil também conhecer os padrões existentes e fazer considerações objetivas e mensuráveis.

A Humanidade vem estudando padrões espaciais há séculos, mas a necessidade de reduzir tais informações a números é recente. O olho e o cérebro

humano formam um mecanismo maravilhoso com o qual analisam e reconhecem padrões, mas eles são subjetivos, suscetíveis a fadiga e assim, também, a erros (Ripley, 2004). Assim, o estudo dos mesmos torna-se frequente e desafiador.

Um dos exemplos pioneiros com enfoque espacial, segundo Câmara et al. (2004), foi o trabalho desenvolvido pelo médico inglês John Snow, no século XIX. Ele fez a associação da cólera com ingestão/consumo de água insalubre. Em 1854, Londres estava com uma epidemia de cólera. Como Snow sabia que a doença ocorria em certas localizações, então com o mapa de Londres foi marcando a residência do indivíduo morto pela doença, informação obtida do registro de óbitos, e também a localização de poços de água existentes na rua. Assim ele pode visualizar que em torno do poço da Broad Street concentrava-se a maioria dos casos, e identificado a origem da contaminação, ordenou o fechamento desse poço, o que contribuiu para combater a epidemia. Este foi um dos primeiros relatos, onde a análise de dados espaciais teve um ação direta, contribuindo para identificação do epicentro da epidemia e também para diminuir a epidemia de cólera da época, através da eliminação da principal origem da doença, a água contaminada do poço.

O mapa de Snow, figura 1, foi um instrumento de visualização dos dados da cólera. Mas, neste caso não se fez qualquer análise estatística. Os pontos em vermelho são as residências dos casos de cólera e os balões azuis representam a localização dos poços de abastecimento de água. (A figura está disponível em <http://www.ufrj.br/institutos/it/de/acidentes/sr10.htm> )

O mapa obtido por Snow foi bastante evidente e mostrou essa associação entre água e cólera. Mas em diversos outros casos do dia-a-dia da Vigilância em Saúde, o padrão espacial dos pontos não está tão claro e precisa-se usar técnicas de análise espacial para demonstrar e testar os padrões (Brasil. Ministério da Saúde, 2007)

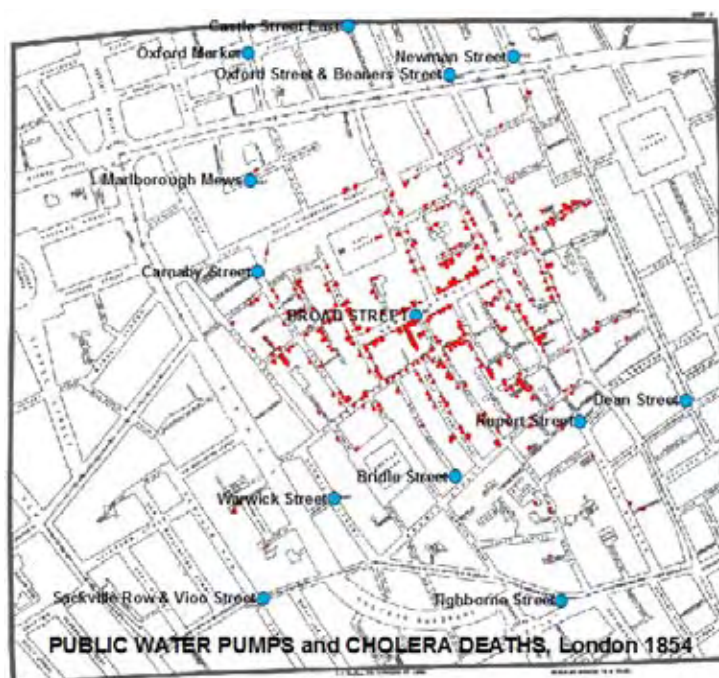


Figura 1 - Mapa de Londres com casos de cólera e poços.

Os fenômenos espaciais podem ser analisados conceitualmente em três categorias, de acordo com a natureza dos mesmos: dados pontuais, dados por áreas e dados contínuos.

- Dados por Área

Dados por área são aqueles cuja localização está associada a áreas delimitadas por polígonos, onde não necessariamente se tem a localização exata do evento, mas um valor, que representa toda a área. Esse valor pode ser uma taxa por área, por exemplo, taxas de mortalidade e natalidade de determinada região, proporção de analfabetos nos municípios do Estado do Paraná, etc. Desse modo, as análises sobre esses dados são imprescindíveis nas áreas de saúde, em estudos demográficos, em atividades políticas e outras afins.

- Dados contínuos

Diferentemente de eventos discretos que estão associados a ocorrências pontuais, os estudos sobre esses tipos de dados recorrem a amostras da superfície, isto é, valores representativos do fenômeno na área de estudo. Segundo Bailey & Gatrell (1995), os fenômenos naturais, tais como temperatura, pressão atmosférica, características de solo, etc podem ser, em princípio, observados e medidos em qualquer parte da superfície terrestre.

A geoestatística é capaz de descrever a continuidade espacial e possui ferramentas para analisar as suas variações .

- Dados pontuais

Os fenômenos que geram dados pontuais estão intimamente ligados a localização espacial. Os dados estão na forma de um conjunto de pontos, distribuídos dentro de uma região de estudo, determinado por este fenômeno, por exemplo, a localização de árvores numa floresta ou a localização de ninhos de pássaros. Essa distribuição pode ser analisada e descrita por métodos de análise de processos pontuais, embora possam também ser aplicados, através da técnica de Krigagem indicativa.

## 1.1 Alguns conceitos estatísticos importantes

Antes de se iniciar os estudos na estatística espacial pontual, deve-se ter em mente alguns conceitos estatísticos espaciais importantes.

A dependência espacial é uma noção na qual a estatística espacial está fundamentada. Essa estrutura de dependência espacial pode ser detectada nos dados e quantificada, por medidas de autocorrelação espacial. Além disso, a variação

da dependência espacial é feita analisando a amostra com sua vizinhança.

Os dados espaciais, em geral, não formam um conjunto de dados independentes. Uma importante diferença em relação a dados sem essa peculiaridade é que cada observação não traz uma informação independente, e o conjunto de todas as observações é utilizado integralmente para descrever o padrão do fenômeno estudado (Brasil. Ministério da Saúde, 2007).

Segundo Cressie (1993), dados que estão mais próximos no espaço são mais parecidos que aqueles que estão mais distantes. Ou seja, os objetos mais próximos tendem a ser mais correlacionados e não podem ser tratados como independentes.

A estacionariedade é um outro conceito muito explorado dentro da estatística espacial. Ela diz respeito ao modelo espacial, ao processo que dá origem às ocorrências. Os processos estacionários são processos estocásticos que não são afetados pela passagem do tempo ou espaço; em termos estatísticos, a média e a variância do processo permanecem constante. Por exemplo, considerando o número de mortes por doença em alguma região, se a mesma for dividida em sub-áreas, ocorrerá o mesmo número de mortes por sub-área, ou seja, a média do processo será constante em toda região. A estacionariedade será estudada mais profundamente em análise de padrões pontuais.

Tem-se, também, o conceito de completa aleatoriedade espacial (Complete Spatial Randomness - CSR), que implica que no processo espacial, os eventos estão distribuídos de forma aleatória e cobrem uniformemente a área de estudo. Ou seja, não há locais onde o evento seja mais provável de ocorrer com maior ou menor frequência, nem que a ocorrência de um evento modifique a probabilidade de ocorrência dos outros eventos da vizinhança.

## 1.2 Análise de Padrões Pontuais

Os processos pontuais advem de fenômenos identificados como pontos localizados no espaço. Assim, os dados em estudo são formados por coordenadas dos eventos de interesse em alguma região  $R$  e sua representação é dada da forma,  $(s_1, s_2, \dots)$ , onde  $s_i$  representa a localização do  $i$ -ésimo evento. Segundo Câmara & Carvalho (2004), o termo evento refere-se a qualquer fenômeno localizável no espaço, que dentro da escala de investigação, possa estar associado a uma representação pontual. É possível, por exemplo, fazer uma distribuição espacial dos centros das células em uma unidade quadrada, como o exemplo da figura 2C documentada por Ripley (1977).

É de grande interesse verificar se o padrão de ocorrência do fenômeno na área de estudo é aglomerado (as ocorrências estão concentradas em regiões no espaço), regular (uniforme) ou aleatório. A figura 2 apresenta esses diferentes tipos de possibilidades.

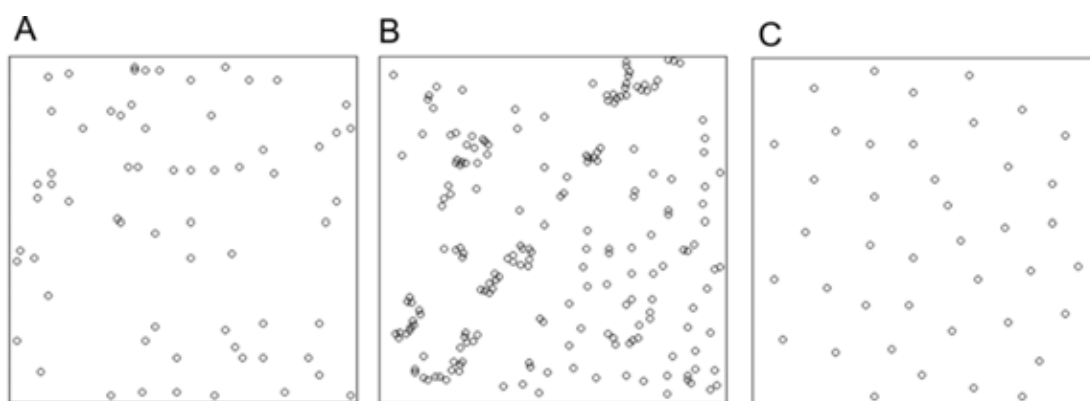


Figura 2 - Tipologia da distribuição de padrões pontuais. A) Aleatório. B) Aglomerado. C) Regular. (Diggle, 1983)

A figura 2A mostra a área quadrada de  $32,49m^2$  de 65 mudas de pinheiros japoneses do artigo de Numata (1961), a figura 2B traz 62 mudas de uma espécie de sequoia em uma área quadrada de  $529m^2$  do artigo de Ripley(1977) e a figura 2C mostra 42 núcleos de células biológicas também do artigo de Ripley(1977); todos os dados estão disponíveis em Diggle (1983).

No padrão aglomerado, os pontos tendem a ocorrer de forma agrupada (*cluster* ou "áreas quentes") na área (figura 2B); no padrão regular, a distribuição das ocorrências dos pontos na área é uniforme (figura 2C); enquanto que no padrão aleatório, as ocorrências se dão de forma aleatória (figura 2A). Este último consiste numa mistura dos dois padrões anteriores (Ward & Carpenter, 2000).

Caso haja uma fonte que cause a ocorrência dos dados, como é o caso do poço da Broad Street de Londres, no exemplo da ocorrência de cólera, a tendência é que estas se apresentem na forma aglomerada. O padrão uniforme, mais raro de aparecer, pode ocorrer. Por exemplo, na distribuição de escolas de ensino fundamental em uma cidade, onde a distribuição da população nos bairros é homogênea quanto à faixa etária e quanto ao número de crianças por família. Neste caso, espera-se que a distribuição das escolas seja aproximadamente homogênea. Como exemplo do padrão aleatório, pode-se citar a distribuição em uma cidade da ocorrência de uma doença não contagiosa e que não seja afetada por fatores ambientais, sociais ou demográficos.

Os processos pontuais podem ser vistos como resultado de um processo estocástico; então o comportamento espacial do fenômeno é, em geral, o resultado da combinação de efeitos de primeira e segunda ordem.

Os efeitos de primeira ordem, de escala global ou larga escala caracterizam-se pela variação do valor médio do processo, calculando a intensidade,

$\lambda(s)$ , do processo na região de estudo, isto é, o número médio de eventos por unidade de área.

Define-se  $Y(A)$  como sendo número de eventos que ocorrem em uma área arbitrária ou sub-área,  $A$ , de uma região de estudo,  $R$ . Desse modo, o processo pode ser representado pelo conjunto  $\{Y(A), A \subset R\}$ .

A intensidade é estimada por:

$$\lambda(s) = \lim_{ds \rightarrow 0} \frac{E(Y(\mathbf{ds}))}{ds},$$

onde  $\mathbf{ds}$  é uma pequena região em torno do ponto  $s$ ,  $ds$  é a área dessa região e  $E(\cdot)$  é o operador Esperança.

Os efeitos de segunda ordem, locais ou de pequena escala, remetem à dependência espacial, que é estimada pelo relacionamento de pares de eventos, o que se assemelha ao cálculo da covariância de variáveis aleatórias.

Isto pode ser formalmente descrito pela intensidade de segunda ordem,  $\gamma(s_i, s_j)$ , do processo, que envolve novamente eventos por unidade de área e é formalmente definido por:

$$\gamma(s_i, s_j) = \lim_{ds_i, ds_j \rightarrow 0} \frac{E(Y(\mathbf{ds}_i), Y(\mathbf{ds}_j))}{ds_i ds_j}$$

Um processo é dito estacionário numa região  $R$  se a média e variância do processo for constante, ou seja, a intensidade deste é constante,  $\lambda(s) = \lambda$  para todo ponto  $s$  em  $R$ . Isto é, não há tendências. O processo é estacionário se a covariância entre valores em quaisquer dois pontos depende da distância e direção entre ambos e não de suas localizações em  $R$ . O processo é isotrópico se for estacionário e a



dependência for em função de distância entre os eventos e não da direção.

É possível verificar o padrão espacial das distribuições considerando esses efeitos de primeira e segunda ordem, ou seja, em escalas globais ou considerando a interação dos eventos.

### 1.3 Método dos Quadrats

Um método bastante simples de sumarizar o padrão da distribuição é particionar a região  $R$  em sub-áreas,  $A_i$ , de tamanhos iguais ou em quadrats (quadrados ou retângulos) e observar quantos eventos ocorrem em cada sub-área. Em geral, supõe-se uma grade regular sobre  $R$  e conta-se o número de eventos nas unidades quadradas, obtendo-se, assim, uma medida de intensidade, a qual é o número de eventos pela área do quadrado.

Desse modo, tem-se um histograma bidimensional ou uma distribuição de frequências das ocorrências, com o que se pode ter indicação da variação da intensidade do processo por toda a região. Se fossem aleatórios teriam distribuição de Poisson.

Como obtem-se uma caracterização das sub-áreas em alta ou baixa intensidades, perde-se muitos detalhes do padrão espacial observado. Quadrats menores retem mais informação espacial, mas obtem-se alta variabilidade em quadrats, o que pode degenerar-se dentro de um mosaico com muitos quadrats vazios, podendo tornar a interpretação impossível (Bailey & Gatrell, 1995). Desse modo, essa técnica pode ser não muito precisa.

### 1.4 Estudo da Dependência Espacial

O método do vizinho mais próximo e a função  $K$  de Ripley são técnicas utilizadas para se estudar a dependência espacial de eventos, a qual está associada as propriedades de segunda ordem do processo pontual.

O método do vizinho mais próximo indica preliminarmente a distribuição espacial e trabalha apenas com escalas pequenas. Para as escalas maiores, a função de  $K$  de Ripley é mais eficiente para o estudo de padrão espacial dos eventos. (Câmara et al., 2004)

#### 1.4.1 Método do Vizinho mais Próximo

O método do vizinho mais próximo investiga as propriedades de segunda ordem usando as distâncias entre os eventos observados na região de estudo. Poder-se-ia medir a distância entre eventos de diferentes formas, mas interessa ao estudo as distâncias de vizinhos mais próximos. A distância de evento-evento trata-se da medida de um evento escolhido ao acaso a outro evento mais próximo, e denota-se por  $W$ . Já, a distância ponto-evento trata-se da medida de um ponto aleatório a um evento mais próximo, e denota-se por  $X$ .

A função  $G$  mede a distribuição da distância,  $W$ , de um evento arbitrário a seu evento vizinho mais próximo. Formalmente, seja  $w_i = \min\{d_{ij}, \forall j \neq i\}$ ,  $i = 1, 2, \dots, n$ , onde  $d_{ij}$  é a distância do  $i$ -ésimo evento ao  $j$ -ésimo evento. Então a função de probabilidade acumulada ou função distribuição pode ser estimada empiricamente por:

$$\hat{G}(w) = \frac{\#\{w_i : w_i \leq w, \forall i\}}{n},$$

onde o numerador é o número de distâncias  $w_i$  no conjunto das distâncias, tais que sejam menores ou iguais a  $w$  e o denominador  $n$  é o número total de eventos.

A função  $F$  mede a distribuição de todas as distâncias  $X$  de um ponto

arbitrário na região plana a eventos vizinhos mais próximos. Na realidade esta função mede o espaço médio dentre dois eventos. A função de probabilidade acumulada pode ser estimada por:

$$\hat{F}(x) = \frac{\#\{x_j : x_j \leq x, \forall j\}}{m}, \quad j = 1, 2, \dots, m,$$

onde o numerador é o número de distâncias  $x_i$  no conjunto das distâncias, tais que sejam menores ou iguais a  $x$  e o denominador  $m$  é o número total de pontos aleatórios amostrados.

Ao plotar-se as funções  $\hat{G}(w)$  contra  $W$ , ou  $\hat{F}(x)$  contra  $X$ , pode-se ter evidências da interação dos eventos. Se o gráfico crescer rapidamente para pequenos valores de distância, tem-se uma maior interação entre os eventos dando indícios de agrupamentos; caso apresente pequenos valores no início e depois cresça para valores maiores de distância tem-se uma distribuição regular.

Pode-se plotar  $\hat{G}(w)$  contra  $\hat{F}(x)$  para verificar a interação dos eventos. Se o gráfico for uma reta com ângulo próximo de  $45^\circ$ , então essas distribuições se assemelham e não há interação. Haverá interação positiva (padrão de aglomeração) se  $\hat{G}(w)$  exceder os valores de  $\hat{F}(x)$ , ou seja, as distâncias de evento-evento são menores que as de ponto-evento; a situação oposta implicaria num padrão regular.

O objetivo do método em si é verificar o grau de dependência dos eventos, comparando a curva empírica com curvas supondo completa aleatoriedade espacial dos eventos. Na situação de completa de aleatoriedade, a função de  $G(w)$  é dada por um processo de Poisson:

$$G(h) = 1 - e^{-\lambda\pi h^2}, \quad h \geq 0$$

Assim, o estimador de  $G(h)$  é dado pela expressão:

$$\bar{G}(h) = \frac{\sum_{i=1}^s \widehat{G}_i(h)}{s}, \quad i = 1, 2, \dots, s$$

e  $\widehat{G}_i(h)$  são funções de distribuição empíricas, que podem ser obtidas através de  $s$  simulações independentes dos  $n$  eventos, considerando-se a hipótese que são independentes entre si e uniformemente distribuídos.

Calcula-se os chamados envelopes de simulação inferior e superior, como:

$$I(h) = \min\{\widehat{G}_i(h)\}, \quad i = 1, 2, \dots, s$$

$$S(h) = \max\{\widehat{G}_i(h)\}, \quad i = 1, 2, \dots, s$$

Constrói-se um gráfico cartesiano  $\widehat{G}(h)$  versus  $\bar{G}(h)$ , compreendendo também os envelopes inferior e superior. Caso haja aleatoriedade, o gráfico da curva deve ser aproximadamente linear, formando um ângulo de  $45^\circ$ . Caso  $\widehat{G}(h)$  esteja acima da linha de  $45^\circ$ , há tendências para agrupamentos na distância  $h$ , enquanto que, se  $\widehat{G}(h)$  estiver abaixo da linha de  $45^\circ$ , há indicativo de um padrão de regularidade, na distância considerada. (Câmara et al., 2004)

#### 1.4.2 Função $K$

Um dos problemas do método do vizinho mais próximo é que ele desconsidera padrões de escalas maiores; afinal, as abordagens são feitas com eventos vizinhos mais próximos e perde-se informação para grandes distâncias. Para obter um panorama geral da dependência espacial sob diversas escalas, pode-se usar as medidas de segundo momento ou a função  $K$  do processo observado.

Para que as técnicas de estimação espacial façam sentido na prática, deve-se supor, implicitamente, que o processo em estudo seja homogêneo ou isotrópico em pequenas escalas na região. Caso contrário, as propriedades de

segunda ordem não são necessariamente constantes sobre uma escala considerada e não faz sentido estimar os efeitos do processo.

Se, por exemplo, estiver claro que existe muita variação na intensidade de um padrão pontual sobre a região inteira  $R$ , e isso for realmente um efeito de primeira ordem e não resultado de dependência espacial, então seria sensato estudar somente efeitos de segunda ordem em escalas suficientemente pequenas para valer a suposição de isotropia na região  $R$ . Por outro lado, se a variação na intensidade parece não acontecer, ou aquela que existe possivelmente seja devido a dependência espacial e não um autêntico efeito de primeira ordem, então pode-se justificar o exame dos efeitos de segunda ordem sobre escalas maiores na região de estudo.

A função  $K$  pode ser definida como

$$K(h) = \frac{E(\#(\text{eventos dentro da distância } h \text{ de um evento arbitrário}))}{\lambda}$$

onde  $\#$  significa contagem,  $E()$  é o operador de esperança e  $\lambda$  é a intensidade ou número de eventos por unidade de área, assumido constante em toda região  $R$ .

Seja  $R^*$  a área de  $R$ . Então o número esperado de eventos em  $R$  é  $\lambda R^*$ , e assim, da definição de função  $K$ , o número esperado de pares numa distância máxima  $h$  é  $\lambda^2 R^* K(h)$ .

Sejam  $d_{ij}$ , a distância entre o  $i$ -ésimo e o  $j$ -ésimo evento observados em  $R$  e  $I_h(d_{ij})$ , uma função indicatriz cujo valor é 1 se  $d_{ij} \leq h$  e 0 em caso contrário. Então, o número total de pares observados, é:

$$\sum I_h(d_{ij})$$

Portanto, segundo Bailey & Gatrell (1995), uma estimativa de  $K(h)$  é:

$$\widehat{K}(h) = \frac{1}{\lambda^2 R^*} \sum \sum_{i \neq j} I_h(d_{ij})$$

Uma correção do efeito da borda de  $R$  deve ser feita na estimativa, com o objetivo de considerar pares de eventos onde o segundo evento está externo à região  $R$ , sendo não observável. Seja, então, um círculo centrado em  $i$  passando por  $j$  e  $w_{ij}$  a proporção da circunferência deste círculo dentro de  $R$ .

$w_{ij}$  representa a probabilidade condicional de que um evento  $j$  seja observado em  $R$  dado que está à uma distância  $d_{ij}$  do evento  $i$ . Desta maneira, uma correção de borda para o estimador de  $K(h)$  é:

$$\widehat{K}(h) = \frac{1}{\lambda^2 R^*} \sum \sum_{i \neq j} \frac{I_h(d_{ij})}{w_{ij}}$$

Uma estimativa de intensidade de  $\lambda$  é  $\widehat{\lambda} = \frac{n}{R^*}$ . Finalmente,

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

Como ilustração, veja a figura 3

Seja um evento qualquer e em torno dele é construído um conjunto de círculos concêntricos de pequeno espaçamento. O número acumulado de eventos dentro das faixas de distância é contado. Todo qualquer outro ponto é similarmente "visitado" e o número de eventos cumulativos dentro de faixas de distância até um raio  $h$ , em torno de todos os eventos torna-se a estimativa de  $K(h)$  quando multiplicado por  $\frac{R}{n^2}$ .

Obtida a estimativa, pode-se plotar  $\widehat{K}(h)$  para se verificar a dependência espacial, para diferentes valores de  $h$ . Entretanto, há certa dificuldade no processo pontual, em se concluir a respeito da dependência espacial, observando-se o gráfico, pois diferentemente do método do vizinho mais próximo (cuja independência representa uma reta de aproximadamente 45°), não se sabe exatamente qual figura

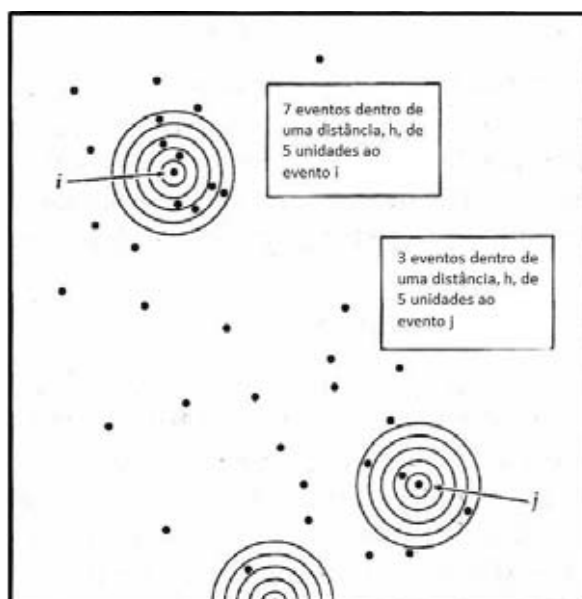


Figura 3 - Estimação da função K. Fonte:(Bailey & Gatrell, 1995)

acontecerá no caso de independência.

Pode-se detectar através da função  $K$  se um padrão aleatório ou regular ou em forma de *clusters* está presente nos dados. Um padrão aleatório implica que a ocorrência de um evento em qualquer ponto da região  $R$  independe da condição de terem ou não ocorridos outros eventos e é igualmente provável em toda a região. Assim, para um processo aleatório, o número esperado de eventos à uma distância  $h$  de qualquer ponto aleatoriamente escolhido é  $\lambda\pi h^2$ . Desta forma, para um processo homogêneo, onde não haja dependência espacial,

$$K(h) = \pi h^2$$

Sob regularidade,  $K(h) < \pi h^2$  e sob a formação de *clusters*,  $K(h) > \pi h^2$ .

Portanto, para conhecer o comportamento de um evento em uma área, a estimativa de  $K(h)$  pode ser comparada com  $\pi h^2$ . A comparação pode ser feita

através do plot de  $\widehat{L}(h)$  em função de  $h$ , onde

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

Assim, de acordo com Bailey & Gatrell (1995),

- $\widehat{L}(h) > 0$  indica uma tendência dos eventos aparecerem em *clusters*;
- $\widehat{L}(h) = 0$  indica aleatoriedade dos eventos;
- $\widehat{L}(h) < 0$  indica distribuição regular, na distância  $h$ .

Uma expressão alternativa a  $\widehat{L}(h)$  é calcular

$$\widehat{l}(h) = \frac{1}{2} \log\left(\frac{\widehat{K}(h)}{\pi}\right) - \log h,$$

onde os valores altos desta função tendem a *clusters* na distância  $h$ .

Numa situação em que  $\widehat{L}(h)$  ou  $\widehat{l}(h)$  forem positivos para todos os valores de  $h$ , então há um indicativo que os eventos aparecem em *clusters* em todas as escalas de  $h$ .

## 1.5 Estimação da Intensidade: Kernel

O padrão espacial do processo pode ser descrito pela estimação da densidade estatística espacial dos dados observados. Essa densidade tem as mesmas propriedades da densidade univariada e apenas o domínio que se modifica, que passa a ser a área de estudo do processo. Mede-se, assim, a distribuição dos eventos, calculando a intensidade do processo,  $\lambda(x)$ .

Dentre as diversas técnicas estatísticas existentes para mapear a forma de distribuição dos pontos, destaca-se a técnica kernel, de grande uso nos anos recentes, a qual consiste num estimador probabilístico de intensidade do processo pontual não-paramétrico através de função kernel. As entradas para aplicação são



as ocorrências da variável (na área , através de um sistema de coordenadas).

Se  $s$  representa uma localização qualquer em  $R$  e  $s_1, s_2, \dots, s_n$  são as localizações dos  $n$  eventos observados, então um estimador para  $\lambda$ , em  $s$  é dado por:

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

Onde  $k(\cdot)$  é uma função kernel bivariada e simétrica chamada de função de estimação ou alisamento e o parâmetro  $\tau > 0$  é conhecido como largura da banda (ou raio de influência) e determina o grau de suavização; essencialmente ele é o raio de um disco centrado em  $s$  em que os  $s_i$  irão contribuir significativamente para  $\hat{\lambda}(s)$ .

A técnica suaviza a superfície, calculando a densidade para cada região da área de estudo, utilizando interpolação. Isto permite a construção de uma superfície contínua de ocorrências das variáveis, inferindo para toda a área de estudo a variação espacial da variável, mesmo nas regiões onde o processo não tenha gerado nenhuma ocorrência real, permitindo verificar, em escala global, possíveis tendências de dados.

A figura 4 explicita os passos para o cálculo de densidade dos pontos, através da Kernel.

Como mostra a figura 4, é realizada a contagem dos eventos de interesse dentro de uma região de influência, ponderando-os pela distância, de cada um, à localização de interesse. (Câmara & Carvalho, 2004)

A escolha do raio  $\tau$  é crucial, produzindo significantes alterações na estimativa final.(Brasil. Ministério da Saúde, 2007).

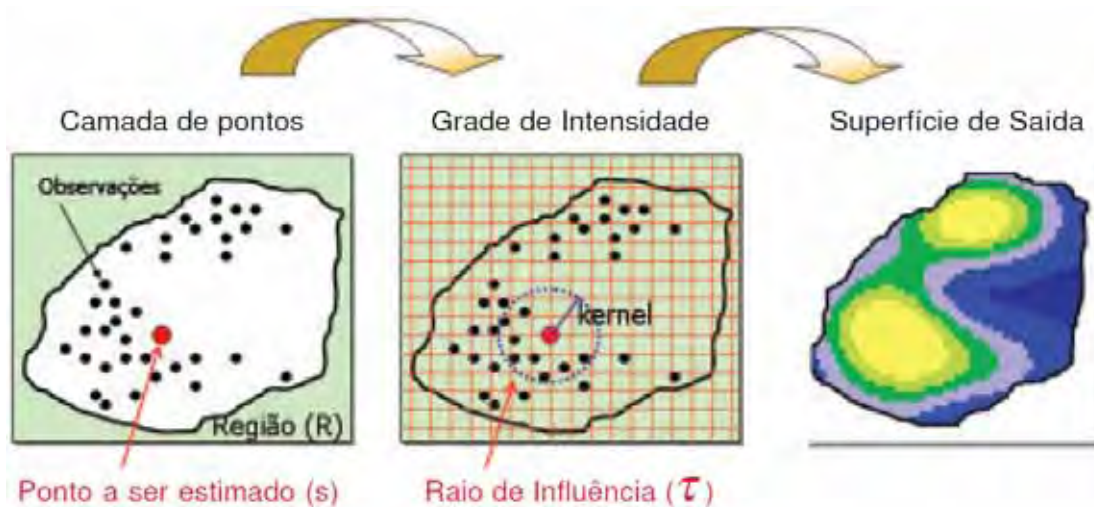


Figura 4 - Passos para o cálculo de intensidade de pontos segundo a técnica de Kernel. Extraído de Brasil. Ministério da Saúde (2007)

Já a escolha da função kernel não é muito complicada. A função de estimação kernel executa a contagem dos eventos  $s_i$  dentro do raio de influência,  $\tau$ , ponderando cada um deles pela distância ao ponto  $s$ ,  $h_i$ . (Câmara & Carvalho, 2004)

Para o kernel quártico, a expressão torna-se

$$k\left(\frac{s - s_i}{\tau}\right) = \frac{3}{\pi} \left(1 - \left(\frac{s - s_i}{\tau}\right)^2\right)^2 \quad \text{para } \left(\frac{s - s_i}{\tau}\right)^2 \leq 1$$

Desta forma,

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

onde  $h_i$  é distância entre os pontos  $s$  e  $s_i$ ,  $i = 1, 2, \dots, n$ . O fator  $\frac{3}{\pi \tau^2}$  suaviza a função em valores  $s_i$  próximos à  $s$ , onde a distância tende a zero.

As outras funções de estimação kernel  $k()$  são:

- kernel gaussiano ou normal:

$$k(h) = \frac{1}{2\pi\sigma} e^{-\frac{h^2}{2\sigma^2}}, \text{ onde } \sigma = \tau$$

- kernel triangular:

$$k(h) = (1 - |h|)$$

- kernel exponencial negativo:

$$k(h) = \frac{1}{2\pi} e^{-h}$$

- kernel uniforme:

$$k(h) = \frac{1}{2}$$

A distribuição normal pondera os pontos dentro do círculo de forma que os pontos mais próximos têm maior peso comparados com os mais afastados.

A função quártica pondera com maior peso os pontos mais próximos do que pontos distantes, mas o decréscimo é gradual.

A função triangular dá maior peso aos pontos próximos do que os pontos distantes dentro do círculo, mas o decréscimo é mais rápido. A função exponencial negativa pondera os pontos próximos com peso muito mais intenso do que os pontos distantes. A função uniforme pondera todos os pontos dentro do

círculo igualmente.

Usa-se  $\delta_\tau(s) = \int_R \frac{1}{\tau^2} k\left(\frac{s-u}{\tau}\right) du$  como fator de correção de borda para compensar as observações perdidas que ocorrem quando  $s$  está próximo a borda de  $R$ , a região de estudo.

Desse modo, a intensidade  $\lambda$  pode ser reescrita como:

$$\lambda_\tau(s) = \frac{1}{\delta_\tau(s)} \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s-s_i}{\tau}\right)$$

O problema da escolha do raio  $\tau$  é delicado. Há alguns estudos sugerindo aproximações. Por exemplo, uma escolha aproximada seria  $\tau = 0,68n^{-0,2}$ , para estimar a intensidade quando a região  $R$  for um quadrado unitário e  $n$  for o número de eventos observados em  $R$ . Esta sugestão pode ser estendida a regiões de estudo de outros formatos (Bailey & Gatrell, 1995).

Se a variável em estudo ocorrer de forma diversa na área de análise, de forma a produzir regiões esparsas e regiões mais densas, o uso de um único raio pode não ser adequado. Neste caso, o ajuste do kernel é feito localmente, por uma técnica conhecida como estimação kernel adaptativo. Para cada ponto  $s_i$ , seja a expressão:

$$\lambda_\tau(s) = \sum \frac{1}{\tau_{s_i}^2} k\left(\frac{s-s_i}{\tau_{s_i}}\right),$$

onde  $\tau_{s_i}$  é uma função da presença de eventos na vizinhança de  $s_i$ .

Bailey & Gatrell (1995) sugerem ajustar inicialmente com um kernel não-adaptativo através de um raio  $\tau_0$ , encontrando um estimativa piloto  $\lambda(s)$ .

Calcula-se a média geométrica  $\lambda_g$  das estimativas piloto  $\lambda(s_i)$  em cada  $s_i$ . Na sequência, calcula-se  $\tau(s_i) = \tau_0(\frac{\lambda_g}{\lambda(s_i)})^\alpha$ , onde  $\alpha$  é conhecido com parâmetro de sensibilidade.

A escolha de  $\alpha = 0$  corresponde a nenhum ajuste local;  $\alpha = 1$ , corresponde ao máximo ajuste local. Na prática, a escolha de  $\alpha = 0,5$  tem produzido bons resultados.

## 1.6 Superfícies de Tendência

Uma outra técnica que possibilita o estudo do comportamento espacial de uma variável é a da Superfícies de Tendência, onde ajusta-se uma superfície numa área que melhor explique o comportamento espacial da variável, bem como a existência de anomalias locais, segundo Davis (1986). Esta técnica consiste de uma análise de superfície de tendência, por regressão polinomial.

Ajusta-se um polinômio de grau  $k$  ( $k = 1, 2, 3, \dots$ ), na área em estudo, segundo o modelo:

$$z(x_i, y_i) = a_0 + a_1x_i + a_2y_i + a_3x_i^2 + a_4y_i^2 + a_5x_iy_i + \dots + e(x_i, y_i),$$

onde,  $z$ : variável mapeada, em função das coordenadas  $(x_i, y_i)$ ;  
 $a_0, a_1, a_2, \dots$  : coeficientes estimados pelo método de mínimos de quadrados;  
 $e(x_i, y_i)$ : resíduo do modelo no ponto  $(x_i, y_i)$

O grau  $k$  do polinômio que melhor se ajusta aos dados pode ser escolhido levando-se em consideração:

- o valor do coeficiente de explicação  $R^2$  do modelo (que deve ser alto);
- a análise do gráfico de resíduos;
- a simplicidade do modelo.

Deseja-se, assim, um polinômio de grau não muito elevado, mas com alto valor de  $R^2$  e com um gráfico de resíduos exibindo aleatoriedade, dentre outras propriedades.

Quando o interesse for o melhor ajuste aos dados, ajusta-se uma superfície de maior grau possível; se o objetivo for apenas na detecção de anomalias locais, o interesse maior é pelos resíduos, e assim, pode-se ajustar uma superfície de baixo grau (grau 1, por exemplo), obtendo-se o mapa de resíduos. Neste caso, como o foco de interesse está nos resíduos, é pouco relevante o modelo do ajuste da superfície (Lourenço et al., 2001).

O modelo das superfícies de tendência é polinomial, que é muito simples comparado com o ajuste de outros modelos. Entretanto, se o grau do polinômio for muito elevado, a matriz solução pode tornar-se instável devido a erros de arredondamento, já que haverá um grande número de equações simultâneas para ser resolvido, cujos coeficientes poderão ser muito elevados.

## 2 MATERIAL E MÉTODOS

O principal objetivo deste estudo foi uma compreensão maior do estimador de intensidade Kernel, para mapear a ocorrência das dados e fazer análise dos aglomerados obtidos. Utilizou-se para as análises do kernel, o *software* Terraview 4.0 (2010)

Comparações entre os resultados para diferentes funções de densidades e diferentes raios no estimador Kernel, foram feitas, utilizando conjunto de dados reais. Comparou-se as seguintes funções kernel, disponíveis no programa-Terraview 4.0 (2010): quártico, normal, triangular, uniforme e exponencial negativo.

Essa aplicação foi feita sobre dois conjuntos de dados retirados da cidade de Rio Claro-SP.

Uma comparação dos resultados do estimador kernel com superfícies de tendência foi feita, utilizando 4 conjuntos hipotéticos de dados em uma pequena malha quadrada. Para determinar a superfície de tendência utilizou-se o *software* Surfer 7.0 (1999)

### 2.1 Conjuntos de dados

Um conjunto de dados refere-se a acidentes de trânsito ocorridos no perímetro urbano da cidade de Rio Claro-SP, notificados através de Boletim de Ocorrência, junto à Polícia Militar, no período de Janeiro de 2009 a Setembro de 2010, cedidos pela Secretaria de Mobilidade Urbana e Sistema Viário da Prefeitura

de Rio Claro. Tratam-se da localização de 4544 ocorrências de acidentes de trânsito, havendo vítimas ou não, e desconsiderando-se o tipo de acidente (figura 5). As ocorrências com vítimas foram 1871 e as que não tiveram qualquer vítima foram 2673 (figura 6). Os dados foram georreferenciados para identificação dos locais das ocorrências e aplicação da técnica de estimação kernel para mapear a distribuição espacial das ocorrências de acidentes de trânsito. Também realizou-se essa análise discriminando com e sem vítima.



Figura 5 - Distribuição de acidentes de trânsito em Rio Claro. Fonte: Secretaria de Mobilidade Urbana- P. M. Rio Claro-SP.





Figura 6 - Distribuição de acidentes de trânsito em Rio Claro, com vítima ou não.

Fonte: Secretaria de Mobilidade Urbana- P. M. Rio Claro-SP.

O outro conjunto de dados refere-se a casos positivos da dengue na cidade de Rio Claro-SP, registrados no primeiro semestre de 2011. Trata-se de 1656 localizações de residências de pessoas que tiveram a dengue, não tendo mais informações disponíveis. O conjunto de dados foi fornecido pela Defesa Civil, órgão subordinado à Secretaria de Segurança Pública do Município de Rio Claro-SP. Foram usados 1357 localizações (figura 7), e os outros casos de dengue foram descartados da análise por problemas no seu georrefenciamento.

Através dos endereços georreferenciados, foi possível aplicar a técnica kernel com o objetivo de mapear a distribuição espacial da dengue na cidade. Embora



Figura 7 - Distribuição de casos positivos de dengue em Rio Claro. Fonte: Secretaria de Segurança Pública - P.M. Rio Claro- SP.

nem todo paciente tenha contraído o vírus causador do dengue em seu local de residência, devido à mobilidade das pessoas, considerou-se o endereço da vítima como o local onde se contraiu a doença. Acredita-se que o erro é desprezível, devido ao grande tamanho amostral envolvido.

Finalmente, para maior facilidade de comparação da técnica Kernel com a Superfície de Tendência, gerou-se 4 conjuntos de hipotéticos de dados, numa

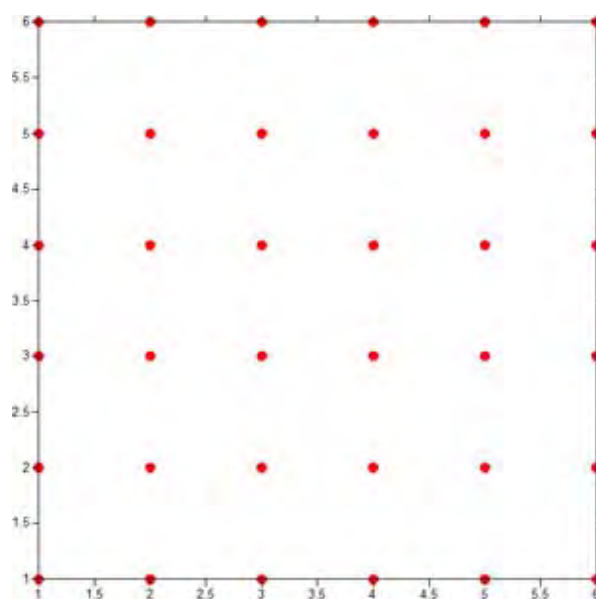


Figura 8 - Malha de pontos dos valores hipotéticos

malha quadrada de 36 pontos, conforme (figura 8) e apêndice. Os conjuntos foram denotados por:

- valor1: tendência de crescimento dos dados em forma de uma superfície linear;
- valor2: tendência de rápido crescimento dos dados em forma de superfície linear;
- valor3: tendência de distribuição dos dados na área de acordo com uma superfície de 2º grau;
- valor4: tendência de distribuição aleatória dos dados da área

Para cada conjunto de dados, gerou-se os gráficos de kernel e de superfície de tendência, comparando visualmente os mapas, bem como os resíduos gerados por cada uma das técnicas.

### 3 RESULTADOS E DISCUSSÃO

Apresenta-se a seguir as intensidades estimadas pelo estimador kernel.

As figuras 9 e 10 apresentam os resultados da aplicação do Kernel para diferentes valores de raios de influência, para os dados de acidentes de trânsito em Rio Claro, referentes ao período 2009-2010.

Os valores  $\tau$  do raio de influência testados são os seguintes: 250 metros, 375 metros, 500 metros e 1000 metros, pela função quártico.

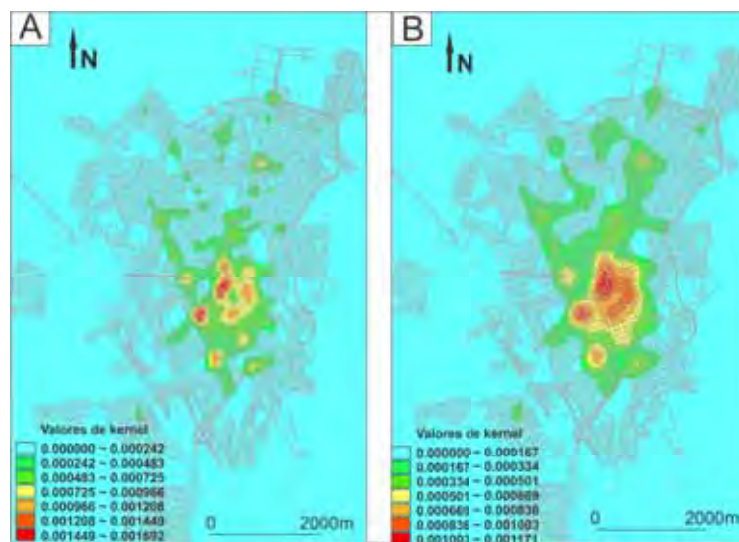


Figura 9 - Estimação kernel quártico aplicado aos dados de acidentes de trânsito. A) raio de influência = 250m. B) raio de influência = 375m

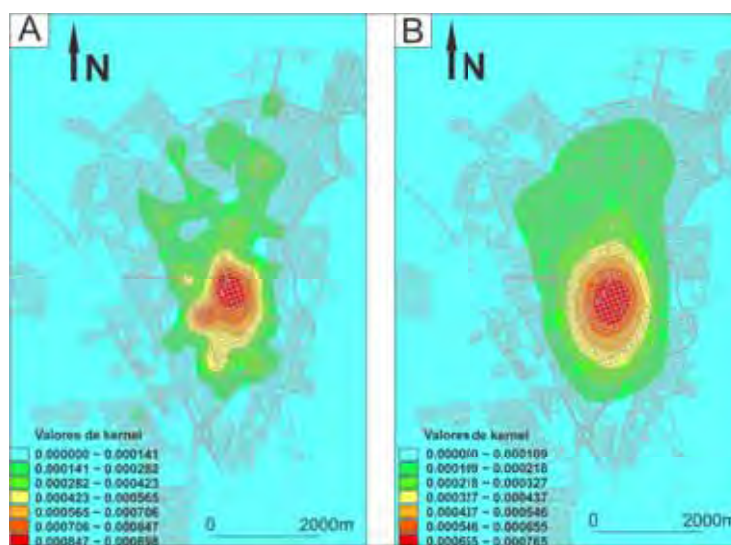


Figura 10 - Estimação kernel quártico aplicado aos dados de acidentes de trânsito.

A) raio de influência = 500m. B) raio de influência = 1000m

Os acidentes, em geral ocorrem em pontos específicos da cidade, como ruas ou avenidas de grande movimento, cruzamentos e rotatórias.

A grande prevalência das ocorrências está na região central da cidade, destacada na figura 11, em especial na Avenida Visconde de Rio Claro, que é uma avenida de grande movimento e que praticamente atravessa toda a parte central da cidade.

Os melhores resultados obtidos para  $\tau$  aproximadamente 250 metros, onde pode-se observar valores altos do Kernel (correspondente às chamadas "áreas quentes") em torno do centro da cidade, compreendendo a Avenida Visconde de Rio Claro (figura 11), o que discrimina adequadamente os acontecimentos.



Figura 11 - Estimaco kernel qurtico aplicado aos dados de acidentes de trnsito, destacando a Avenida Rio Claro

Para valores de raios muito maiores que 250 metros ( figura 10A e figura 10B), h uma suavizaco das ocorrncias dos eventos, o que no apresenta interesse prtico, neste caso. Para valores muito pequenos de  $\tau$ , os dados aparecem em alguns picos de maiores ocorrncias. Pode tambm ser interessante para se detectar cruzamentos ou outros locais de grandes ocorrncias. Entretanto para raios prximos a 250 metros, a suavizaco da distribuio espacial dos eventos  mais ntida.

A partir de comparações da função kernel quártico e função kernel normal (figura 12, figura 13), utilizando os mesmos raios de influência, e utilizando raios menores para a normal (figura 14) percebe-se que as funções quárticas fornecem mais informações *in loco* do que a normal, para raios maiores. Apesar das funções normal e quártica fazerem ponderações aos pontos mais próximos dentro do raio de influência, a função quártica tem seu decréscimo gradual, o que faz com que as suas percepções visuais fiquem mais evidentes para raios maiores

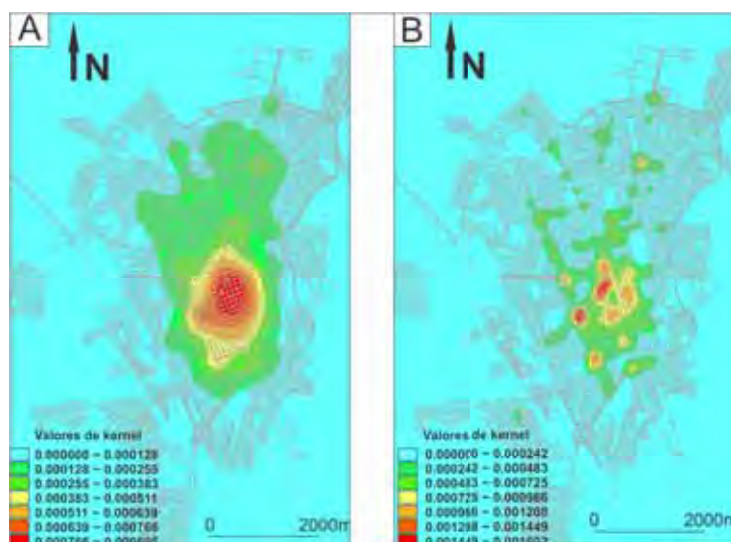


Figura 12 - Comparação de funções kernel com raio de influência igual a 250m aplicado aos dados de acidentes de trânsito. A)Kernel Normal. B)Kernel Quártico

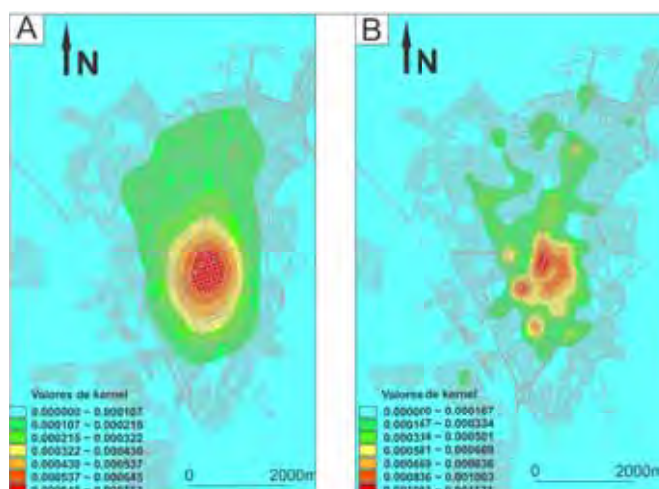


Figura 13 - Comparação de funções kernel com raio de influência igual a 375m aplicado aos dados de acidentes de trânsito. A)Kernel Normal. B)Kernel Quártico

A função normal forneceu uma boa visualização dos aglomerados nos seguintes raios: 100m, 125m e 150m (figura 14). Obteve-se aglomerados semelhantes aos realizados com a função quártico, ainda, comparando as figuras 9A e 14A, nota-se que o raio de 250m para a função quártica produziu os mesmos efeitos que o raio 100m para a função normal .

A função uniforme pesa todos os pontos igualmente, o que torna sua visualização uniforme e não útil para análises de kernel.

A função triangular pesa pontos próximos mais do que pontos distantes dentro do círculo, mas o decréscimo é mais rápido. Já, a função exponencial negativa pesa pontos próximos muito mais intensamente do que pontos distantes. Desse modo, a visualização é também uniforme e neste caso inútil para a estimação de kernel.



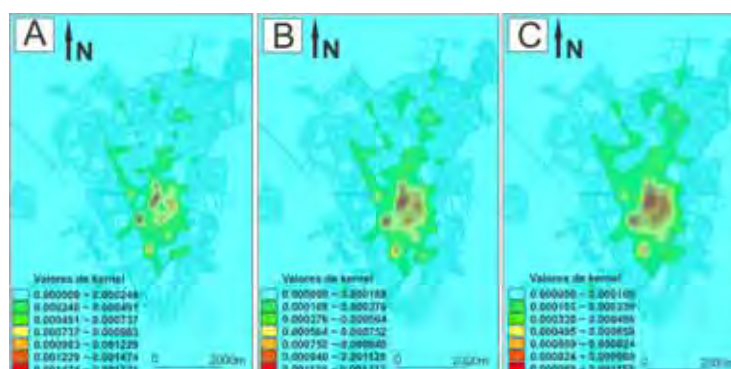


Figura 14 - Estimação kernel normal aplicado aos dados de acidentes de trânsito.

A) raio de influência = 100m. B) raio de influência = 125m. C) raio de influência = 150m.

Obteve-se também os mapas de estimação kernel para os casos de acidente de trânsito, com e sem vítima (figura 15A e 15B), observa-se que os padrões de distribuição espacial dos acidentes com vítimas e sem vítima são semelhantes.

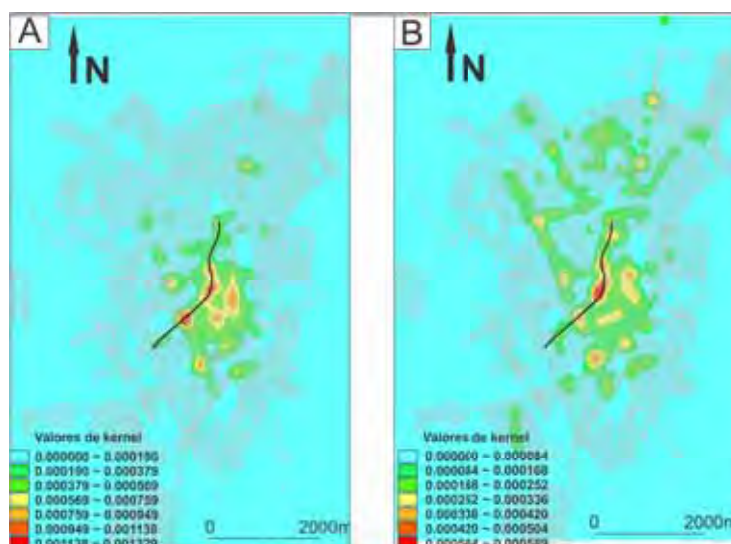


Figura 15 - Mapa de Estimação de kernel quártico com raio de influência igual a 250m aplicado a dados de acidentes de trânsito: A) sem vítimas; B) com vítimas.

Para os eventos de dengue (figura 16, figura 17), o raio que gera melhores visualizações é em torno de 500 metros (figura 17A). Nota-se uma grande área de intensas ocorrências do evento, área esta situada na região norte da cidade (figura 18), o que certamente indica focos de infestação de dengue nesta região e alertando ao Poder Público Municipal os locais que merecem uma maior atenção para política de prevenção e combate da dengue .

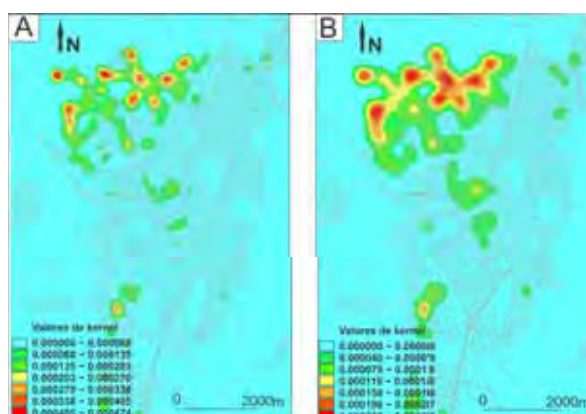


Figura 16 - Estimação kernel quártico aplicado aos dados de dengue. A) raio de influência = 250m. B) raio de influência = 375m

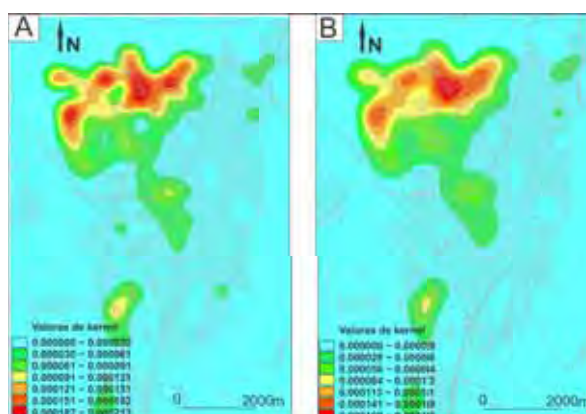


Figura 17 - Estimação kernel quártico aplicado aos dados de dengue. A) raio de influência = 500m. B) raio de influência = 625m

Outras áreas de menores intensidades também foram evidenciadas e devem ser analisadas e atendidas pelas medidas citadas anteriormente, como a parte oeste da cidade, onde tem-se um aglomerado de intensidade média (figura 18).

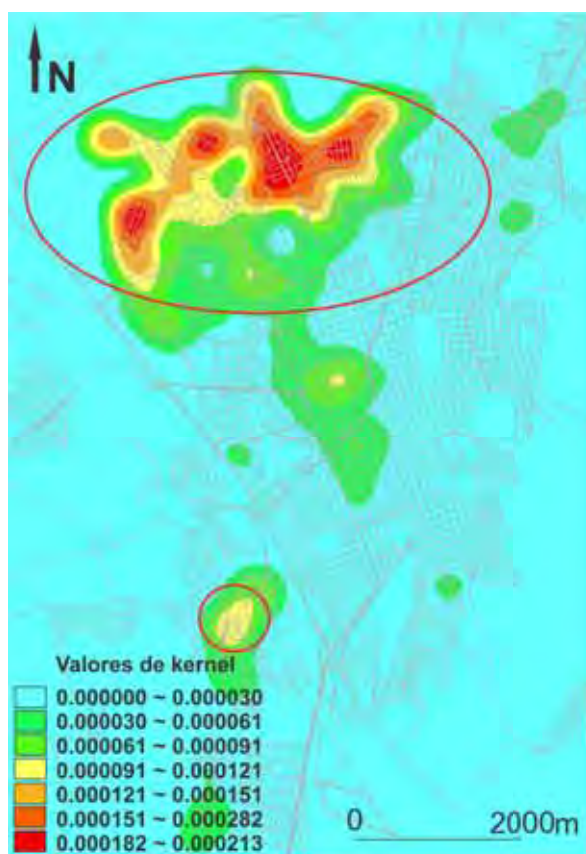


Figura 18 - Estimação kernel quártico aplicado aos dados de dengue ao raio de influência de 500m, destaque para algumas aglomerações.

A comparação de algumas funções kernel com a função normal (figura 19, figura 20), utilizando o mesmo raio de influência, aplicadas a dados de dengue indicam que a função quártica fornece mais informações detalhadas (mais pontuais) do que a normal. Considerando a figura 21, a função normal também fornece uma boa visualização, mas para raios menores.

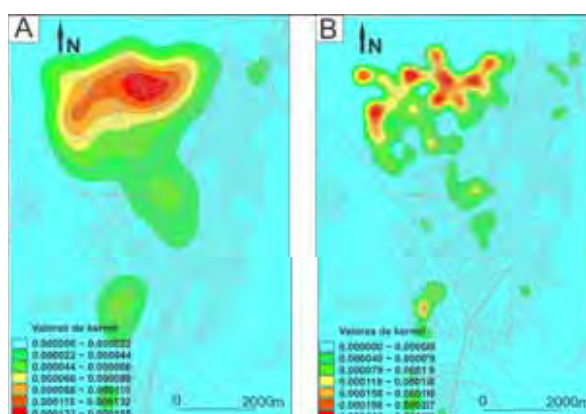


Figura 19 - Comparação de funções kernel com raio de influência igual a 375m aplicado aos dados de dengue. A)Kernel Normal. B)Kernel Quártico

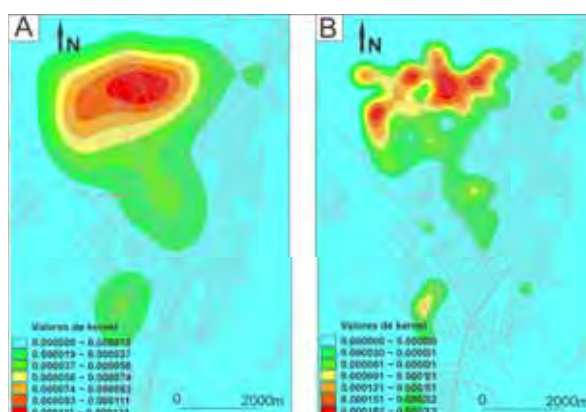


Figura 20 - Comparação de funções kernel com raio de influência igual a 500m aplicado aos dados de dengue. A)Kernel Normal. B)Kernel Quártico

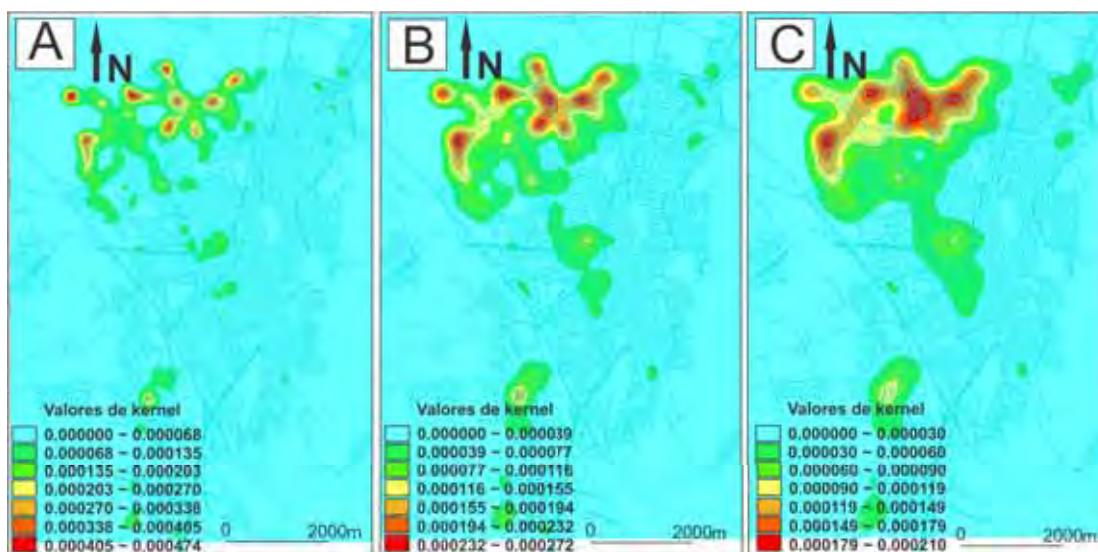


Figura 21 - Estimação kernel normal aplicado aos dados de dengue. A) raio de influência = 100m. B) raio de influência = 150m. C) raio de influência = 200m.

Considerando os melhores valores de raios para os dois conjuntos de dados, nota-se que no caso dos acidentes de trânsito o ponto ótimo de valores de raio é menor que o caso de dengue. Provavelmente isso se deve ao fato dos acidentes de trânsito ocorrerem mais em pequenos grupos (pontuais) que os valores de dengue.

A comparação dos métodos de intensidade pelo kernel e a análise da superfície de tendência, pela regressão polinomial foi feita pelos resíduos das técnicas. Nas figuras 22, 23, 24, 25 tem-se os mapas de estimação kernel e a superfície de tendência gerado pelo conjunto de dados hipotéticos.

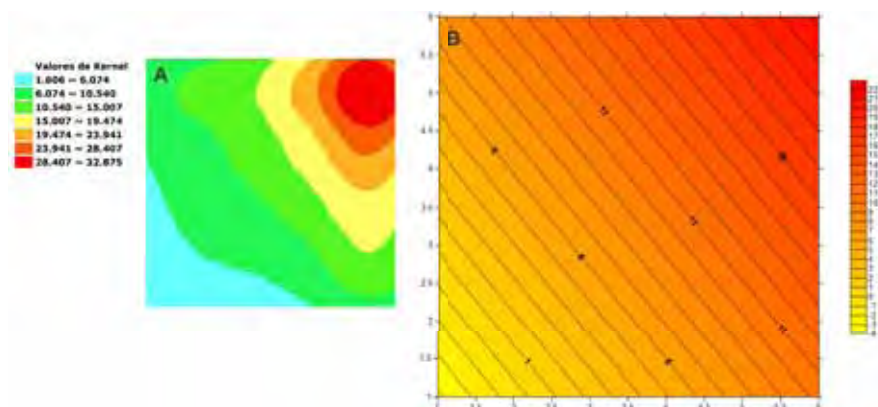


Figura 22 - A) Mapa de estimação kernel quártico aplicado ao valor 1, com raio 2.  
B) Superfície de tendência linear aplicado ao valor 1.

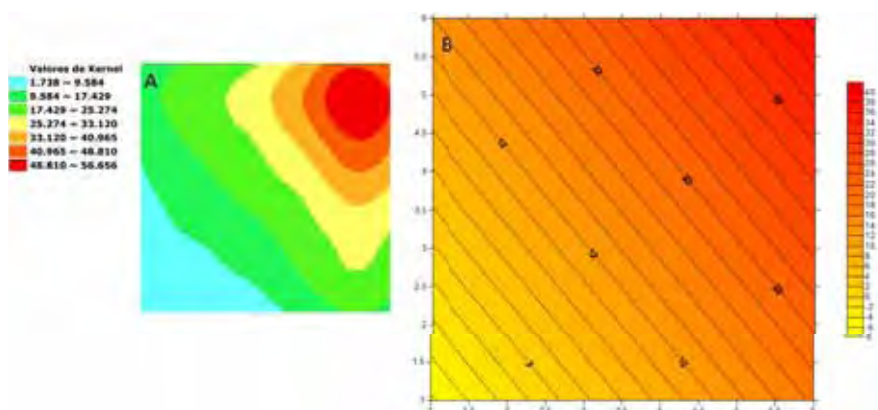


Figura 23 - A) Mapa de estimação kernel quártico aplicado ao valor 2, com raio 2.  
B) Superfície de tendência linear aplicado ao valor 2.

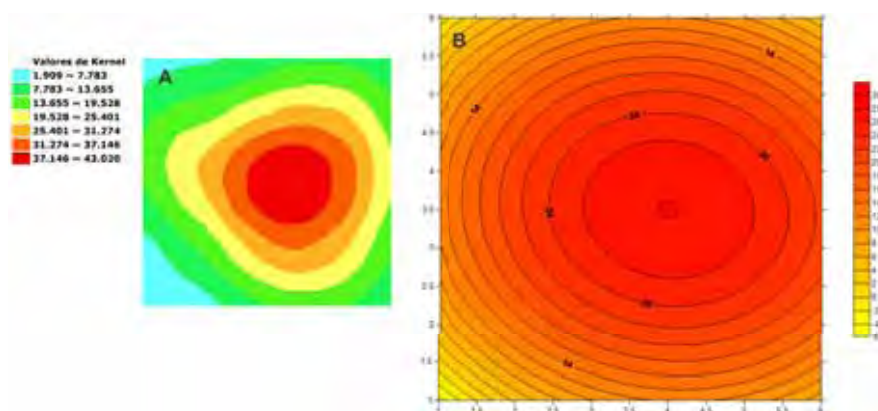


Figura 24 - A) Mapa de estimação kernel quártico aplicado ao valor 3, com raio 2 .  
B) Superfície de tendência quádrático aplicado ao valor 3.

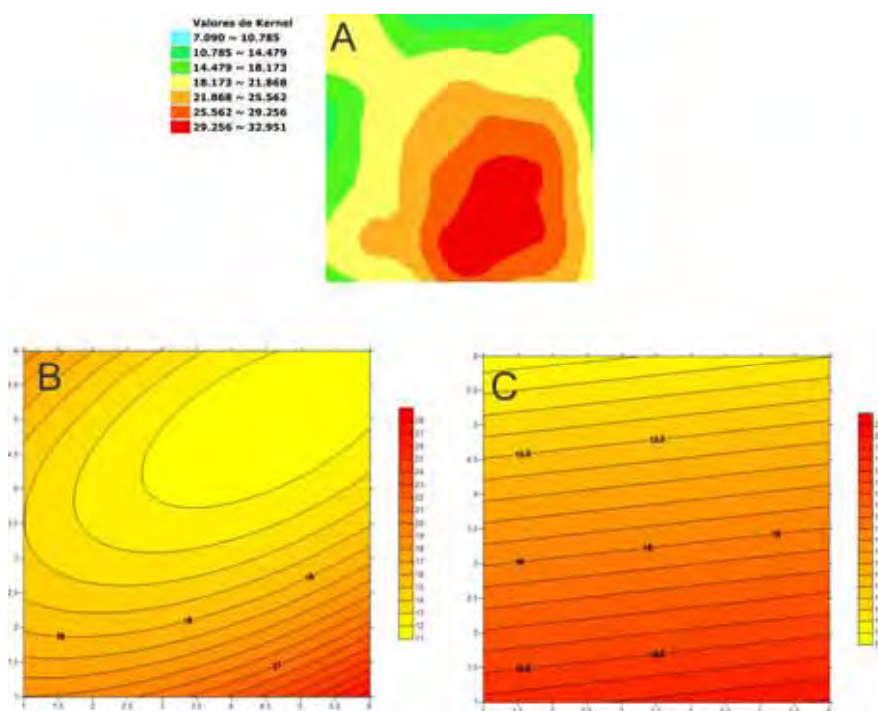


Figura 25 - A) Mapa de estimação kernel quártico aplicado ao valor 4, com raio 2 .  
B) Superfície de tendência quádrático aplicado ao valor 4. C) Superfície de tendência linear aplicado ao valor 4.

Obteve-se os mapas de resíduos, figuras 26, 27 , para a análise da superfície de tendência. Os resíduos da estimação kernel foram calculados pelo estimador  $\lambda$  sem considerar o efeito de borda.

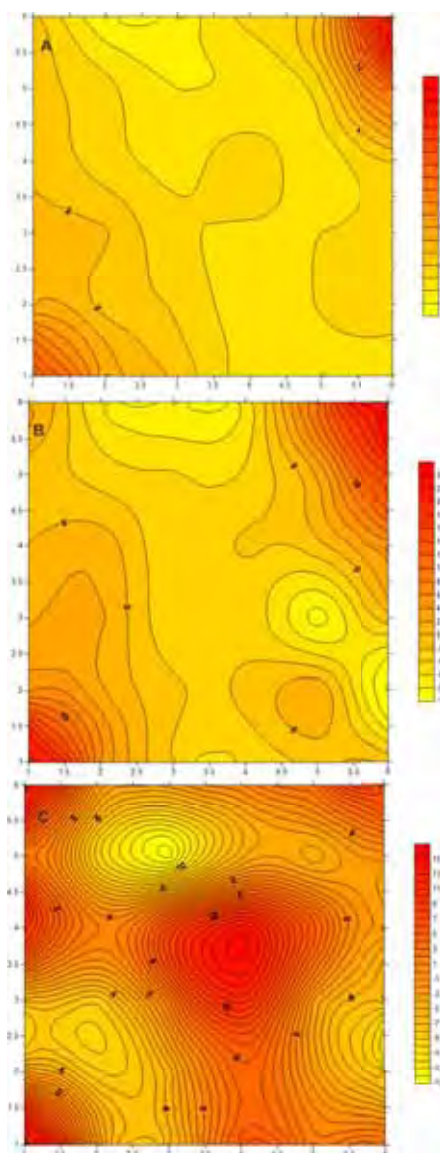


Figura 26 - A) Mapa de resíduos aplicado ao valor 1. B) Mapa de resíduos aplicado ao valor 2. C) Mapa de resíduos aplicado ao valor 3.



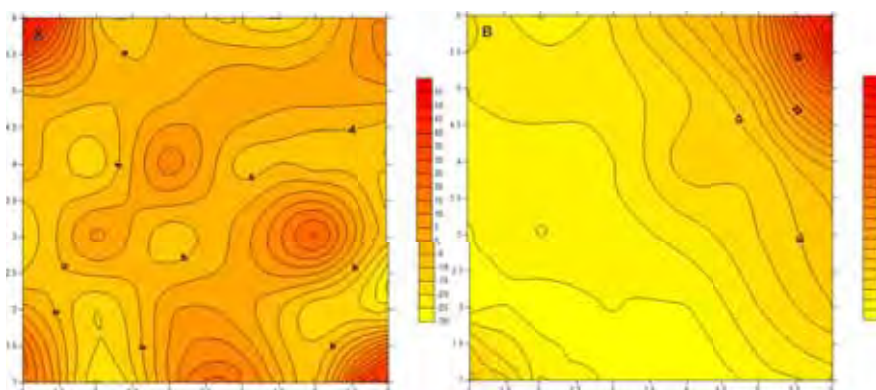


Figura 27 - A) Mapa de resíduos aplicado ao valor 4, supondo superfície quadrática.  
B) Mapa de resíduos aplicado ao valor 4, supondo superfície linear.

A comparação dos resíduos feita para os dados hipotéticos apresenta-se nas tabelas 1 e 2. Para o valor 4 foram ajustadas duas superfícies: linear e quadrática.

Tabela 1. Comparação dos resíduos, quanto ao desvio padrão.

	valor1		valor2		valor3	
	Kernel	Superf. Tend	Kernel	Superf. Tend	Kernel	Superf. Tend
Desvio Padrão	3,68	3,28	5,71	4,35	3,82	5,81

Tabela 2. Comparação de resíduo do valor4, quanto ao desvio padrão.

	valor4		
	Kernel	Superf. Tend. Linear	Superf. Tend. Quadrat.
Desvio Padrão	9.69	9.75	9.33

Nota-se que, em relação ao desvio padrão dos resíduos, que no ajuste das superfícies lineares (valor 1 e valor 2) o desempenho das superfícies de tendência foi superior ao do kernel, enquanto que, no ajuste da superfície do segundo grau (valor 3), ocorreu o inverso. Quanto ao valor 4, os desvios padrões

são aproximadamente iguais, tanto para entre o kernel e a superfície de tendência linear, como para a quadrática.

As figuras 28, 29, 30, 31 e 32 apresentam os gráficos Box-plot para os resíduos, comparando as duas técnicas de ajuste.

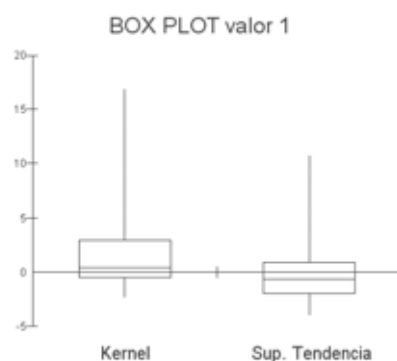


Figura 28 - Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência linear considerando valor 1

Pode-se notar, da figura 28, alta assimetria positiva dos resíduos, em ambos os casos; maior variabilidade no gráfico kernel; a mediana dos resíduos no kernel é positiva, enquanto que, no método das superfícies de tendência, é negativa.

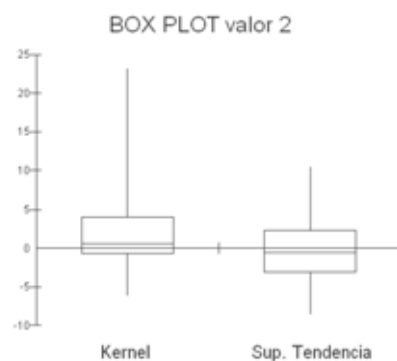


Figura 29 - Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência linear considerando valor 2

Da figura 29, observa-se que há alta assimetria positiva dos resíduos no caso kernel; maior variabilidade no gráfico kernel; a mediana dos resíduos no kernel é positiva, enquanto que, no método das superfícies de tendência, é negativa.

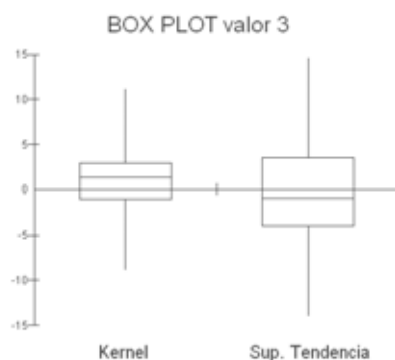


Figura 30 - Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência quadrática considerando valor 3

Da figura 30, nota-se que há alta assimetria negativa dos resíduos no caso kernel; assimetria à direita no gráfico das superfícies de tendência; maior variabilidade no gráfico das superfícies de tendência; a mediana dos resíduos no kernel é positiva, enquanto que, no método das superfícies de tendência, é negativa.

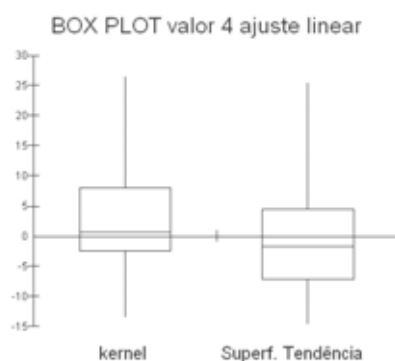


Figura 31 - Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência linear considerando valor 4

Pode-se notar, da figura 31, assimetria positiva dos resíduos, em ambos os casos; variabilidades aproximadamente iguais nos dois conjuntos; a mediana dos resíduos no kernel é positiva (muito próxima de zero), enquanto que, no método das superfícies de tendência, é negativa.

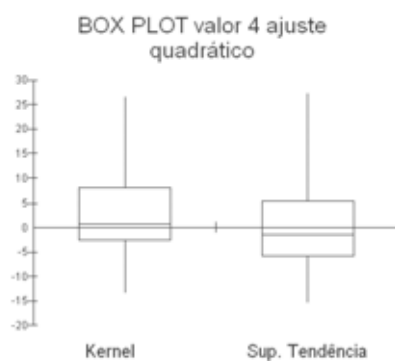


Figura 32 - Gráfico Box Plot dos Resíduos, comparando com o Kernel e a superfície de tendência quadrática considerando valor 4

Pode-se notar, da figura 32, assimetria positiva dos resíduos, em ambos os casos; variabilidades aproximadamente iguais nos dois conjuntos; a mediana dos resíduos no kernel é positiva (muito próxima de zero), enquanto que, no método das superfícies de tendência, é negativa.

A tabela 1 e os gráficos das figuras 28,29 e 30 sugerem melhor desempenho do método das superfícies de tendência, nos casos de ajuste linear, e do kernel, no caso do ajuste quadrático.

Quanto as correlações dos resíduos das duas técnicas, nos casos dos valores 1, 2 e 3, são, respectivamente:  $r = 0,72$ ,  $r = 0,61$  e  $r = 0,88$ . No caso do valor 4, a correlação entre os resíduos do estimador kernel e a superfície de tendência(linear e quadrática), respectivamente, é igual a  $r = 0,95$  e  $r = 0,88$ .

As figuras 33, 34, 35, 36 e 37 mostram os gráficos cartesianos de pontos dos resíduos gerados pelas duas técnicas, para os dados do valor 1, valor 2, valor 3 e valor 4( ajustado a superfície linear), valor 4( superfície quadrática).

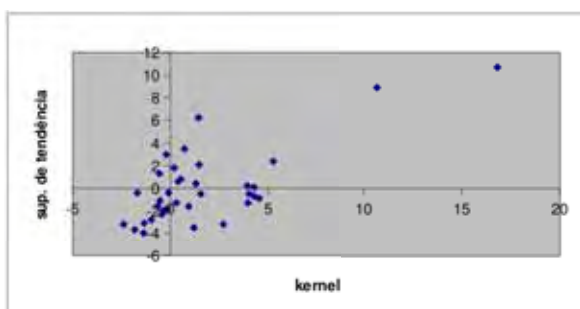


Figura 33 - Gráfico dos Resíduos do Kernel vs. superfície de tendência para os dados do valor 1

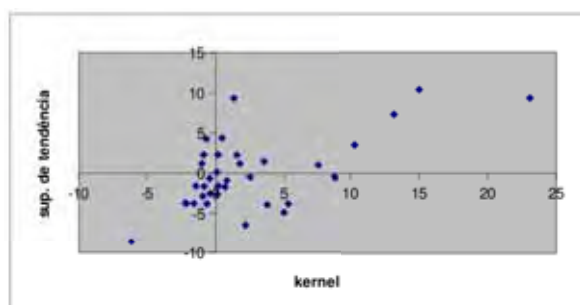


Figura 34 - Gráfico dos Resíduos do Kernel vs. superfície de tendência para os dados do valor 2

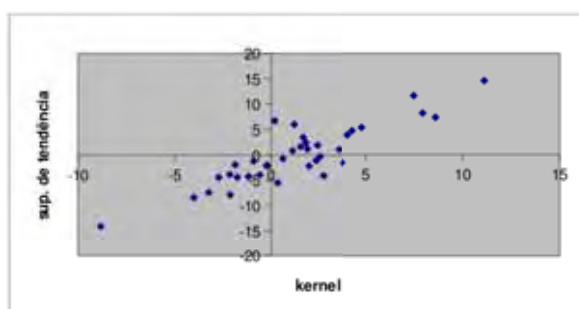


Figura 35 - Gráfico dos Resíduos do Kernel vs. superfície de tendência para os dados do valor 3

A partir das figuras 36 e 37, pode-se observar uma alta correlação positiva entre os resíduos, tanto na superfície de tendência linear, quanto na quadrática.

No caso estudado nota-se também uma grande semelhança entre o método kernel e a superfície quadrática, pela tabela 2 e figura 31, quando os valores foram atribuídos aleatoriamente nos 36 pontos .

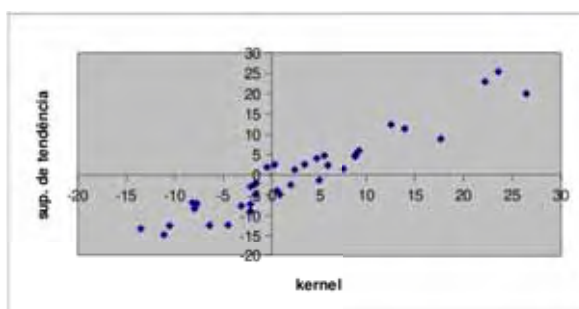


Figura 36 - Gráfico dos Resíduos do Kernel vs. superfície de tendência linear para os dados do valor 4

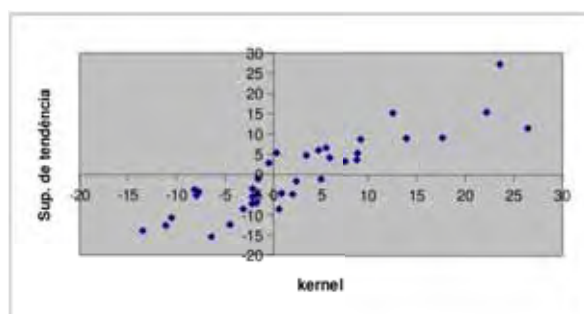


Figura 37 - Gráfico dos Resíduos do Kernel vs. superfície de tendência quadrática para os dados do valor 4

## 4 CONCLUSÕES

A utilização do estimador kernel com auxílio do *software* Terraview 4.0 (2010) forneceu bons resultados. A variação da intensidade do processo foi facilmente visualizado nas aplicações aos dados.

A concentração de casos, os aglomerados puderam facilmente ser localizados. Os focos da dengue foram evidenciados, o que pode melhorar na eficácia, imediata, das medidas de eliminação de vetores, no ápice da doença. Algumas regiões da cidade necessitavam de ações do Poder Público, medidas preventivas e de conscientização.

As regiões em amarelo indicam regiões que não se evidenciam, mas são regiões que merecem estado de alerta, para medidas futuras.

No caso dos acidentes de trânsito, o método evidenciou a região próxima a Avenida Rio Claro, num raio de aproximadamente 250 metros, para função quártico. Nessa situação, é facilitado o direcionamento de medidas do Poder Público para reduzir o número de acidentes.

Para o Poder público é de extrema importância ter conhecimento dos pontos críticos da cidade sobre acidentes de trânsito. As análises *in loco* podem gerar estratégias de intervenção viária, fiscalização e ou educação para melhor mobilidade urbana no município.



Também foi evidenciado que os melhores raios de influência  $\tau$  são diferentes nas duas aplicações. Os melhores raios para os acidentes de trânsito são menores que aqueles para os casos de dengue, ou seja os raios de influência dependem do conjunto de dados.

A melhor função kernel para visualização das regiões "quentes" no Teraview 4.0 (2010) foi a função quártica.

A tabela 1 e os gráficos das figuras 20, 21 e 22 sugerem melhor desempenho do método das superfícies de tendência, nos casos de ajuste linear, e do kernel, no caso do ajuste quadrático. Visto, que neste estudo foi apenas escolhido um conjunto de dados hipotéticos bem simples, mais estudos devem ser realizados para comparar estes métodos.

#### 4.1 Considerações finais

A técnica kernel apresentou bons resultados nos dois conjuntos de dados analisados, evidenciando as distribuições espaciais das variáveis acidentes de trânsito e de casos positivos de dengue.

Quando comparada com uma técnica da literatura, a análise de superfícies de tendência, para pequenos conjuntos de dados hipotéticos, a mesma mostrou-se mais adequada em algumas situações e menos adequada em outras.

Portanto, o kernel é uma técnica importante na análise espacial, porém nem sempre produz os melhores resultados, devendo as análises serem feitas sempre com espírito crítico. Maiores estudos são necessários.

## REFERÊNCIAS BIBLIOGRÁFICAS

BAILEY, T. C.; GATRELL, A. C. **Interactive Spatial Data Analysis**. London: Longman, 1995. 413p.

BRASIL. MINISTÉRIO DA SAÚDE. **Introdução à Estatística Espacial para a Saúde Pública**. Brasília: Ministério da Saúde, 2007.

CÂMARA, G.; CARVALHO, M. S. Análise Espacial de Eventos. In: **Análise Espacial de Dados Geográficos**, EMBRAPA, 2004.

CÂMARA, G.; MONTEIRO, A. M.; FUCKS, S. D.; CARVALHO, M. S. Análise Espacial e Geoprocessamento. In: **Análise Espacial de Dados Geográficos**, EMBRAPA, 2004.

CRESSIE, N. A. C. **Statistics for Spatial Data**. New York: John Wiley & Sons, 1993.

DAVIS, J. C. **Statistics and Data Analysis in Geology**. London: John Wiley & Sons, 1986. 646p.

DIGGLE, P. J. **Statistical Analysis of Spatial Point Patterns**. Londres: Academic Press, 1983. 148p.

LOURENÇO, P. W.; LANDIM, P. M. B.; FERREIRA, M. C. Análise da Distribuição Espacial da Produção de Monóxido de Carbono(CO) em Áreas Urbanas a partir de Superfícies de Tendência. **Geografia**, v.26, n.2, p.127–138, 2001.

NUMATA, M. Forest vegetation in the vicinity of Choshi, Coastal flora and vegetation at Choshi, Chiba Prefecture IV. **Bull. Choshi Marine Lab. Chiba Univ.**, v.3, p.28–48, 1961.

RIPLEY, B. D. Modelling spatial patterns. **Journal of the Royal Statistical Society**, v.39, p.172–212, 1977.

RIPLEY, B. D. **Spatial Statistics**. Hoboken: John Wiley & Sons, 2004. 241p.

SURFER 7.0. **Surfer @ version 7.0**. New York: Golden Software, Inc., 1999.

TERRAVIEW 4.0 . **Terraview 4.0**. São José dos Campos: INPE, 2010.

WARD, M. P.; CARPENTER, T. E. Analysis of time–space clustering in veterinary epidemiology. **Preventive Veterinary Medicine**, v.43, n.4, p.225–237, 2000.

## APÊNDICES

Valores hipotéticos gerados numa malha de 36 pontos

coord. X	coord. Y	valor 1	valor 2	valor 3	valor 4
1	1	3	3	2	28
1	2	2	2	4	19
1	3	3	3	8	7
1	4	4	6	20	13
1	5	6	9	10	14
1	6	8	15	5	34
2	1	3	3	3	7
2	2	3	5	7	3
2	3	6	9	13	22
2	4	5	12	22	2
2	5	7	12	8	10
2	6	7	15	6	9
3	1	3	2	6	24
3	2	4	6	16	22
3	3	5	8	28	8
3	4	6	12	34	27
3	5	7	16	7	11
3	6	10	19	9	9

coord. X	coord. Y	valor 1	valor 2	valor 3	valor 4
4	1	3	6	16	31
4	2	4	11	27	20
4	3	7	14	39	19
4	4	11	19	42	8
4	5	10	21	18	18
4	6	15	25	12	4
5	1	6	13	12	16
5	2	8	19	19	5
5	3	10	13	24	42
5	4	11	25	31	8
5	5	16	32	16	16
5	6	21	41	12	7
6	1	9	18	8	39
6	2	12	16	11	6
6	3	14	26	17	12
6	4	17	34	19	8
6	5	28	45	14	16
6	6	32	48	8	17