



**RODRIGO FERREIRA DE ABREU**

**DETECÇÃO DE AGRUPAMENTOS ESPAÇO-TEMPORAIS DE  
OCORRÊNCIAS DE DENGUE UTILIZANDO PROCESSOS  
PONTUAIS**

**LAVRAS – MG**

**2017**

**RODRIGO FERREIRA DE ABREU**

**DETECÇÃO DE AGRUPAMENTOS ESPAÇO-TEMPORAIS DE OCORRÊNCIAS DE  
DENGUE UTILIZANDO PROCESSOS PONTUAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Prof. Dr. João Domingos Scalon  
Orientador

**LAVRAS – MG**

**2017**

**Ficha catalográfica elaborada pela Coordenadoria de Processos Técnicos  
da Biblioteca Universitária da UFLA**

Abreu, Rodrigo Ferreira de.

Detecção de agrupamentos espaço-temporais de ocorrências de dengue utilizando processos pontuais / Rodrigo Ferreira de Abreu. – Lavras : UFLA, 2017.

73 p. : il.

Orientador: Prof. Dr. João Domingos Scalon.

Dissertação (mestrado acadêmico)–Universidade Federal de Lavras, 2017.

Bibliografia

1. Dengue. 2. Processos Pontuais. 3. Espaço-tempo. 4. Testes globais. 5. Estatística Scan. I. Scalon, João Domingos. .  
II. Título.

**RODRIGO FERREIRA DE ABREU**

**DETECÇÃO DE AGRUPAMENTOS ESPAÇO-TEMPORAIS DE OCORRÊNCIAS DE  
DENGUE UTILIZANDO PROCESSOS PONTUAIS  
DETECTION OF SPATIO-TEMPORAL CLUSTERS OF DENGUE OCCURRENCES  
USING POINT PROCESSES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 20 de janeiro de 2017.

Prof. Dr. Renato Ribeiro de Lima UFLA  
Prof. Dr. Denismar Alves Nogueira UNIFAL-MG

Prof. Dr. João Domingos Scalon  
Orientador

**LAVRAS – MG  
2017**

*A todos os homens e mulheres que procuram fazer deste mundo um lugar melhor.  
Dedico.*

## **AGRADECIMENTOS**

À Deus, por quem todas as coisas são possíveis.

À Maria mãe de Cristo, pela sua proteção.

Aos meus pais Maria da Glória e Vantuir, pelo exemplo e incentivo.

Aos meus irmãos Matheus e Letícia, por todo amor e carinho.

À minha namorada Fiorita, pelo apoio incondicional.

Ao professor Dr. João Domingos Scalon, por toda a atenção e confiança no meu trabalho.

À banca examinadora, pelas contribuições.

Aos colegas de mestrado, companheiros de empreitada.

Ao meu amigo Rafael Agostinho, companheiro de pesquisa.

À Nádia secretária do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, pela atenção.

Aos professores Martin Kulldorff, Marcelo Azevedo Costa, Peter Diggle e Edith Gabriel, pela ajuda com questões teóricas.

Ao professor Júlio Bueno, pela ajuda computacional.

Aos colegas do Grupo de Pesquisa em Estatística Espacial - GPS, pelas contribuições ao meu aprendizado em estatística espacial.

Às secretarias municipais de saúde de Três Corações e Lavras, pela disponibilização de dados para a pesquisa.

À Universidade Federal de Lavras e ao Departamento de Estatística, pela oportunidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela concessão da bolsa de mestrado.

**MUITO OBRIGADO!**

*O melhor meio para alcançar a felicidade é contribuir para a felicidade dos outros.  
(Robert Baden Powell)*

## RESUMO

A dengue é uma das doenças infecciosas que mais afetam a população mundial, onde cerca de 50 milhões de pessoas adquirem a doença todos os anos e, aproximadamente, 2,5 bilhões de pessoas estão em áreas de risco. O Brasil é um dos países em que a população está mais propensa a ser infectada pela dengue. Entender o comportamento espacial e temporal dos casos de dengue é um dos aspectos mais importantes para a tomada de decisão dos gestores de saúde pública. Assim, o objetivo deste trabalho foi apresentar e avaliar diversos métodos estatísticos que podem ser utilizados para detectar a presença de agrupamentos espaço-temporais nos casos de dengue. São abordados os seguintes métodos para detecção de agrupamentos espaço-temporais globais: teste Knox, teste de Mantel, teste de Jacquez, função K homogênea e função K não homogênea. Também foi utilizada a estatística Scan de varredura para detecção de agrupamentos locais e tempos específicos. A performance dos métodos foi avaliada a partir da aplicação dos mesmos em dados de ocorrências de dengue da cidade de Três Corações - MG, durante o período de 01/01/2010 a 31/12/2015. Os testes de Knox, Mantel e Jacquez indicaram a presença de agrupamentos espaço-temporais nas ocorrências de dengue na região de estudo. A partir das análises utilizando as funções K homogênea e não homogênea, foi possível verificar que os padrões de agrupamentos das ocorrências de dengue são resultados de efeitos de primeira ordem (intensidade) e não de efeitos de segunda ordem (dependência espaço-temporal). A análise com a estatística de varredura Scan possibilitou a identificação de seis agrupamentos espaço-temporais locais significativos na cidade de Três Corações. Os resultados mostram que cada método tem suas particularidades e, portanto, não devem ser usados individualmente para a detecção de agrupamentos espaço-temporais dos casos de dengue. Recomenda-se a utilização dos métodos combinados para uma descrição mais precisa dos agrupamentos espaço-temporais dos casos de dengue.

**Palavras-chave:** Dengue. Processos pontuais. Espaço-tempo. Testes globais. Estatística scan.

## ABSTRACT

Dengue is one of the most infectious diseases affecting the world's population, where around 50 million people get the disease every year and, approximately, 2.5 billion people are in risky areas. Brazil is one of the countries where the population is most prone to be infected with dengue. Understanding the spatial and temporal behavior of dengue cases is one of the most important aspects for the decision making of public health managers. Thus, the aim of this work was to present and evaluate several statistical methods that can be used to detect the presence of space-time clusters in dengue cases. The following methods are presented for the detection of global spatiotemporal clusterings: Knox test, Mantel test, Jacquez test, homogeneous K function and non-homogeneous K function. The Scan statistic was also used to detect clusters at specific times and locations. The performance of the methods was evaluated from the application of them in data of occurrences of dengue in the city of Três Corações - MG, during the period from 01/01/2010 to 12/31/2015. The Knox, Mantel and Jacques tests indicated the presence of spatio-temporal clusters in dengue occurrences in the study region. From the analyzes using the homogeneous and non-homogeneous K functions, it was possible to verify that the patterns of clustering of dengue occurrences are results of first order effects (intensity) and not of second order effects (spatio-temporal dependence). The analysis with scan statistic allowed the identification of six significant local spatio-temporal clusters in the city of Três Corações. The results show that each method has its peculiarities and, therefore, should not be used individually for the detection of space-time clusters of dengue cases. It is recommended to use the combined methods for a more precise description of the spatio-temporal clustering of dengue cases.

**Keywords:** Dengue. Point Process. Space-time. Global tests. Scan statistics.

## LISTA DE FIGURAS

Figura 2.1 – A fêmea do mosquito <i>Aedes Aegypti</i> picando a pele de uma pessoa. . . . .	12
Figura 2.2 – Países ou áreas onde a dengue foi reportada ou áreas em risco de infecção destacados em verde. . . . .	13
Figura 2.3 – Distribuição dos números de ocorrências da dengue no Brasil e nas cinco grandes regiões no período de 2010 a 2015. . . . .	16
Figura 2.4 – Representação de padrão pontual com intensidade: (a) Homogênea, (b) Não homogênea . . . . .	19
Figura 2.5 – Realizações típicas de processos pontuais: (a) completamente aleatório, (b) regular, (c) agrupamentos. . . . .	20
Figura 2.6 – Exemplificação para o caso de estacionariedade e isotropia. . . . .	21
Figura 2.7 – Tempos de chegada $T_i$ . . . . .	23
Figura 2.8 – Tempos entre chegadas $S_i$ . . . . .	23
Figura 2.9 – Processo de contagem $N_t$ associado a um processo pontual. . . . .	24
Figura 2.10 – Situação hipotética de um cilindro localizado sobre um possível agrupamento. Os pontos representam casos e a sua altura em relação ao plano $xy$ é o tempo decorrido durante o período de estudo. . . . .	37
Figura 3.1 – Município de Três Corações destacado em vermelho no sul do estado de Minas Gerais. . . . .	40
Figura 3.2 – Gráfico da precipitação acumulada média mensal (a) e das temperaturas mínima, máxima e média mensal (b) para a cidade de Três Corações. . . . .	40
Figura 3.3 – Contorno da cidade de Três Corações obtido através do <i>Software</i> Google Earth. . . . .	42
Figura 4.1 – Distribuição mensal das ocorrências de Dengue em Três Corações no período de janeiro de 2010 a dezembro de 2015. . . . .	46
Figura 4.2 – Histogramas das distribuições empíricas das estatísticas calculadas para os testes (A) Knox e (B) Mantel partir de um procedimento de Monte Carlo com $N = 999$ repetições e a localização da estatística observada para cada um dos três testes. . . . .	47

Figura 4.3 – Boxplots das estatísticas $T_k$ obtidas no teste Jacquez para $k = 1, \dots, 20$ vizinhos mais próximos a partir de um procedimento de Monte Carlo com $N = 999$ repetições, e em vermelho a localização da estatística calculada para os dados originais. . . . .	48
Figura 4.4 – Histograma da distribuição empírica das estatísticas calculadas $D$ a partir de um procedimento de Monte Carlo com $N = 999$ repetições utilizando a função $K$ homogênea e a localização da estatística observada. . . . .	49
Figura 4.5 – Gráfico da diferença entre a função $K$ homogênea observada e o esperado pelo processo de Poisson homogêneo. . . . .	49
Figura 4.6 – Histograma da distribuição empírica das estatísticas calculadas $D$ a partir de um procedimento de Monte Carlo com $N = 999$ repetições utilizando a função $K$ não homogênea e a localização da estatística observada. . . . .	50
Figura 4.7 – Resumo dos resultados obtidos pela análise do métodos de permutação espaço-temporal com o <i>software</i> SaTScan <sup>TM</sup> (unidades do mapa dadas em quilômetros). . . . .	51

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>12</b>
<b>2.1</b>	<b>A dengue</b>	<b>12</b>
<b>2.1.1</b>	<b>Características epidemiológicas</b>	<b>12</b>
<b>2.1.2</b>	<b>Histórico</b>	<b>14</b>
<b>2.2</b>	<b>Estatística espacial</b>	<b>16</b>
<b>2.3</b>	<b>Processos pontuais espaciais</b>	<b>18</b>
<b>2.4</b>	<b>Processos pontuais no tempo</b>	<b>22</b>
<b>2.5</b>	<b>Processos pontuais espaço-temporais</b>	<b>24</b>
<b>2.6</b>	<b>Métodos para detecção de agrupamentos espaço-temporais</b>	<b>29</b>
<b>2.6.1</b>	<b>Função K</b>	<b>29</b>
<b>2.6.2</b>	<b>Teste Knox</b>	<b>33</b>
<b>2.6.3</b>	<b>Teste Mantel</b>	<b>35</b>
<b>2.6.4</b>	<b>Teste Jacquez</b>	<b>36</b>
<b>2.6.5</b>	<b>Estatística <i>scan</i> - modelo de permutação no espaço-tempo</b>	<b>37</b>
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>40</b>
<b>3.1</b>	<b>Área de estudo</b>	<b>40</b>
<b>3.2</b>	<b>O conjunto de dados</b>	<b>41</b>
<b>3.3</b>	<b>Métodos estatísticos</b>	<b>43</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>45</b>
<b>5</b>	<b>CONCLUSÕES</b>	<b>53</b>
	<b>REFERÊNCIAS</b>	<b>54</b>
	<b>APENDICE A – Estimação das intensidades das ocorrências de dengue por mês no período de 2010 a 2015 utilizando uma kernel gaussiana. Escala dada em número de casos por unidade de área (km<sup>2</sup>) e unidade de tempo (dia)</b>	<b>58</b>
	<b>ANEXO A – Estimação da intensidade por kernel</b>	<b>61</b>
	<b>ANEXO B – Códigos do R para os testes</b>	<b>64</b>

## 1 INTRODUÇÃO

Estudar e compreender os mecanismos pelos quais as doenças infecciosas se propagam nas populações humanas é de fundamental importância no combate e controle das mesmas. Nesse sentido, os estudos epidemiológicos têm lugar central. De acordo com Bonita, Beaglehole e Kjellström (2006), a epidemiologia pode ser definida como o estudo da distribuição e dos determinantes de estados ou eventos relacionados à saúde em populações específicas e sua aplicação na prevenção e controle dos problemas de saúde.

No Brasil uma doença infecciosa que tem causado grandes transtornos desde a década de 1980 é a dengue. Milhões de pessoas já foram infectadas com a doença e muitas pessoas vieram a óbito em função disso. O Brasil está localizado na zona tropical com clima quente e úmido e esses são ingredientes que contribuem na propagação da dengue. Além disso, de acordo com Mendonça, Souza e Dutra (2009) diversos fatores concorrem para a formação das epidemias de dengue nos países tropicais e subtropicais, destacando-se a proliferação do mosquito transmissor, a intensa e desordenada urbanização, infra-estrutura inadequada e despreparo da população e dos agentes de saúde no controle da doença. Os autores destacam ainda que o vetor da dengue desenvolve cada vez mais mecanismos de resistências às diversas formas de controle.

É necessário então, que se busque sempre mais mecanismos que possam auxiliar na prevenção e no combate da dengue e a compreensão da forma como o mosquito transmissor se propaga. Assim, o estudo da distribuição espaço-temporal da dengue por meio das análises dos padrões pontuais das ocorrências pode ser uma ferramenta útil nesse processo.

Em relação a estudos sobre os padrões de dispersão de insetos, de acordo com Scalon et al. (2011), os mesmos geralmente envolvem a contagem do número de eventos (ocorrências), em subconjuntos da região de estudo e tem sido descritos usando índices como a razão de variância das médias de Fisher e índice de aglomeração de Morisita. Os autores salientam, no entanto que existe uma perda de informação ao tratar os dados desse modo, pois não existe qualquer consideração a respeito das localizações pontuais dos eventos, de modo que a maior parte da informação espacial é perdida. Além disso, esses métodos analisam apenas as informações espaciais dos eventos, não levando em conta a informação temporal da ocorrência do evento.

O objetivo desta dissertação é apresentar e avaliar diversos métodos estatísticos baseados na teoria de processos pontuais que podem ser utilizados para detectar a presença de agrupa-

mentos espaço-temporais nos casos de dengue. Para atingir esses objetivos, tais métodos foram aplicados em dados constituídos pelas informações espaciais e temporais das ocorrências de dengue na cidade de Três Corações - MG no período de 01/01/2010 a 31/12/2016.

Deste modo, este trabalho tem a seguinte organização:

- Revisão bibliográfica: é apresentada uma revisão de literatura contendo informações sobre as características epidemiológicas e o histórico da dengue; uma breve introdução sobre a estatística espacial e as diferentes áreas de estudo; uma abordagem da teoria de processos pontuais sob os aspectos puramente espacial e temporal e de modo conjunto, sob o aspecto espaço-temporal. Também são apresentados os métodos utilizados para o estudo do padrão espaço-temporal das ocorrências.
- Material e métodos: são apresentadas as informações sobre a cidade de Três Corações relacionadas à localização, dados populacionais e clima; informações sobre o conjunto de dados e a forma como foram tratados, e os *softwares* necessários para a execução das análises.
- Resultados e discussão: são apresentados os resultados obtidos em cada análise executada e uma discussão a respeito dos mesmos.
- Conclusão: são apresentadas a conclusão e as considerações a respeito do trabalho.
- Referências: são apresentadas as referências dos trabalhos citados no texto que foram utilizados como base para esta pesquisa.
- Apêndice: são expostos os mapas de intensidade das ocorrências de dengue em três corações para todos os meses de janeiro de 2010 a dezembro de 2015.
- Anexos: são apresentados os códigos de programação utilizados para a execução dos testes abordados e uma pequena abordagem a respeito da estimação de intensidades utilizando estimadores kernel.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 A dengue

#### 2.1.1 Características epidemiológicas

A dengue é uma doença viral que é transmitida através da picada dos mosquitos *Aedes Aegypti* e *Aedes Albopictus*. De acordo com a Organização Mundial da Saúde (WORLD HEALTH ORGANIZATION, 2009) existem quatro sorotipos do vírus - DENV1, DENV2, DENV3 e DENV4 - que apresentam características genéticas distintas. No Brasil, o principal vetor da doença é o mosquito *Aedes Aegypti* e não há casos registrados de dengue causada pelo *Aedes Albopictus*, cujas ocorrências estão predominantemente na Ásia e na região do oceano Índico.

De acordo com Holmes e Twiddy (2003), a dengue é integrante do gênero *Flavivirus*, ou seja, é um vírus que causa doenças no homem e em outros vertebrados e é classificado como um *Arbovirus*, pois é transmitido por artrópodes, neste caso, mosquitos.

Em relação à infecção da doença, é importante destacar que apenas a fêmea do mosquito (Figura 2.1) transmite o vírus pela picada, uma vez que faz isso para sugar sangue para amadurecer seus ovos. Durante a picada, o mosquito libera saliva contendo substâncias anti-coagulantes e analgésicas, que o permite não ser notado e assim, sugando a maior quantidade de sangue possível. Neste processo o vírus é injetado na corrente sanguínea junto com a saliva do mosquito (FUNDAÇÃO INSTITUTO OSWALDO CRUZ, 2016). De acordo a Organização Mundial da Saúde (WORLD HEALTH ORGANIZATION, 2016) a maioria das fêmeas do *Aedes Aegypti* podem passar a vida dentro ou em torno das casas onde emergem como adultas, voam normalmente, em média, 200 metros e podem picar por toda a vida adulta que dura aproximadamente 30 dias.

Figura 2.1 – A fêmea do mosquito *Aedes Aegypti* picando a pele de uma pessoa.



Fonte: <http://www.ioc.fiocruz.br/dengue/textos/aedesvetoredoenca.html>

De acordo com a Organização Mundial da Saúde (WORLD HEALTH ORGANIZATION, 2009), no mundo existem em torno de 2,5 bilhões de pessoas vivendo em áreas de risco (Figura 2.2) de infecção por dengue e estima-se, que todos os anos, em média, 50 milhões de pessoas são infectadas pela doença.

Figura 2.2 – Países ou áreas onde a dengue foi reportada ou áreas em risco de infecção destacados em verde.



Fonte: <http://apps.who.int/ithmap/index.html>

De acordo com Holmes e Twiddy (2003), apesar do alto número de pessoas infectadas todos os anos, uma grande quantidade dessas infecções podem ser assintomáticas, ou seja, podem não apresentar nenhum sintoma. Nesses casos, uma pessoa pode ter dengue e não perceber, descobrindo isso através de exames somente após contrair o vírus novamente. Segundo esses autores, se a infecção pela dengue resultar em doença evidente, o resultado mais comum é uma doença febril aguda, semelhante à gripe, sendo esta a Dengue Clássica ou Febre de Dengue (DF - *Dengue Fever*). Em uma minoria de casos, isso evolui para hemorragia espontânea, que é a Febre Hemorrágica de Dengue (DHF - *Dengue Hemorrhagic Fever*) e, em um caso mais sério, à Síndrome de Choque de Dengue (DSS - *Dengue Shock Syndrome*), podendo nesses últimos casos levar uma pessoa ao óbito.

Ao contrair um dos sorotipos da doença e se recuperar dessa infecção primária, Holmes e Twiddy (2003) afirmam que essas pessoas desenvolvem respostas robustas de anticorpos que reagem cruzadamente com os quatro sorotipos. Porém, apesar dessa reatividade cruzada, os anticorpos impedem apenas a reinfeção pelo mesmo sorotipo e, assim, os indivíduos são suscetíveis a uma segunda infecção por um sorotipo diferente. Esses autores alertam ainda, que se uma pessoa adquirir uma infecção secundária com um novo sorotipo, ela estará sob um

risco muito maior de desenvolvimento de DHF, indicando que a imunidade pré-existente pode exacerbar a doença.

Em relação à vacinação contra a dengue, atualmente está disponível a vacina Dengvaxia<sup>®</sup> que é produzida pelo laboratório francês Sanofi Pasteur, cuja aprovação do registro no Brasil foi publicada em 28 de dezembro de 2015 no Diário Oficial da União (BRASIL, 2015). Essa vacina, no entanto, não é acessível a toda a população, podendo o preço variar de R\$132,76 até R\$138,53. Além disso, a vacina apresenta eficácia global de apenas 60% e são necessárias três doses, aplicadas com intervalos de seis meses, para que o efeito seja atingido completamente, sendo indicada para pessoas com idades entre 9 e 45 anos.

No Brasil, está sendo produzida a Vacina Dengue Butantan pelo Instituto Butantan em parceria com o NIH (*National Institutes of Health*), o Instituto Adolfo Lutz, o Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HCFMUSP) – Instituto Central e Instituto da Criança, com o fomento do Banco Nacional de Desenvolvimento Econômico (BNDES) e da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp). No presente momento, de acordo com o Instituto Butantan (BUTANTAN, 2016), a vacina se encontra na fase III de ensaios clínicos e, na próxima fase, serão realizados os testes em humanos, abrangendo pessoas de todo o território nacional.

### 2.1.2 Histórico

Através de uma análise filogenética, ou seja, através de uma representação da história evolutiva, Holmes e Twiddy (2003) mostram que a dengue é originariamente um vírus do macaco e que, alguns autores como Gaunt et al. (2001) sugerem uma origem africana para o vírus, em função de que muitos dos mais divergentes *Flavivirus* transmitidos por mosquitos circulam exclusivamente na África e muitas vezes infectam primatas.

Ainda de acordo com Holmes e Twiddy (2003), parece que o mosquito *Aedes Aegypti* seria originário da África. Contudo, a presença de todos os quatro sorotipos da dengue em macacos e humanos na Ásia, sugere através da análise filogenética, que o vírus tem origem asiática e não africana. Os autores argumentam que para a definição do local de origem do vírus da dengue é necessário ainda uma análise filogenética molecular mais precisa.

A respeito do termo dengue como o nome de uma doença, Rigau-Pérez (1998) afirma que a origem do mesmo tem sido atribuída a fontes tão diversas quanto a doença está geograficamente dispersa. Apoiado por uma definição do Oxford English Dictionary (SIMPSON;

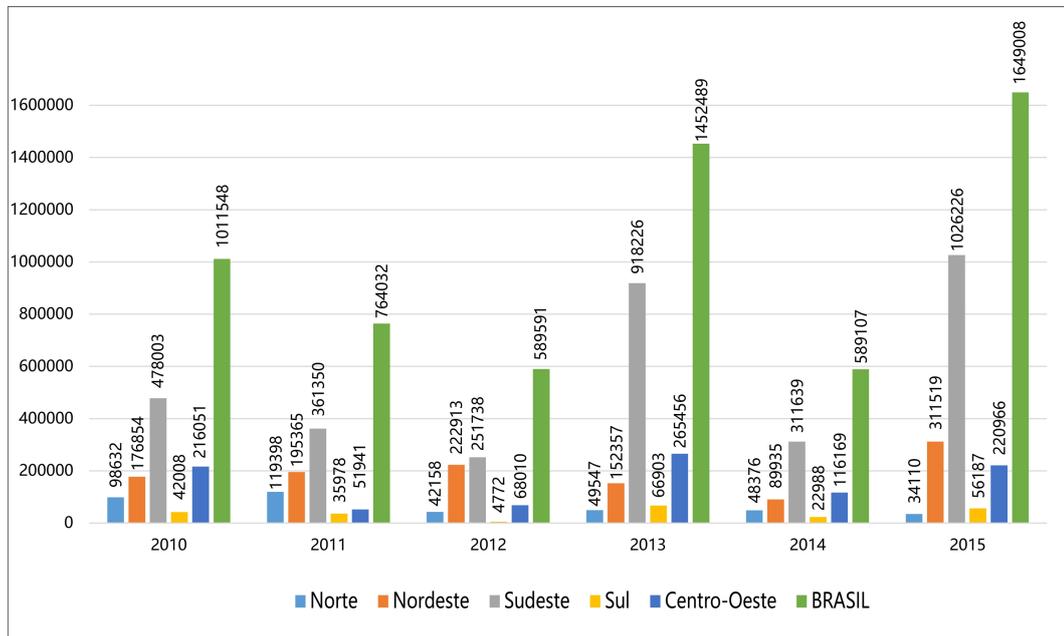
WEINER, 1989), o autor mostra que o termo se originou na frase “*Swahili Ka dinga pepo*” (uma espécie de gripe súbita como um ataque de um espírito maligno ou praga) e que de acordo com o dicionário a frase e a doença atravessaram da África Oriental para o Caribe em 1827. Rigau-Pérez (1998) afirma também, que em Cuba essa frase foi popularmente identificada com a palavra espanhola “dengue” e nas Índias Ocidentais Britânicas com a palavra “dândi”. Na Espanha em 1801, a dengue foi utilizada para descrever uma doença febril aguda com dores ósseas e articulares, hemorragia e icterícia.

No continente americano, a chegada da dengue se deu, de acordo com Dick et al. (2012), no século XVII, sendo que os primeiros relatos de suspeitas de epidemias foram em Martinica e Guadalupe em 1635 e em 1699 no Panamá. No Brasil os primeiros relatos são para o período de 1845 a 1849. Surtos subsequentes aconteceram no Brasil, segundo os autores, entre 1880 e 1912 na cidade de Curitiba - PR, em 1916 no estado do Rio Grande do Sul e em 1922 em Niterói - RJ. Os primeiros casos de dengue confirmados por exame laboratorial no Brasil ocorreram, conforme Viana e Ignotti (2013), na cidade de Boa Vista - RR em meados de 1981-1982. Nessa ocasião foram isolados os sorotipos DENV1 e DENV4, sendo 11000 pessoas infectadas. Desde então ocorreram muitos outros surtos da doença que atingiram todas as regiões do país.

Em relação a períodos recentes como de 2010 a 2015, dados do Ministério da Saúde (BRASIL, 2016) mostram que as ocorrências no Brasil têm apresentado números expressivos. No ano de 2010 o número de ocorrências foi de 1.011.548 casos. Já nos dois anos seguintes houve uma queda e os números foram de 764.032 para 2011 e 589.591 para 2012. No ano de 2013 foram registrados 1.452.489 casos, um aumento de mais de 246% em relação ao ano anterior. Em 2014 o número de casos voltou a cair, sendo registradas 589.107 ocorrências e em 2015, novamente um aumento expressivo com 1.649.008 casos, o que representa um aumento de quase 280% em relação a 2014. Durante esse intervalo de tempo, teve destaque a região Sudeste, que apresentou em todos os anos o maior número de casos, como se pode observar no gráfico da Figura 2.3, sendo que nos anos de 2013 e 2015, a região Sudeste chegou a ser responsável por mais de 62% de todas as ocorrências.

Para a cidade de Três Corações, que foi objeto de estudo deste trabalho, considerando o mesmo período de 2010 a 2015, em 2010 e 2011 foram registrados respectivamente, 381 e 471 casos de dengue. Nos dois anos seguintes esse número caiu para 55 casos em 2012 e 58 em 2013. Já nos anos de 2014 e 2015 o aumento foi substancial, sendo que em 2014 foram 1056 ocorrências e em 2015 o número total de casos registrados foi de 1435.

Figura 2.3 – Distribuição dos números de ocorrências da dengue no Brasil e nas cinco grandes regiões no período de 2010 a 2015.



Fonte: Gráfico construído a partir das informações de BRASIL (2016)

## 2.2 Estatística espacial

Qualquer tipo de dado utilizado em análises estatísticas é proveniente de alguma posição geográfica no espaço e que foi coletado em um determinado instante de tempo. Essas informações espaciais e temporais, para muitos tipos de estudo, não são importantes para as análises. Entretanto, em algumas situações pode-se estar interessado em estudar o fenômeno, ou experimento em questão, considerando as informações espaciais ou temporais dos dados, ou ainda as duas informações simultaneamente (análise espaço-temporal).

De acordo com Cressie (1993) existem diferentes tipos de dados que têm uma estrutura espacial e que podem ser categorizados do seguinte modo:

- **Geoestatística:** São dados de superfície contínua, em que o fenômeno está distribuído continuamente sobre uma superfície, como, por exemplo, teor de um determinado nutriente no solo, temperatura, estimativa da extensão de uma jazida mineral, etc.
- **Dados de área (*lattice*):** São dados indexados a sub-regiões (polígonos) que constituem uma partição de um domínio contínuo. São exemplos os dados agregados por município, setor censitário, bairros, etc, em que não se sabe exatamente onde os eventos ocorrem mas dispõe-se de um valor agregado a uma área.

- **Dados de processos pontuais:** São dados em que a informação é a própria localização do evento. Ocorrência de crimes, epidemias e localização de plantas de uma determinada espécie são alguns exemplos desse tipo de dados. Cada localização pode estar associada a alguma variável marca<sup>1</sup>.

Esses tipos de dados são objetos de estudo da Estatística Espacial, que pode ser caracterizada como o conjunto de conceitos e métodos de análise de dados em que a localização geográfica é utilizada explicitamente, podendo assim ser modelada como um processo estocástico (CRESSIE, 1993).

De acordo com Snyder e Miller (2012), um processo estocástico é uma família ou uma coleção de variáveis aleatórias que podem ser identificadas ou localizadas (indexadas), por um determinado parâmetro  $\theta$  que toma valores em um conjunto de parâmetros  $\Theta$  chamado de conjunto de índices do processo. Se  $\Theta$  é um conjunto contável  $\{\theta_1, \theta_2, \dots, \theta_n\}$ , tem-se então, segundo os autores, um processo de parâmetro discreto. Se  $\Theta \subset \mathbb{R}^d$  o processo é dito de parâmetro contínuo. Esse parâmetro pode ser o tempo, o espaço ou ainda uma combinação desses dois (espaço-tempo).

A partir disso, um processo estocástico pode ser representado no espaço, no tempo e no espaço-tempo, respectivamente por:

$$\{Z(\mathbf{s}) : \mathbf{s} \in S \subset \mathbb{R}^d\}, \quad (2.1)$$

$$\{Z(t) : t \in T \subset \mathbb{R}\} \text{ e} \quad (2.2)$$

$$\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in S \times T \subset \mathbb{R}^d \times \mathbb{R}\}. \quad (2.3)$$

Considerando  $t$  como um parâmetro de tempo contínuo, de acordo com Snyder e Miller (2012),  $T$  em (2.2) e (2.3) é um subconjunto de  $\mathbb{R}$  correspondente a um intervalo de tempo e pode ser considerado como o tempo decorrido após algum instante inicial,  $T = \{t : t \geq t_0\}$ .

A partir da equação (2.1) tem-se, segundo Shabenberger e Gotway (2005), que um processo espacial é uma coleção de variáveis aleatórias, que são indexadas por algum conjunto  $S \subset \mathbb{R}^d$  contendo as coordenadas espaciais  $\mathbf{s} = [s_1, s_2, \dots, s_d]'$  e se  $d = 2$ , ou seja, (2.1) é um processo no plano bidimensional, as coordenadas podem ser representadas por  $\mathbf{s} = [x, y]'$ , sendo  $x$  as coordenadas no eixo das abscissas e  $y$  as coordenadas no eixo das ordenadas.

<sup>1</sup> Marca é uma variável aleatória atribuída a cada ocorrência. São exemplos: tipos de crimes, idade da pessoa que contraiu a doença, etc.

Em (2.3) tem-se, conforme Gneiting e Guttorp (2010), que o processo varia tanto em função das localizações espaciais,  $\mathbf{s} \in S \subset \mathbb{R}^d$ , como no tempo,  $t \in T \subset \mathbb{R}$  e considerando  $d = 2$ , tem-se que a análise espaço temporal está definida em  $\mathbb{R}^2 \times \mathbb{R}$ .

As análises desses processos estão relacionadas basicamente ao estudo dos efeitos de primeira e segunda ordem, ou seja, nas variações do valor médio do processo e na verificação da presença ou ausência de dependência espacial, respectivamente.

Para facilitar as análises considera-se, em geral, as suposições de estacionariedade e isotropia sobre o processo estocástico em estudo. Tais conceitos serão abordados nos tópicos seguintes. Neste trabalho, as análises foram feitas a partir dos métodos da teoria de processos pontuais. Desse modo, a seguir está apresentada, em um primeiro momento, uma introdução à teoria de processos pontuais considerando somente os aspectos puramente espacial e puramente temporal e posteriormente, um maior detalhamento para a teoria de processos pontuais espaço-temporais, que foi o foco desta investigação.

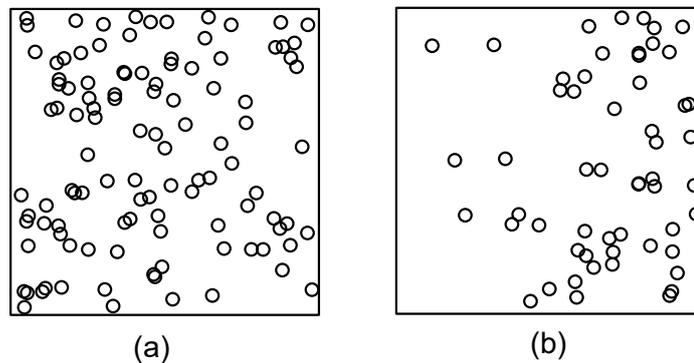
### 2.3 Processos pontuais espaciais

Estudos pautados na abordagem de processos pontuais espaciais são úteis para se estudar o padrão de ocorrências de eventos em diferentes áreas, como por exemplo, arqueologia, silvicultura, astronomia, ecologia e biologia, entre outras. Formalmente, um processo pontual espacial é, segundo Diggle (2013), um mecanismo estocástico que gera um conjunto contável finito ou infinito de pontos (denominados eventos)  $\{\mathbf{s}_i : i = 1, 2, \dots\}$  tais que  $\mathbf{s}_i \in S \subset \mathbb{R}^2$ , ou seja, cada variável aleatória denota a localização de um evento específico no espaço bidimensional.

De acordo com Baddeley (2007), pode-se descrever os processos pontuais em termos de momentos como o valor esperado (efeitos de primeira ordem) e variância ou covariância (efeitos de segunda ordem), de modo análogo ao que é feito para uma variável aleatória. Essas quantidades, segundo o autor, são úteis no estudo teórico de processos pontuais e na inferência estatística sobre padrões pontuais.

Os efeitos de primeira ordem, de acordo com Câmara e Carvalho (2003, p. 3), são aqueles “considerados globais ou de larga escala, e correspondem a variações no valor médio do processo no espaço. Neste caso, estamos interessados na intensidade do processo, isto é, no número de eventos por unidade de área”. A intensidade mede a frequência dos eventos e pode ser constante (homogênea) ou variar na região de estudo (não homogênea), conforme pode ser visto na Figura 2.4 .

Figura 2.4 – Representação de padrão pontual com intensidade: (a) Homogênea, (b) Não homogênea



Formalmente as propriedades de primeira ordem do processo pontual espacial podem ser definidas, conforme Diggle (2013), pela função intensidade dada por

$$\lambda(\mathbf{s}) = \lim_{|ds| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(ds)]}{ds} \right\}, \quad (2.4)$$

em que  $ds$  define uma pequena região em torno da localização  $\mathbf{s}$ ,  $|ds|$  é a sua área e  $N(ds)$  é número de eventos localizados dentro de  $ds$ .

Os efeitos de segunda ordem, ainda de acordo Câmara e Carvalho (2003, p.3), são aqueles “denominados locais ou de pequena escala, representam a dependência espacial no processo, proveniente da estrutura de correlação espacial”.

Na análise dos efeitos de segunda ordem procura-se responder se os eventos estão mais próximos ou mais distantes entre si, exibindo algum grau de relação do que seria esperado no caso de completa aleatoriedade espacial (CAE).

Segundo Schabenberger e Gotway (2005), a CAE ocorre se o padrão pontual atende aos seguintes critérios:

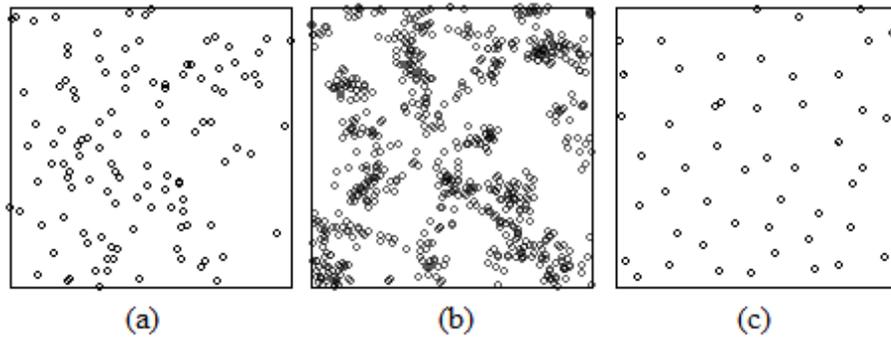
- O número médio de eventos por unidade de área, ou seja, a intensidade  $\lambda(\mathbf{s})$  é homogênea em  $S$ ,
- o número de eventos em duas sub-regiões sem sobreposição  $A_1$  e  $A_2$  são independentes,
- o número de eventos em qualquer sub-região segue distribuição Poisson.

De acordo com esses autores, desse modo, os eventos distribuem-se uniformemente e independentemente ao longo do domínio e a manifestação matemática da completa aleatoriedade espacial é o chamado processo de Poisson homogêneo. Este é um processo que não exibe

qualquer estrutura espacial e serve como hipótese nula para muitas investigações estatísticas em padrões pontuais. Os padrões de pontos observados são testados inicialmente contra a hipótese de CAE e se essa hipótese for rejeitada, então o pesquisador geralmente segue com análises mais específicas para investigar a natureza do padrão pontual espacial.

Na Figura 2.5 estão representadas as três configurações típicas de padrões pontuais: (a) independência (CAE), (b) agrupamentos e (c) regularidade.

Figura 2.5 – Realizações típicas de processos pontuais: (a) completamente aleatório, (b) regular, (c) agrupamentos.



Formalmente, as propriedades de segunda ordem podem ser definidas, conforme Diggle (2013), pela função intensidade de segunda ordem dada por

$$\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lim_{|d\mathbf{s}_i|, |d\mathbf{s}_j| \rightarrow 0} \frac{\mathbb{E}[N(d\mathbf{s}_i)N(d\mathbf{s}_j)]}{|d\mathbf{s}_i||d\mathbf{s}_j|}, \quad (2.5)$$

em que  $N(d\mathbf{s}_i)$  e  $N(d\mathbf{s}_j)$  representam o número de eventos dentro pequenos círculos centrados nas localizações  $\mathbf{s}_i$  e  $\mathbf{s}_j$  com áreas iguais a  $|d\mathbf{s}_i|$  e  $|d\mathbf{s}_j|$ , respectivamente.

De acordo com Bivand et al. (2008) a intensidade de segunda ordem de dois pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$  representa a probabilidade de qualquer par de eventos ocorrer nas vizinhanças de  $\mathbf{s}_i$  e  $\mathbf{s}_j$ , respectivamente.

Outras medidas essencialmente equivalentes de propriedades de segunda ordem de um processo pontual espacial são dadas, conforme Diggle (2013) pela função de correlação entre pares de eventos

$$g(\mathbf{s}_i, \mathbf{s}_j) = \frac{\lambda_2(\mathbf{s}_i, \mathbf{s}_j)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \quad (2.6)$$

e a densidade de covariância dada por

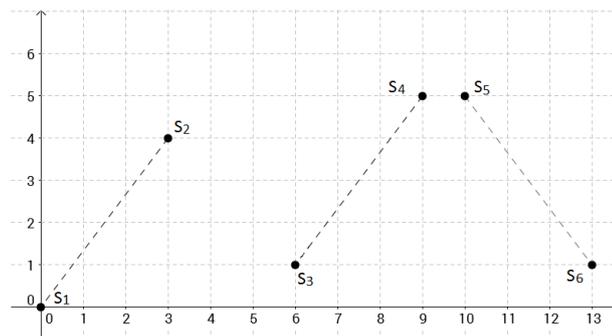
$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\mathbf{s}_i, \mathbf{s}_j) - \lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j). \quad (2.7)$$

No estudo das propriedades de segunda ordem, as suposições de estacionariedade e isotropia sobre um processo pontual são feitas a fim de facilitar as análises. De acordo com Lloyd (2010), o termo estacionariedade é muitas vezes tomado para se referir ao resultado de algum processo que tem propriedades semelhantes em todos os locais na região de interesse. Mais especificamente, o processo é dito estacionário se as suas propriedades estatísticas são invariantes sobre translação, ou seja, a intensidade de primeira ordem é constante em toda parte ( $\lambda(\mathbf{s}) = \lambda$ ) e a relação entre dois eventos, medida pela intensidade de segunda ordem, depende somente da sua distância relativa ( $\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\mathbf{s}_i - \mathbf{s}_j)$ ).

Schabenberger e Gotway (2005) apontam que, se além de estacionário, o processo for isotrópico, a função intensidade de segunda ordem depende então apenas das distâncias entre os eventos e não da direção entre eles, ou seja,  $\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\|\mathbf{s}_i - \mathbf{s}_j\|) = \lambda_2(u)$ .

Para exemplificar essa situação, considere a configuração pontual mostrada na Figura 2.6. Os pares de pontos  $\mathbf{s}_1\mathbf{s}_2$ ,  $\mathbf{s}_3\mathbf{s}_4$  e  $\mathbf{s}_5\mathbf{s}_6$  estão todos separados por uma distância de 5 unidades de medida, sendo que  $\mathbf{s}_1\mathbf{s}_2$  e  $\mathbf{s}_3\mathbf{s}_4$  estão apenas transladados um em relação ao outro, ou seja, estão em uma mesma direção, e  $\mathbf{s}_5\mathbf{s}_6$  está em uma direção diferente. Se o processo pontual que gerou esses pontos for estacionário, então a relação existente entre  $\mathbf{s}_1$  e  $\mathbf{s}_2$  é a mesma entre  $\mathbf{s}_3$  e  $\mathbf{s}_4$ , mas é diferente de  $\mathbf{s}_5$  e  $\mathbf{s}_6$ . Se o processo for isotrópico, mesmo as direções entre esses três pares de pontos não sendo todas iguais, a relação entre eles será a mesma.

Figura 2.6 – Exemplificação para o caso de estacionariedade e isotropia.



Para o caso em que o processo pontual espacial não é homogêneo, ou seja, a intensidade de primeira ordem não é constante, faz-se necessário considerar a tendência espacial nos procedimentos de análise. Nesse sentido, Baddeley, Møller e Waagepetersen (2000) introduziram uma suposição mais fraca de estacionariedade, onde o processo pontual espacial é dito estacionário de intensidade de segunda ordem reponderada, em que na estimação das propriedades de segunda ordem, cada ponto é ponderado pela intensidade inerente a ele. Uma propriedade

associada a essa suposição é que a função de correlação entre pares de eventos só depende das distâncias espaciais, ou seja,

$$g(u) = \frac{\lambda_2(u)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)}, \quad (2.8)$$

em que  $u = \|\mathbf{s}_i - \mathbf{s}_j\|$ .

Esses conceitos possibilitam a estruturação da função K proposta por Ripley (1976,1977), que é uma medida das características de segunda ordem de um processo pontual espacial, sendo definida por

$$K(u) = \lambda^{-1}E[N(u)], \quad (2.9)$$

em que  $N(u)$  é o número de eventos encontrados dentro de uma distância  $u$  e  $\lambda$  é a intensidade de primeira ordem do processo pontual. Essa abordagem, no entanto, não foi detalhada aqui, o leitor interessado pode se aprofundar, além do autor citado, em Baddeley, Møller, e Waagepetersen (2000) e Diggle (2013).

## 2.4 Processos pontuais no tempo

Um processo pontual temporal pode ser descrito como um processo estocástico em que a realização do mesmo consiste em um conjunto de eventos indexados no tempo, dados por  $\{t_i : i = 1, \dots\}$  com  $t_i \in \mathbb{R}^+$ . De acordo com Baddeley (2007), um processo pontual no tempo é um modelo útil para uma sequência de tempos aleatórios de ocorrência de um evento particular, como por exemplo, os tempos em que o Corpo de Bombeiros recebe chamadas de emergência podem ser modelados como um processo pontual. Cada chamada de emergência acontece em um instante, ou ponto, do tempo. Desse modo haverá um número aleatório de tais chamadas em qualquer período de tempo e elas ocorrerão em instantes de tempo aleatórios.

De acordo com Cox e Isham (1980), um processo pontual temporal pode ser estudado da mesma maneira que o processo espacial, ou seja, pelas propriedades de primeira e segunda ordem. Cox e Isham (1980) mostram que a função intensidade que descreve o número médio de eventos por unidade de tempo pode ser expressa por

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{E}[N(dt)]}{|dt|}, \quad (2.10)$$

em que  $N(dt)$  é o número de eventos dentro de um pequeno intervalo em torno do tempo  $t$ , com comprimento igual a  $|dt|$ .

Se essa intensidade for constante em todo o intervalo, ou seja,  $\lambda(t) = \lambda$ , então o processo pontual temporal é dito homogêneo.

As propriedades de segunda ordem podem ser descritas, pela função intensidade de segunda ordem dada por

$$\lambda(t_i, t_j) = \lim_{|t_i|, |t_j| \rightarrow 0} \frac{\mathbb{E}[N(t_i)N(t_j)]}{\lambda(dt_i)\lambda(dt_j)}. \quad (2.11)$$

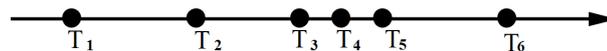
Outra maneira de se analisar as propriedades de segunda ordem de um processo pontual temporal, assim como no caso espacial, pode ser feito pela função K temporal. Diggle et al. (1995) adaptaram a função K para o caso temporal como sendo

$$K(v) = \lambda^{-1} E[N(v)], \quad (2.12)$$

em que, de modo análogo ao caso espacial,  $N(v)$  é o número de eventos encontrados dentro de um intervalo de tempo  $v = |t_i - t_j|$ .

De acordo com Baddeley (2007), matematicamente falando, um processo pontual temporal pode ser tratado de várias maneiras diferentes. Podemos estudar os tempos de chegada  $T_1 < T_2 < \dots$ , em que  $T_i$  é o instante em que chega o  $i$ -ésimo evento, como chamadas de emergência, por exemplo (Figura 2.7). De acordo com esse autor, usar essas variáveis aleatórias é a maneira mais direta de lidar com o padrão de pontos, mas seu uso é complicado pelo fato de serem fortemente dependentes, uma vez que  $T_i < T_{i+1}$ .

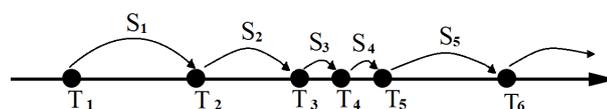
Figura 2.7 – Tempos de chegada  $T_i$ .



Fonte: Adaptado de Baddeley (2007)

Baddeley (2007) mostra que alternativamente, podemos estudar os tempos de interchegada  $S_i = T_{i+1} - T_i$ , como ilustrado na Figura 2.8. Estes têm a vantagem de que, para alguns modelos especiais como o Poisson, as variáveis aleatórias  $S_1, S_2, \dots$  são independentes.

Figura 2.8 – Tempos entre chegadas  $S_i$ .



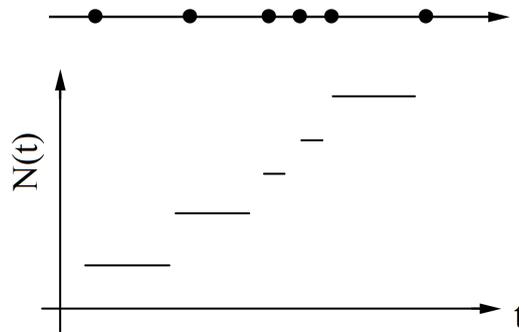
Fonte: Adaptado de Baddeley (2007)

Outro modo apresentado pelo autor, que está ilustrado na Figura 2.9, consiste em formular um processo pontual, em termos de um processo de contagem cumulativa

$$N_t = \sum_{i=1}^{\infty} \mathbb{I}_{\{T_i \leq t\}}, \quad (2.13)$$

em que  $N_t$  é o número de ocorrências até o tempo  $t$  e  $\mathbb{I}_{\{\cdot\}}$  é uma função indicadora que é igual a 1, se o argumento “ $\cdot$ ” é verdadeiro e 0, caso contrário. Essa abordagem segundo Baddeley (2007) tem a vantagem de converter o processo numa função aleatória de tempo contínuo  $t$ , mas tem a desvantagem de que os valores  $N_t$  para diferentes  $t$  são altamente dependentes.

Figura 2.9 – Processo de contagem  $N_t$  associado a um processo pontual.



Fonte: Adaptado de Baddeley (2007)

Há ainda outras formas de se analisar um processo pontual no tempo e um aprofundamento nessa teoria pode ser encontrado em Cox e Isham (1980), Daley e Vere-Jones (2003) e Baddeley (2007).

## 2.5 Processos pontuais espaço-temporais

Diferentemente de uma abordagem puramente espacial ou puramente temporal, o estudo de processos pontuais a partir de uma perspectiva espaço-temporal é importante, quando a distribuição das ocorrências tanto no espaço como no tempo são necessárias para uma melhor compreensão do fenômeno que está sendo estudado. A distribuição de uma determinada doença infecciosa ou a ocorrência de crimes em uma cidade ao longo de um período específico de tempo são alguns exemplos de situações em que a abordagem de processos pontuais espaço-temporais pode ser aplicada.

Formalmente, um processo pontual espaço-temporal é um processo estocástico que gera um conjunto contável de pontos  $\{(s_i, t_i) : i = 1, \dots, n\}$ , em que  $s_i \in S \subset \mathbb{R}^2$  e  $t_i \in T \subset \mathbb{R}^+$ . Desse

modo, o estudo dos processos pontuais espaço-temporais é considerado no produto cartesiano  $\mathbb{R}^2 \times \mathbb{R}^+$  (DIGGLE, 2013).

Um processo pontual espaço-temporal também pode ser analisado em função das propriedades de primeira e segunda ordem e, conforme afirmam Gabriel, Rowlingson e Diggle (2013), as propriedades de primeira ordem de um processo pontual espaço-temporal podem ser descritas pela sua função de intensidade espaço-temporal, dada por

$$\lambda(\mathbf{s}, t) = \lim_{|ds|, |dt| \rightarrow 0} \left\{ \frac{E[N(ds, dt)]}{|ds||dt|} \right\}, \quad (2.14)$$

em que  $ds$  define uma pequena região em torno da localização  $\mathbf{s}$ ,  $|ds|$  é a sua área,  $dt$  é um pequeno intervalo contendo o tempo  $t$ ,  $|dt|$  é o comprimento desse intervalo e  $N(ds, dt)$  é número de eventos em  $ds \times dt$ . Um processo em que  $\lambda(\mathbf{s}, t) = \lambda$  para todo  $(\mathbf{s}, t)$  é chamado e processo homogêneo. Informalmente,  $\lambda(\mathbf{s}, t)$  pode ser entendida como o número médio de eventos por unidade de volume em torno da localização  $(\mathbf{s}, t)$ .

De acordo com Gabriel, Rowlingson e Diggle (2013), as propriedades de segunda ordem de um processo pontual espaço-temporal descrevem a relação entre números de eventos em pares de sub-regiões dentro de  $S \times T$  e são dadas em termos da função de intensidade

$$\lambda_2\left((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\right) = \lim_{|D_i|, |D_j| \rightarrow 0} \left\{ \frac{E[N(D_i)N(D_j)]}{|D_i||D_j|} \right\} \quad (2.15)$$

em que  $D_i = ds_i \times dt_i$  e  $D_j = ds_j \times dt_j$  são pequenos cilindros contendo os pontos  $(\mathbf{s}_i, t_i)$  e  $(\mathbf{s}_j, t_j)$ , respectivamente.

Outros descritores essencialmente equivalentes das propriedades de segunda ordem são a densidade de covariância dada por

$$\gamma\left((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\right) = \lambda_2\left((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\right) - \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) \quad (2.16)$$

e a função distribuição radial (ou função de correlação de pares de eventos) dada por

$$g\left((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\right) = \frac{\lambda_2\left((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)\right)}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}. \quad (2.17)$$

Ainda, de acordo com Gabriel, Rowlingson e Diggle (2013), a densidade de covariância pode ser interpretada de modo análogo a covariância de variáveis aleatórias e, a função de correlação entre pares pode ser interpretada, informalmente, como a densidade de probabilidade

padronizada de ocorrer um evento em cada um de dois pequenos volumes centrados nos pontos  $(\mathbf{s}_i, t_i)$  e  $(\mathbf{s}_j, t_j)$ .

Considerando o estudo dessas propriedades, procura-se então comparar os eventos do padrão pontual observado, com o padrão esperado pela completa aleatoriedade espaço-temporal (CAET), que é o caso do processo de Poisson homogêneo. A CAET ocorre, segundo Gabriel, Rowlingson e Diggle (2013), quando

- Para algum  $\lambda > 0$ , o número  $N(S \times T)$  de eventos dentro da região  $S \times T$  segue distribuição Poisson com média  $\lambda|S||T|$ , em que  $|\cdot|$  denota a área (bidimensional) ou o comprimento (unidimensional), dependendo do contexto.
- Dado  $N(S \times T) = n$ , os  $n$  eventos em  $S \times T$  formam uma amostra aleatória independente da distribuição uniforme em  $S \times T$ .

De acordo com esses autores, para o caso do processo de Poisson homogêneo, as intensidades de primeira e de segunda ordem reduzem-se a constantes  $\lambda(\mathbf{s}, t) = \lambda$  e  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda^2$ . Desse modo, a partir de 2.16 e 2.17 temos que densidade de covariância que é dada por

$$\begin{aligned} \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \underbrace{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}_{=\lambda^2} - \underbrace{\lambda(\mathbf{s}_i, t_i)}_{=\lambda} \underbrace{\lambda(\mathbf{s}_j, t_j)}_{=\lambda} \\ &= \lambda^2 - \lambda^2 \\ &= 0, \end{aligned} \tag{2.18}$$

é identicamente zero e a função de correlação entre pares dada por

$$\begin{aligned} g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \frac{\overbrace{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}^{=\lambda^2}}{\underbrace{\lambda(\mathbf{s}_i, t_i)}_{=\lambda} \underbrace{\lambda(\mathbf{s}_j, t_j)}_{=\lambda}} \\ &= \frac{\lambda^2}{\lambda^2} \\ &= 1, \end{aligned} \tag{2.19}$$

é identicamente 1. Valores maiores ou menores do que esses parâmetros indicariam portanto, informalmente, quanto mais ou menos provável é a ocorrência de um par de eventos em localizações e tempos específicos em um processo de Poisson com a mesma intensidade.

Considerando as suposições de estacionariedade e isotropia temos, conforme Gabriel, Rowlingson e Diggle (2013), que um processo pontual espaço-temporal é estacionário de segunda ordem:

- no espaço, se  $\lambda(\mathbf{s}, t) \equiv \lambda(t)$  e  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda_2(\mathbf{s}_i - \mathbf{s}_j, t)$ , ou seja, é invariante sobre translação no espaço (intensidade de primeira ordem homogênea e intensidade de segunda ordem depende apenas da distância relativa), mas varia no tempo;
- no tempo, se  $\lambda(\mathbf{s}, t) \equiv \lambda(\mathbf{s})$  e  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda_2(\mathbf{s}, t_i - t_j)$ , ou seja, a intensidade de primeira ordem é constante no tempo, mas varia no espaço e a intensidade de segunda ordem varia no espaço, mas depende apenas da distância temporal relativa entre os eventos;
- no espaço e no tempo, se  $\lambda(\mathbf{s}, t) \equiv \lambda$  e  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda_2(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)$ , ou seja, é invariante sobre translação no tempo e no espaço.

O processo será também isotrópico se  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda_2(u, v)$ , em que  $(u, v)$  é o vetor distância espaço-temporal,  $u = \|\mathbf{s}_i - \mathbf{s}_j\|$  e  $v = |t_i - t_j|$ , ou seja, é invariante sobre rotação.

Para o caso em que o processo não é estacionário, ou seja, a intensidade de primeira ordem é não homogênea, o padrão pontual pode ser comparado a um processo de Poisson não homogêneo, que apesar de ter uma intensidade que varia, não exhibe uma estrutura de correlação espaço-temporal entre os eventos. Este processo pode ser obtido, de acordo com Gabriel, Rowlingson e Diggle (2013), substituindo a intensidade constante  $\lambda$  do processo de Poisson homogêneo, por uma função intensidade  $\lambda(\mathbf{s}, t)$  que varia no espaço, no tempo ou em ambos.

De acordo com esses autores, um processo de Poisson não homogêneo pode ser definido a partir dos seguintes postulados:

- O número  $N(S \times T)$  de eventos dentro da região  $S \times T$  segue distribuição Poisson com média  $\int_S \int_T \lambda(\mathbf{s}, t) dt ds$ .
- Dados  $N(S \times T) = n$ , os  $n$  eventos em  $S \times T$  formam uma amostra aleatória independente da distribuição em  $S \times T$  com função densidade de probabilidade proporcional a  $\lambda(\mathbf{s}, t)$ .

Gabriel, Rowlingson e Diggle (2013) afirmam que para um processo de Poisson com intensidade  $\lambda(\mathbf{s}, t)$ , a intensidade de segunda ordem é  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)$ . Assim, de modo análogo a (2.18) e (2.19), tem-se que a densidade de covariância

$$\begin{aligned}
\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) - \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) \\
&= \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) - \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) \\
&= 0,
\end{aligned} \tag{2.20}$$

é identicamente zero e a função de correlação entre pares

$$\begin{aligned}
g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \frac{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \\
&= \frac{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \\
&= 1
\end{aligned} \tag{2.21}$$

é identicamente 1.

Assim como no caso espacial, para o caso em que a intensidade de primeira ordem não é constante, a tendência espaço-temporal nas localizações dos eventos deve ser levada em consideração. O conceito de estacionariedade de segunda ordem reponderada tratado para o caso espacial em Baddeley, Møller, e Waagepetersen (2000) – abordado neste trabalho na seção 2.3 – foi estendido diretamente para o caso espaço-temporal em Gabriel e Diggle (2009). Esta suposição mais fraca de estacionariedade permite uma intensidade não constante, que é obrigada a ser limitada a partir de zero e assume que a função de correlação entre pares, dada em 2.17, depende apenas do vetor de diferenças espaço-temporais  $(u, v)$ , ou seja,

$$g(u, v) = \frac{\lambda_2(u, v)}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}. \tag{2.22}$$

Essas suposições permitem a estruturação da função K para investigar a dependência espaço-temporal do processo.

Na literatura existem diversos métodos para investigar o padrão de ocorrência de eventos que tem uma distribuição espaço-temporal. A função K, proposta por Diggle et al. (1995) para o caso homogêneo e por Gabriel e Diggle (2009) para o caso não homogêneo, são medidas de dependência espacial que utilizam as propriedades de segunda ordem do processo pontual e são considerados métodos de detecção global, ou seja, identificam agrupamentos na região como um todo e não em locais e tempos específicos. Outros tipos de análises pautadas por testes globais que verificam a existência de interação espaço-temporal (neste caso, interação se refere exclusivamente a agrupamentos), são muitas vezes utilizados quando deseja-se verificar se uma

determinada doença em estudo é ou não infecciosa. Neste trabalho são utilizados os testes Knox (1963), Mantel (1967) e Jacquez (1996). Além das análises globais, também existem as análises focais que são pautadas na detecção de agrupamentos locais. Um dos métodos mais utilizado é o de varredura da estatística *scan* proposta por Kulldorff et al. (2005). Todas essas abordagens estão detalhadas a seguir.

## 2.6 Métodos para detecção de agrupamentos espaço-temporais

### 2.6.1 Função K

Diggle et al. (1995) estenderam a ideia da função K, em (2.9) e (2.12) para o caso espaço temporal, definindo-a como

$$K(u, v) = \lambda^{-1} E[N(u, v)], \quad (2.23)$$

em que  $N(u, v)$  é o número de eventos encontrados dentro de uma distância  $u$  e tempo  $v$  e  $\lambda$  é a intensidade homogênea do processo pontual.

A função  $K$  pode ser definida diretamente da esperança como em (2.23), ou de acordo com Diggle e Gabriel (2010) a partir da função de correlação entre pares de eventos, por

$$K(u, v) = 2\pi \int_0^v \int_0^u g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) ds dt. \quad (2.24)$$

Sob a hipótese da completa aleatoriedade espaço-temporal, ou seja, quando se tem um processo de Poisson espaço-temporal homogêneo, a função K teórica pode ser deduzida como:

$$\begin{aligned} K(u, v) &= 2\pi \int_0^v \int_0^u g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) ds dt \\ &= 2\pi \int_0^v \int_0^u \frac{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} ds dt \\ &= 2\pi \int_0^v \int_0^u \underbrace{\left( \frac{\lambda^2}{\lambda^2} \right)}_{(2.19)} ds dt \end{aligned} \quad (2.25)$$

$$\begin{aligned}
&= 2\pi \int_0^v \int_0^u s ds dt \\
&= \pi u^2 \int_0^v dt \\
&= \pi u^2 v.
\end{aligned}$$

Um estimador da função  $K$  homogênea espaço-temporal, proposto por Diggle et al. (1995), com base em dados que fornecem a localização dos eventos  $\mathbf{s}_i : i = 1, \dots, n$  numa região espaço-temporal  $S \times T$ , onde  $S$  é um polígono arbitrário e  $T = [T_0, T_1]$  pode ser expresso por

$$\hat{K}(u, v) = \frac{|S \times T|}{nn_v} \sum_{i=1}^{n_v} \sum_{j>i}^{n_v} \frac{1}{w_{ij}} \mathbb{I}_{\{\|\mathbf{s}_i - \mathbf{s}_j\| \leq u; t_j - t_i \leq v\}}. \quad (2.26)$$

Na equação (2.26),  $n_v$  é o número de eventos para os quais  $t_i \leq T_1 - v$ ,  $w_{ij}$  é um fator de correção do efeito de bordas e  $\mathbb{I}_{\{\cdot\}}$  é uma função indicadora que retorna o valor 1 se o argumento “.” é verdadeiro e zero, caso contrário.

O fator de correção do efeito de bordas  $w_{ij}$  é dado por Diggle et al. (1995) como  $w_{ij} = w_{ij}^{(t)} w_{ij}^{(s)}$ . A correção do efeito de bordas espacial  $w_{ij}^{(s)}$  consiste em ponderar a função pela proporção da circunferência de um círculo centrado na posição  $\mathbf{s}_i$  com raio  $u$  encontrando-se em  $S$ . O fator de correção de borda temporal  $w_{ij}^{(t)}$  é igual a 1 se ambas as extremidades do intervalo de comprimento  $2v$  centrado em  $t_i$  estiverem dentro de  $T$  e  $1/2$  caso contrário. A correção do efeito de bordas se deve ao fato de que, caso não seja levada em consideração, o estimador da função  $K$  tem um viés, uma vez que para distâncias  $u$  e  $v$  maiores que a distância (espacial e temporal) entre a borda e o seu evento mais próximo, a contagem do número de eventos dentro de um cilindro de raio  $u$  e altura  $2v$  será menor devido à não contabilização dos eventos que estão fora da região de análise.

De acordo com Diggle et al. (1995), para qualquer processo pontual espaço-temporal estacionário, tanto a componente espacial quanto a temporal teriam uma intensidade infinita e, nesse sentido, não estão bem definidas. Além disso, o que se observa na prática é que o processo espaço-temporal ocorre em uma região grande, mas finita,  $S \times T$ . Considerando então, a restrição do processo estacionário subjacente em  $\mathbb{R}^2 \times \mathbb{R}$  a esta região finita, obtém-se processos de componentes espaciais e temporais bem definidas. Se essas componentes são independentes,

de acordo com os autores, segue-se que a função  $K$  pode ser fatorizada como

$$K(\mathbf{s}, t) = K_1(\mathbf{s})K_2(t). \quad (2.27)$$

Para o procedimento de análise com a função  $K$  homogênea, compara-se o resultado da função  $K$  estimada ( $\hat{K}(u, v)$ ) com o esperado pelo processo de Poisson homogêneo ( $\pi u^2 v$ ) através da diferença  $\hat{K}(u, v) - \pi u^2 v$ . Um resultado apresentando excesso de valores positivos significa que o número observado de eventos é maior do que o esperado pela completa aleatoriedade, indicando, portanto, a presença agrupamentos espaço-temporais em uma separação espacial e temporal inferiores a  $u$  e  $v$ , respectivamente. Já um excesso de valores negativos significa que o número de eventos observados é menor do que o esperado pelo processo de Poisson, indicando que há regularidade. Caso os valores sejam relativamente próximos de zero, não se rejeita a hipótese de que o conjunto de dados é devido a um processo de Poisson Homogêneo.

Uma estatística de teste apresentada por Diggle (2013), para testar a hipótese da completa aleatoriedade espaço-temporal é dada por

$$D = \int_0^v \int_0^u \frac{(\hat{K}(u, v) - \pi u^2 v)^2}{\pi u^2 v} dudv. \quad (2.28)$$

A estatística de teste  $D$  não apresenta distribuição amostral conhecida. Assim, utiliza-se simulações de Monte Carlo, cujo procedimento consiste em obter a distribuição empírica das estatísticas  $D$  calculadas a partir de  $N$  simulações independentes de um processo de Poisson homogêneo com intensidade constante  $\lambda$ , dada pelo número médio de eventos por unidade de área e unidade de tempo ( $\lambda = \frac{n}{|S \times T|}$ ), condicionado para ter  $n$  eventos na região de estudo. A partir do nível de significância determinado para o teste, verifica-se então o valor-p obtido como a proporção de estatísticas de teste  $D$  calculadas a partir de processos simulados iguais ou superiores à estatística observada. Deste modo o valor-p é dado por

$$\text{valor-p} = \frac{R}{(N + 1)}, \quad (2.29)$$

em que  $R$  é a classificação da estatística  $D$  observada entre as  $D_{N+1}$  estatísticas ordenadas. A partir desse procedimento, o menor valor-p possível de se obter é então  $1/(N + 1)$ .

Considerando agora o caso em que a intensidade de primeira ordem não é constante, ou seja, o processo não é estacionário, Gabriel e Diggle (2009) se utilizaram da suposição de

estacionariedade de segunda ordem reponderada para definir a função  $K$  não homogênea como sendo

$$K_I(u, v) = 2\pi \int_0^v \int_0^u g(u, v) u du dv, \quad (2.30)$$

em que  $g(u, v)$  depende apenas do vetor de diferenças espaço-temporais  $(u, v)$ , ou seja,  $g(u, v) = \lambda_2(u, v) / ((\lambda(\mathbf{s}_i, t_i))(\lambda(\mathbf{s}_j, t_j)))$ .

Sob a hipótese do processo ser Poisson não homogêneo, sabemos a partir de (2.21), que a função de correlação entre pares é igual a 1, logo a função  $K_I(u, v)$  em (2.30) se reduz a

$$\begin{aligned} K_I(u, v) &= 2\pi \int_0^v \int_0^u 1 u du dv \\ &= \pi u^2 \int_0^v dv \\ &= \pi u^2 v. \end{aligned} \quad (2.31)$$

Um estimador da função  $K_I(u, v)$  proposto por Gabriel e Diggle (2009), com base em dados que fornecem a localização dos eventos  $\mathbf{s}_i : i = 1, \dots, n$  numa região espaço-temporal  $S \times T$ , onde  $S$  é um polígono arbitrário e  $T = [T_0, T_1]$  pode ser expresso por

$$\hat{K}_I(u, v) = \frac{1}{|S \times T|} \frac{n}{n_v} \sum_{i=1}^{n_v} \sum_{j>i}^{n_v} \frac{1}{w_{ij}} \frac{1}{\lambda(\mathbf{s}_i, t_i) \lambda(\mathbf{s}_j, t_j)} \mathbb{I}_{\{\|\mathbf{s}_i - \mathbf{s}_j\| \leq u; t_j - t_i \leq v\}}. \quad (2.32)$$

em que  $n_v$  é o número de eventos para os quais  $t_i \leq T_1 - v$ ,  $w_{ij}$  é um fator de correção do efeito de bordas e  $\mathbb{I}_{\{\cdot\}}$  é uma função indicadora que retorna o valor 1 se o argumento “ $\cdot$ ” é verdadeiro e zero, caso contrário.

Estes estimadores assumem que a intensidade é conhecida, o que segundo Gabriel e Diggle (2009), não ocorre na prática. De acordo com esses autores, contorna-se o problema a partir de uma suposição de que os efeitos de primeira ordem são separáveis, ou seja, a intensidade  $\lambda(\mathbf{s}, t)$  pode ser fatorada como

$$\lambda(\mathbf{s}, t) = m(\mathbf{s}) \mu(t) \quad \forall (\mathbf{s}, t) \in S \times T, \quad (2.33)$$

em que  $m(\mathbf{s})$  e  $\mu(t)$  são funções não negativas.

Sob essa suposição de separabilidade, Gabriel e Diggle (2009) afirmam que quaisquer efeitos não separáveis são interpretados como efeitos de segunda ordem, ao invés de primeira. Ainda de acordo com os autores, estimativas adequadas de  $m(\mathbf{s})$  e  $\mu(t)$  em (2.33) dependerão

das características de cada aplicação. De um modo geral,  $\hat{m}(\mathbf{s})$  e  $\hat{\mu}(t)$  podem ser obtidos a partir de funções kernel<sup>2</sup>.

O procedimento de análise a partir da função K não homogênea é feito utilizando simulações de Monte Carlo de modo análogo à função K homogênea. A estatística de teste tem a mesma estrutura de (2.28), só que utilizando a função K não homogênea. O que difere é o fato das simulações sobre a hipótese nula serem feitas considerando agora um processo de Poisson não homogêneo com uma intensidade variável  $\hat{\lambda}(\mathbf{s}, t) = \hat{m}(\mathbf{s})\hat{\mu}(t)$ . A interpretação dos resultados é a mesma da função K homogênea.

### 2.6.2 Teste Knox

Este teste foi proposto por Knox (1963) cuja ideia é estabelecer uma contagem do número observado de pares de casos próximos no espaço e no tempo e avaliar a significância. Um número significativamente grande indicaria provas do agrupamento espaço-temporal dos pontos em estudo, rejeitando a hipótese nula dada por:

$$H_0 : \text{As distâncias espaciais entre pares de casos são independentes das distâncias temporais entre pares de casos} \quad (2.34)$$

A estatística para o teste Knox é dada por

$$T = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^S a_{ij}^T \quad (2.35)$$

em que  $a_{ij}^S$  e  $a_{ij}^T$  denotam as medidas de proximidade do tipo *hot-spot* no espaço e no tempo dadas por

$$a_{ij}^S = \begin{cases} 1, & \text{se } i \neq j \text{ e } d_{ij}^S < \delta_1 \\ 0, & \text{caso contrário} \end{cases} \quad \text{e} \quad a_{ij}^T = \begin{cases} 1, & \text{se } i \neq j \text{ e } d_{ij}^T < \delta_2 \\ 0, & \text{caso contrário} \end{cases} \quad (2.36)$$

sendo  $d_{ij}^S = \|\mathbf{s}_i - \mathbf{s}_j\|$  e  $d_{ij}^T = |t_i - t_j|$  as distâncias espacial e temporal, respectivamente.

De acordo com Tango (2010), o teste Knox é o mais amplamente utilizado para detectar agrupamentos espaço-temporais e para aplicá-lo, os valores críticos  $\delta_1$  e  $\delta_2$ , que definem a proximidade, devem ser escolhidos *a priori* com base em informações que possam de alguma

<sup>2</sup> Detalhes sobre a estimação de intensidades a partir de funções kernel estão disponíveis no Anexo A.

forma descrever algumas características do fenômeno. No caso da dengue, por exemplo, informações sobre o ciclo de vida e o alcance do mosquito *Aedes Aegypti* podem ser usadas para determinar essas distâncias críticas.

A distribuição de  $T$  sob a hipótese nula (2.34) e o valor-p podem ser obtidos a partir de um procedimento de Monte Carlo de modo análogo ao apresentado para a função  $K$ . Conforme Besag e Diggle (1977), para cada uma das  $N$  replicações, a estatística de teste é computada por uma permutação aleatória dos tempos mantendo fixas as localizações. Isto, segundo Meyer et al. (2016) elimina qualquer interação espaço-tempo sistemática nos dados, mas deixa ambas as distribuições marginais espacial e temporal inalteradas e o valor-p é calculado assim como em (2.29).

Tango (2010) mostra que o valor-p também pode ser obtido utilizando a resultados da esperança e variância de  $T$  e apresenta o resultado obtido por Barton e David (1966) para as mesmas:

$$E[T] = \frac{N_{1S}N_{1T}}{N} \quad (2.37)$$

$$\begin{aligned} Var[T] = & \frac{N_{1S}N_{1T}}{N} + \frac{4N_{2S}N_{2T}}{n(n-1)(n-2)} - \left(\frac{N_{1S}N_{1T}}{N}\right)^2 \\ & + \frac{4[N_{1S}(N_{1S}-1) - N_{2S}][N_{1T}(N_{1T}-1) - N_{2T}]}{n(n-1)(n-2)(n-3)} \end{aligned} \quad (2.38)$$

sendo

$$N = \frac{n(n-1)}{2}, N_{1S} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^S, N_{1T} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^T$$

$$N_{2S} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{j \neq k} a_{ij}^S a_{ik}^S \text{ e } N_{2T} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{j \neq k} a_{ij}^T a_{ik}^T.$$

$N_{2S}$  e  $N_{2T}$  representam o número de vezes que dois casos de pares próximos no espaço e no tempo respectivamente, são adjacentes.

Deste modo, de acordo com Tango (2010), dados os valores dos parâmetros  $\delta_1$  e  $\delta_2$  e os valores observados  $T = t$ , o valor-p pode ser aproximado por qualquer um dos seguintes casos:

- Distribuição Poisson quando  $N_{1T}$  e  $N_{1S}$  são pequenos comparados com  $N$ :

$$\text{valor-p} = 1 - \sum_{k=0}^{t-1} \frac{E[T]^k}{k!} \exp\{-E[T]\} \quad (2.39)$$

ou valor-p médio

$$\text{valor-p} = 1 - \sum_{k=0}^{t-1} \frac{E[T]^k}{k!} \exp\{-E[T]\} + \frac{1}{2} \frac{E[T]^t}{t!} \exp\{-E[T]\}; \quad (2.40)$$

- Distribuição normal com média e variância dadas respectivamente pelas equações (2.37) e (2.38):

$$\text{valor-p} = 1 - \Phi\left(\frac{t - E[T]}{\sqrt{\text{Var}[T]}}\right), \quad (2.41)$$

em que  $\Phi(\cdot)$  denota a função de distribuição normal padrão.

### 2.6.3 Teste Mantel

O teste proposto por Mantel (1967), ao invés de utilizar funções indicadoras para definir medidas de proximidade como feito para o teste Knox, utiliza funções relacionadas ao inverso das distâncias no tempo e no espaço, em que pares de pontos mais próximos tem pesos maiores que pares de pontos mais afastados. Neste teste calcula-se a soma  $T$  de todos os pares de casos, das distâncias espaciais multiplicadas pelas distâncias de tempo e depois avalia-se a sua significância. Um valor significativamente grande indicaria evidências de agrupamento espaço-temporal dos casos em estudo.

A estatística de teste é

$$T = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^S a_{ij}^T \quad (2.42)$$

em que  $a_{ij}^S$  e  $a_{ij}^T$  são as medidas de proximidade no espaço e no tempo respectivamente e são dadas por

$$a_{ij}^S = \frac{1}{d_{ij}^S + c_1} \quad (a_{ii}^S = 0) \text{ e } a_{ij}^T = \frac{1}{d_{ij}^T + c_2} \quad (a_{ii}^T = 0), \quad (2.43)$$

sendo  $d_{ij}^T$  e  $d_{ij}^S$  são as distâncias entre dois pontos no tempo e no espaço respectivamente;  $c_1$  e  $c_2$  são parâmetros desconhecidos e devem ser predefinidos pelo usuário. Mantel (1967) sugere que uma constante deve ser adicionada às distâncias devido à possibilidade de distâncias espaciais e/ou temporais muito pequenas ou iguais a zero.

Apesar de não ser necessário estabelecer distâncias críticas subjetivas como o teste Knox, o teste Mantel também necessita de estabelecer subjetivamente as constantes  $c_1$  e  $c_2$ . Mantel (1967) sugere que essas constantes devem ser proporcionais às distâncias possíveis previstas no tempo e no espaço. Como exemplo, o autor cita um caso em que os eventos acontecem

dentro de separações espacial e temporal de aproximadamente 1 quilômetro e 60 dias respectivamente, e a utilização de  $c_1 = 0,5$  km e  $c_2 = 30$  dias parece razoável.

A distribuição de  $T$  sob a hipótese nula (idem à (2.34)) e o valor-p podem ser obtidos através de permutações em um procedimento de Monte Carlo como já mostrado anteriormente.

#### 2.6.4 Teste Jacquez

O objetivo do teste proposto por Jacquez (1996) é contar o número observado de pares de casos próximos no espaço e no tempo, em que as medidas de proximidade não são dadas pelas distâncias euclidianas, mas sim considerando as relações entre os  $k$  vizinhos mais próximos e avaliar a sua significância. A distância euclidiana entre os eventos serve para verificar a separação (no tempo e no espaço) em que os eventos ocorreram, para computar qual ou quais são os  $k$  vizinhos mais próximos, mas não é em si uma medida de proximidade.

Dado um ponto  $(s_i, t_i)$ , o ponto  $(s_j, t_j)$  será o seu primeiro vizinho mais próximo se a suas separações tanto no tempo como no espaço forem as menores entre todos os pontos do processo. Um ponto  $(s_p, t_p)$  será o segundo vizinho mais próximo de  $(s_i, t_i)$ , se as suas separações tanto no tempo como no espaço forem as segundas menores e assim sucessivamente para outros valores de  $k$ .

Um número significativamente grande de vizinhos próximos no tempo e no espaço indicaria evidências de agrupamentos espaço-temporais dos casos em estudo.

A estatística de teste é dada por

$$T_k = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^S a_{ij}^T \quad (2.44)$$

em que  $a_{ij}^S$  e  $a_{ij}^T$  são as medidas de proximidade no espaço e no tempo respectivamente e são dadas por

$$a_{ij}^S = \begin{cases} 1, & \text{se o evento } j \text{ é um } k \text{ vizinho mais próximo do evento } i (\neq j) \text{ no espaço} \\ 0, & \text{caso contrário} \end{cases} \quad (2.45)$$

$$a_{ij}^T = \begin{cases} 1, & \text{se o evento } j \text{ é um } k \text{ vizinho mais próximo do evento } i (\neq j) \text{ no tempo} \\ 0, & \text{caso contrário} \end{cases} \quad (2.46)$$

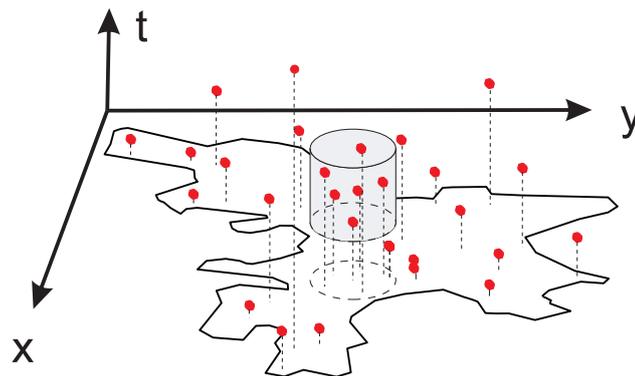
Dados os valores de  $k$ , a análise pelo método de Monte Carlo e a interpretação dos resultados se procede do mesmo modo como nos métodos mencionados anteriormente.

### 2.6.5 Estatística *scan* - modelo de permutação no espaço-tempo

O modelo de permutação no espaço-tempo proposto por Kulldorff et al. (2005) é um método de localização de agregados espaço-temporais que, diferentemente dos testes globais apresentados anteriormente, encontra os agrupamentos em sub-regiões geográficas e intervalos de tempos específicos dentro de toda a região no período de estudo.

Essa estatística de varredura espaço-temporal utiliza milhares (ou milhões) de cilindros sobrepostos, cujo raio e altura variam de tamanho, para definir a janela de varredura, sendo cada cilindro um possível candidato a um agrupamento.

Figura 2.10 – Situação hipotética de um cilindro localizado sobre um possível agrupamento. Os pontos representam casos e a sua altura em relação ao plano  $xy$  é o tempo decorrido durante o período de estudo.



De acordo com Kulldorff et al. (2005), a base circular representa a área geográfica do agrupamento potencial, conforme o exemplo da Figura 2.10. A altura do cilindro representa o intervalo de tempo, neste caso sendo o número de dias. Para cada centro e raio da base circular do cilindro, o método avalia em todas as alturas temporais possíveis.

Esse método não requer dados da população em risco, assim, o número de casos esperado é calculado usando apenas os casos observados. Esse método pode ser utilizado tanto para dados agregados por áreas como para coordenadas únicas, que é o caso deste trabalho.

Seja  $c_{zd}$  o número observado de casos em uma área  $z$  durante o dia  $d$ . O número total de casos observados  $C$  é

$$C = \sum_z \sum_d c_{zd}. \quad (2.47)$$

A área  $z$  aqui é entendida como uma pequena região que compreende apenas uma ocorrência cujas coordenadas correspondem ao centroide da área. Para cada  $z$  e  $d$ , calcula-se o número esperado de casos  $\mu_{zd}$  condicionando nas marginais observadas:

$$\mu_{zd} = \frac{1}{C} \left( \sum_z c_{zd} \right) \left( \sum_d c_{zd} \right). \quad (2.48)$$

Segundo Kulldorff et al. (2005) esta é a proporção de todos os casos que ocorreram em  $z$  multiplicado pelo número total de casos durante o dia  $d$ . O número esperado de casos  $\mu_A$  em um determinado cilindro  $A$  é o somatório dessas esperanças sobre todas as localizações dentro desse cilindro:

$$\mu_A = \sum_{(z,d) \in A} \mu_{zd}. \quad (2.49)$$

De acordo com os autores, a hipótese subjacente ao calcular esses números esperados é que a probabilidade de um caso estar em  $z$ , dado que foi observada no dia  $d$ , é a mesma para todos os dias  $d$ .

Seja  $c_A$  o número observado de casos no cilindro. Condicionado às marginais, e quando não há interação espaço-tempo, Kulldorff et al. (2005) afirmam que  $c_A$  segue distribuição hipergeométrica com média  $\mu_A$  e função de probabilidade dada por

$$P(C_A) = \frac{\binom{\sum_{z \in A} c_{zd}}{c_A} \binom{C - \sum_{z \in A} c_{zd}}{\sum_{d \in A} c_{zd} - c_A}}{\binom{C}{\sum_{d \in A} c_{zd}}}. \quad (2.50)$$

De acordo com os autores, se  $\sum_{z \in A} c_{zd}$  e  $\sum_{d \in A} c_{zd}$  são pequenos em comparação com  $C$ ,  $c_A$  segue uma distribuição aproximadamente Poisson com média  $\mu_A$ . Com base nessa aproximação, se utiliza a razão de verossimilhança generalizada de Poisson  $k$ , como medida da evidência de que o cilindro  $A$  contém um surto:

$$k(A) = \left( \frac{c_A}{\mu_A} \right)^{c_A} \left( \frac{C - c_A}{C - \mu_A} \right)^{C - c_A}. \quad (2.51)$$

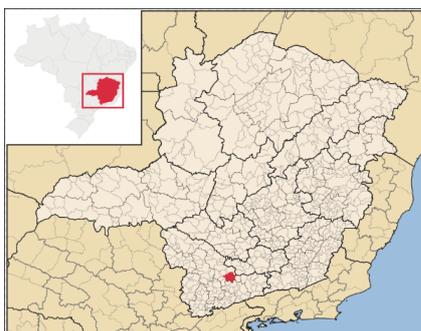
Entre os muitos cilindros avaliados, aquele com o máximo de  $k(A)$  constitui o aglomerado espaço-tempo de casos que é menos provável de ser uma ocorrência completamente aleatória e, portanto, é o candidato para um verdadeiro agrupamento de eventos. Para verificar se esse cilindro exibe um agrupamento espaço-temporal significativo, o método de análise é feito utilizando o procedimento de Monte Carlo abordado anteriormente.

### 3 MATERIAL E MÉTODOS

#### 3.1 Área de estudo

A cidade de Três Corações situa-se na região sul do estado de Minas Gerais na latitude:  $21^{\circ} 41' 49'' S$  e longitude:  $45^{\circ} 15' 12'' W$  a uma altitude de 859 metros acima do nível do mar e está localizada a 293 km ao sul de Belo Horizonte.

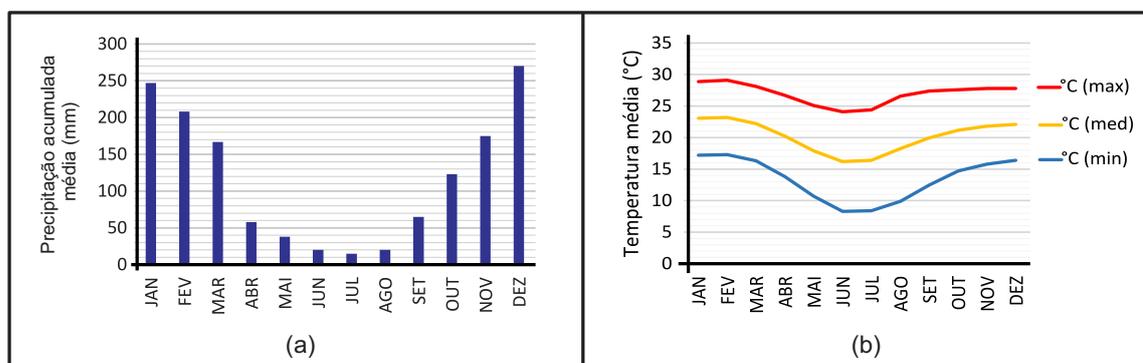
Figura 3.1 – Município de Três Corações destacado em vermelho no sul do estado de Minas Gerais.



Fonte: [https://pt.wikipedia.org/wiki/Três\\_Corações](https://pt.wikipedia.org/wiki/Três_Corações). Acesso em: 06 dez. 2016.

De acordo com o Instituto Brasileiro de Geografia e Estatística (2016), foi registrado no censo de 2010 uma população de 72.765 habitantes e a projeção para o ano de 2016 é de 78.474 habitantes. A área do município é de  $828,038 \text{ km}^2$  e de acordo com o sítio CLIMATE-DATA.ORG,<sup>1</sup> Três Corações possui clima temperado úmido com inverno seco e verão quente pela classificação de Köppen e Geiger. A pluviosidade média anual do município é de 1401 mm e a temperatura média é de  $20,2^{\circ}\text{C}$ . Na Figura 3.2 está apresentado um resumo mensal da temperatura e do acumulado médio de precipitação:

Figura 3.2 – Gráfico da precipitação acumulada média mensal (a) e das temperaturas mínima, máxima e média mensal (b) para a cidade de Três Corações.



Fonte: Adaptado de CLIMATE-DATA.ORG

<sup>1</sup> Disponível em: <<http://pt.climate-data.org/location/25070/>>. Acesso em: 06 dez. 2016.

Conforme a Figura 3.2, o período com maior volume de chuvas ocorre de outubro a março, sendo dezembro o mês com o maior acumulado médio de precipitação (270 mm). Já o mês com menor precipitação é julho com 15 mm. De acordo o sítio CLIMATE-DATA.ORG, fevereiro é o mês mais quente do ano, com média de 23,2 °C e com uma temperatura média de 16,2 °C, junho é o mês mais frio. A diferença de precipitação entre o mês mais seco e o mais chuvoso é de 255 mm, e as temperaturas médias variam em 7 °C.

### 3.2 O conjunto de dados

O conjunto de dados analisados nesta pesquisa foi obtido junto à vigilância epidemiológica da cidade de Três Corações, em que a localização dos eventos (ocorrência dos casos de dengue) foi composta pelo endereço residencial de cada uma das pessoas diagnosticadas e a data da ocorrência de cada evento corresponde ao momento da notificação no serviço de saúde.

Para a preparação dos dados para a análise, a primeira etapa consistiu na transformação dos endereços residenciais em coordenadas geográficas. As ferramentas utilizadas neste procedimento foram a base de dados do Google Earth (GOOGLE INC, 2016) e o pacote Rgoogle-Maps (LOECHER; ROPKINS, 2015) do *software* R (R CORE TEAM, 2016). As coordenadas dos endereços obtidas nesta etapa são dadas na forma geográfica decimal, que são coordenadas na esfera. Contudo, as distâncias calculadas no *software* R são euclidianas, sendo necessário que seja feita uma transformação nos dados para coordenadas UTM (*Universal Transversa de Mercator*), que são coordenadas no plano bidimensional. Nesta etapa utilizou-se o pacote sp (PEBESMA; BIVAND, 2005).

O arquivo de dados continha um total de 3456 ocorrências, sendo que no período de 2010 a 2015, os dados variaram ano a ano, na seguinte sequência: 381, 471, 55, 58, 1056 e 1435. Contudo, para o trabalho foram desconsiderados os casos que ocorreram fora da área urbana e os casos em que as informações de endereço não eram suficientes ou estavam incorretas, impossibilitando a localização das coordenadas. Desse modo, o conjunto de dados analisado no trabalho, passou a ter, respectivamente, ao período de 2010 a 2015, quantidades de casos iguais a 304, 397, 51, 50, 854 e 1162, totalizando 2818 ocorrências de dengue.

Considerando a dimensão temporal do processo pontual, foram admitidas as datas de notificação das ocorrências de dengue, contudo, o horário da notificação não está disponível. Devido a limitações computacionais na execução dos testes no R foi considerado que uma pessoa infectada com dengue pode procurar o serviço de saúde em qualquer hora do dia. Assim,

utilizando-se de programação foi gerado para cada ocorrência um valor aleatório com probabilidade igual para todos os horários de 0h às 24h, a partir de uma distribuição uniforme  $[0, 24]$ . Dividindo o resultado deste procedimento por 24, o valor final irá representar o momento da notificação como a proporção de tempo decorrida naquele dia a partir das 0h. Por exemplo, o valor 0,26 representa aproximadamente 6h14min. O tempo para todo o intervalo será contado em dias decorridos, então um valor como, por exemplo, 98,26 indica que a ocorrência foi as 6h14min do dia 98 do período em estudo.

Para o caso da estatística de varredura com o modelo de permutação espaço-temporal este procedimento não é necessário, devendo ser fornecidas apenas as informações das coordenadas e as datas.

O contorno de Três Corações (necessário para as análises com a função K) foi construído com o auxílio dos *softwares* Google Earth, QGis (QGIS DEVELOPMENT TEAM, 2016) e o pacote *sp* (PEBESMA; BIVAND, 2005) do R. Primeiramente, foi construído um polígono ao redor da cidade conforme a Figura 3.3.

Figura 3.3 – Contorno da cidade de Três Corações obtido através do *Software* Google Earth.



As coordenadas dos vértices do polígono foram extraídas em forma de latitude e longitude e convertidas para UTM, conforme o procedimento já citado, com o pacote *sp* (PEBESMA;

BIVAND, 2005) do R. Utilizando o *software* QGIS (QGIS DEVELOPMENT TEAM, 2016) foi construído um arquivo *shape* a partir das coordenadas UTM obtidas no passo anterior.

### 3.3 Métodos estatísticos

Os *softwares* necessários para as análises com os métodos descritos anteriormente na seção 2.6 foram o R e o SaTScan<sup>TM</sup> (KULLDORF, 2010).

A análise descritiva dos dados e a estimação da intensidade puramente espacial foram feitas utilizando o pacote *spatstat* (BADDELEY; TURNER, 2005) do R. Os testes Knox, Mantel e Jacquez foram analisados no R, cuja implementação dos mesmos foi feita conforme indicação de Tango (2010)<sup>2</sup>.

A distribuição empírica das estatísticas para esses testes sob a hipótese nula e o valor p foram obtidas a partir do procedimento de permutação de Monte Carlo considerando  $N = 999$  repetições e o nível de significância considerado neste trabalho foi de 5%.

Para o teste Knox, as distâncias críticas utilizadas foram  $\delta_1 = 0,2\text{km}$  e  $\delta_2 = 30$  dias. Esses valores foram escolhidos a partir da informação apresentada na seção 2.1, sobre o alcance do voo e o tempo de vida adulta do mosquito *Aedes Aegypti*. Para o teste de Mantel foi utilizado a ideia proposta pelo autor de que as constantes aditivas devem ser proporcionais com as distâncias possíveis previstas no tempo ou no espaço entre os casos relacionados. Assim, a partir das informações de 0,2 km e 30 dias foram utilizadas, de modo proporcional, os valores  $c_1 = 0,1$  e  $c_2 = 15$  dias. Para o teste Jacquez, o procedimento de análise foi considerado para  $k = 1, \dots, 20$  vizinhos mais próximos.

As análises com a função K foram realizadas no R utilizando os pacotes *splancks* (ROWLINGSON; DIGGLE, 1993) e *stpp* (GABRIEL; ROWLINGSON; DIGGLE, 2013) utilizando a função K homogênea e a K não homogênea. Uma comparação dos resultados das duas funções permite verificar se um possível agrupamento espaço-temporal na região é devido a uma tendência (efeito de primeira ordem) ou se existe uma estrutura de correlação espaço-temporal, ou seja, dependência (efeitos de segunda ordem).

A análise utilizando a estatística de varredura com o modelo de permutação espaço-temporal, foi realizada no *software* SaTScan<sup>TM</sup>. A primeira etapa da análise consiste na verificação de todos os possíveis cilindros espaço-temporais que podem representar agrupamentos

<sup>2</sup> Os códigos do R para esses testes estão disponíveis na página da internet do *Department of Technology Assesment and Biostatistics* do *National Institute of Public Health - Japão*, disponível em: <<http://www.niph.go.jp/soshiki/gijutsu/download/Rfunctions/index.html>>.

significativos. Na sequência, através do procedimento de Monte Carlo é obtido o valor-p para cada um dos cilindros e aqueles que foram significativos são detalhados na análise.

Todas as análises descritas aqui foram realizadas nos computadores do Laboratório de Computação Científica, do Departamento de Ciências Exatas, da Universidade Federal de Lavras (UFLA).

## 4 RESULTADOS E DISCUSSÃO

Em relação a distribuição temporal dos dados das ocorrências da dengue, como já mostrado na seção 3.2, as quantidades de casos em cada ano do período de estudo apresentaram uma grande variação. De 2010 a 2015, as quantidades para cada ano foram, respectivamente, 304, 397, 51, 50, 854 e 1162, totalizando 2818 ocorrências de dengue.

Fazendo um refinamento dos dados dessas ocorrências por mês, pode-se observar a partir da Figura 4.1, um padrão de sazonalidade no período em estudo, em que os intervalos de maior intensidade nas ocorrências de dengue ocorrem entre os meses de março a maio, com destaque para o mês de abril, onde acontece o pico das ocorrências.

No trabalho de Viana e Ignotti (2013), os autores fizeram uma revisão de literatura a respeito da ocorrência da dengue e as variações meteorológicas no Brasil e constataram que, nas diferentes localidades o período de maior incidência da dengue ocorre em geral nos períodos de maior intensidade das chuvas e com temperaturas mais altas. Já nos períodos mais frios e secos esse número diminui. Em Três Corações, apesar de nos períodos mais frios (junho a setembro) o número de ocorrências ter sido pequeno em comparação com os demais, o pico das ocorrências em abril e o baixo número de casos nos meses mais chuvosos, que vai, conforme a Figura 3.2, de outubro a março, contrasta com o resultado desses autores.

Uma possível justificativa para a situação observada em Três Corações pode se dar a partir de uma mudança no padrão das chuvas no período de estudo. De acordo com informações do *site*<sup>1</sup> da Agência FAPESP, em pesquisas do Instituto Nacional de Pesquisas Espaciais (Inpe) constatou-se que desde o ano 2010 a quantidade de chuvas vem diminuindo no sudeste do Brasil, cujo agravamento ocorreu nos anos de 2014 e 2015, que coincidem justamente com os períodos de maior número de ocorrências da dengue na cidade de Três Corações. Uma hipótese possível é que a menor quantidade de chuvas interferiu no ciclo de vida do mosquito transmissor, tornando mais tardia a sua proliferação. Contudo, para verificar essa hipótese ou outras possíveis causas das ocorrências de dengue é necessário um estudo mais aprofundado, que leve em consideração outras variáveis, o que não foi objetivo deste trabalho.

<sup>1</sup> Disponível em: <[http://agencia.fapesp.br/deficit\\_de\\_chuvas\\_no\\_brasil\\_vem\\_aumentando\\_nas\\_ultimas\\_decadas/21532/](http://agencia.fapesp.br/deficit_de_chuvas_no_brasil_vem_aumentando_nas_ultimas_decadas/21532/)>. Acesso em: 13 dez. 2016.

Figura 4.1 – Distribuição mensal das ocorrências de Dengue em Três Corações no período de janeiro de 2010 a dezembro de 2015.

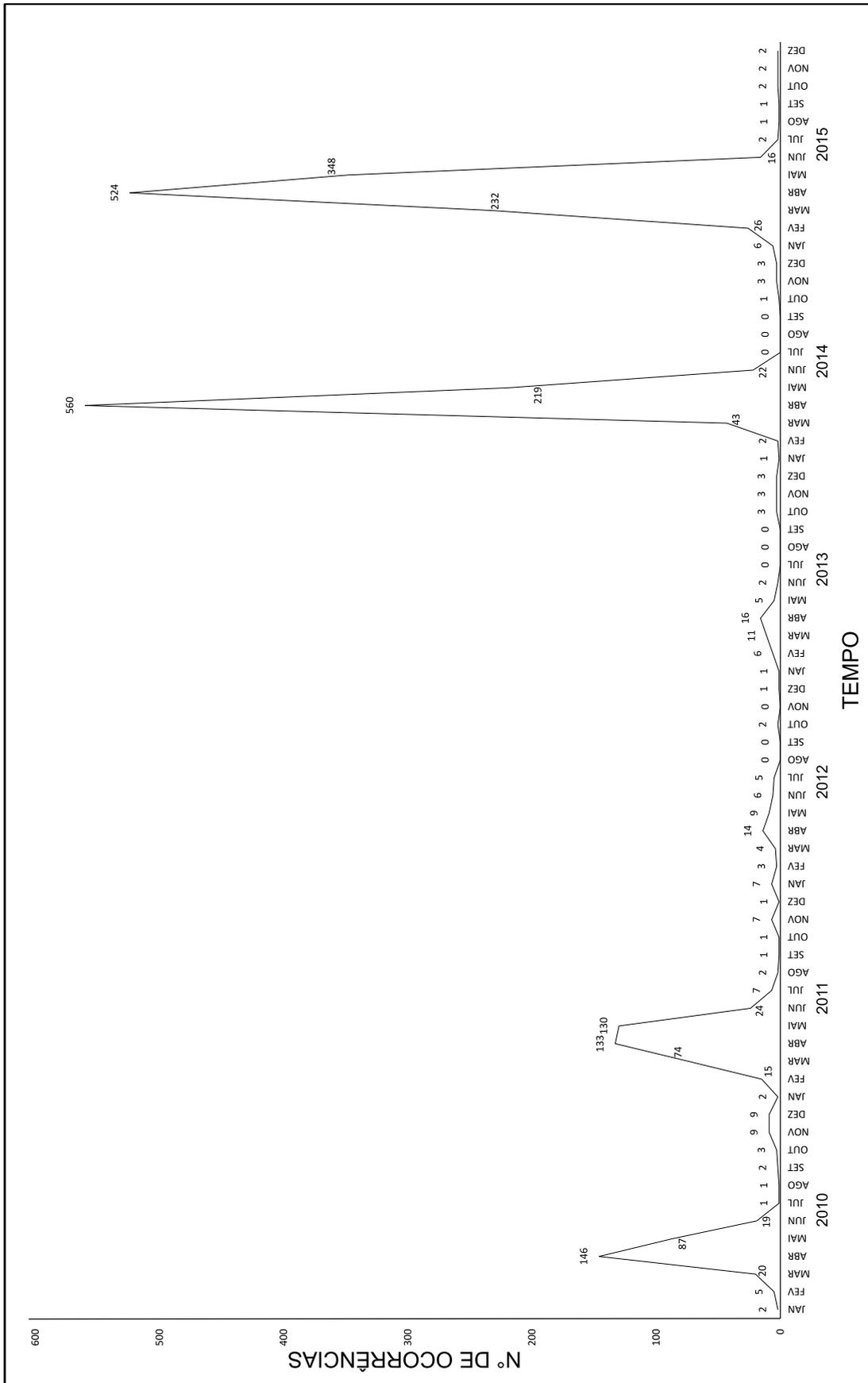
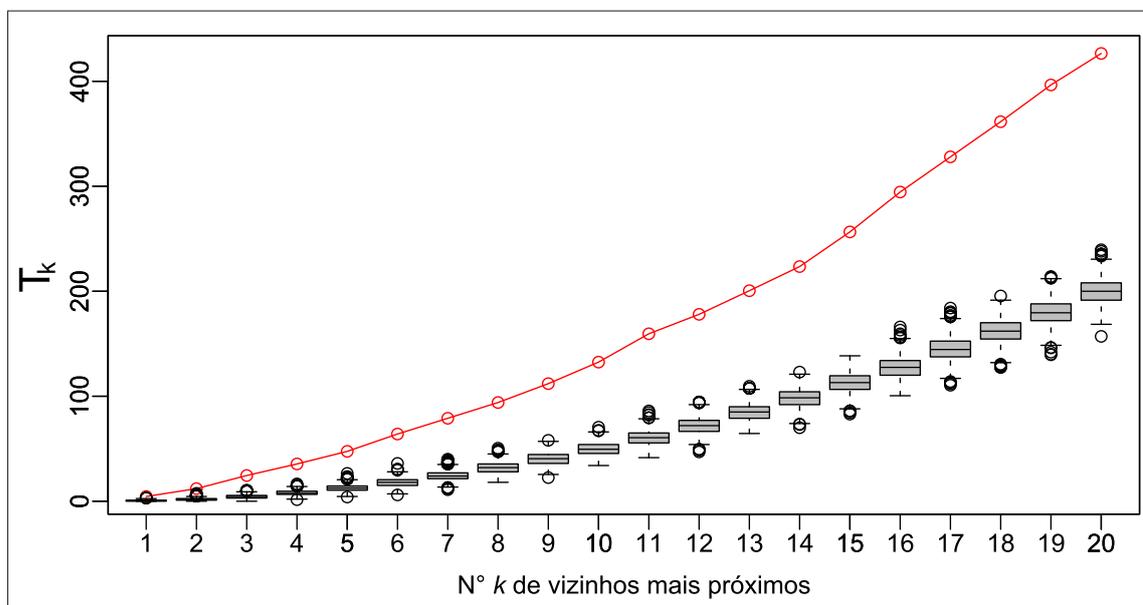




Figura 4.3 – Boxplots das estatísticas  $T_k$  obtidas no teste Jacquez para  $k = 1, \dots, 20$  vizinhos mais próximos a partir de um procedimento de Monte Carlo com  $N = 999$  repetições, e em vermelho a localização da estatística calculada para os dados originais.



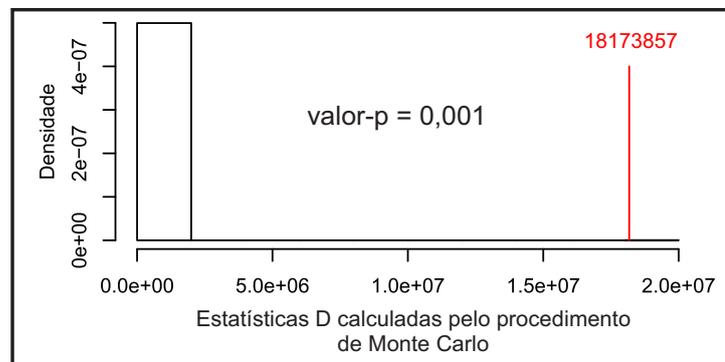
Para todos os três testes os valores-p foram iguais a 0,001, e portanto, a hipótese nula de que as componentes espacial e temporal atuam de modo independente uma da outra foi rejeitada, considerando o nível de significância de 5%. Isso indica então, que durante o período de 01/01/2010 a 31/12/2015 em Três Corações as ocorrências de dengue exibiram padrão de agrupamento espaço-temporal.

Em relação ao poder desses testes, Jacquez (1996) mostrou que o teste proposto por ele apresenta melhor performance que os demais, sendo seguido pelo teste de Mantel e depois o teste Knox. A respeito dos testes Knox e Mantel, Meyer et al. (2016) afirmam que os mesmos têm como vantagem o fato de trabalharem apenas com as informações dos casos e dos tempos associados, sem a necessidade de informações adicionais. Contudo, eles não levam em conta possíveis variações no espaço ou no tempo, como a densidade populacional ou efeitos sazonais e, portanto são propensos a viés. Uma vantagem do teste Jacquez em relação aos testes Knox e Mantel é que ele não necessita de medidas críticas ou adição de constantes. No entanto, Tango (2010) afirma que um problema sobre a abordagem dos  $k$  vizinhos mais próximos é que esse método atribui um peso igual a todos os vizinhos mais próximos, apesar de existirem alguns casos situados distantes um do outro, o que pode causar um aparente agrupamento significativo. Outro fator a ser destacado é que as análises utilizando o teste Jacquez são computacionalmente

mais trabalhosas e necessitam de um tempo relativamente maior para se obter os resultados se comparado aos testes Knox e Mantel.

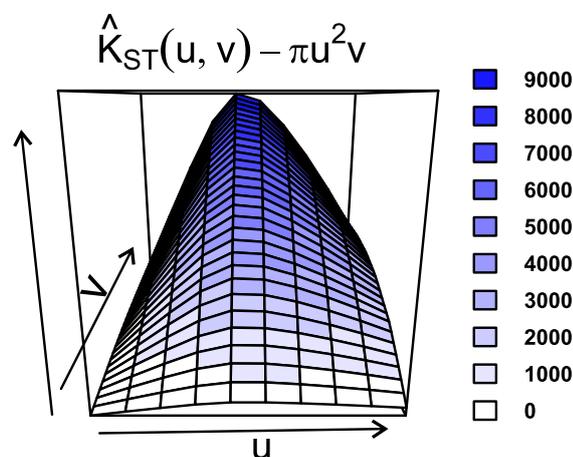
Na Figura 4.4 está apresentado o resultado da análise utilizando a estatística de teste D baseada na função K homogênea. A partir do procedimento de Monte Carlo, o valor-p obtido foi igual a 0,001, portanto, foi rejeitada ao nível de 5% de significância, a hipótese de que os casos de dengue apresentam uma configuração completamente aleatória no tempo e no espaço.

Figura 4.4 – Histograma da distribuição empírica das estatísticas calculadas D a partir de um procedimento de Monte Carlo com  $N = 999$  repetições utilizando a função K homogênea e a localização da estatística observada.



Ao executar a função K homogênea e compará-la ao que seria esperado pela completa aleatoriedade espaço-temporal, o gráfico dessa análise, exposto na Figura 4.5 mostra que houve um excesso de valores positivos, indicando assim, que o comportamento das ocorrências de dengue em Três Corações exibiu um padrão de dependência, com agrupamentos de casos no espaço-tempo.

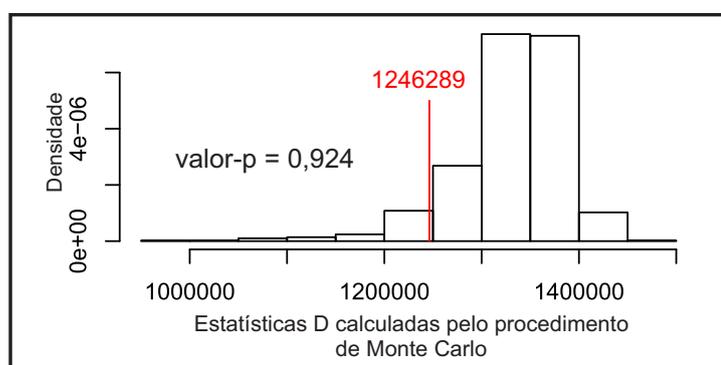
Figura 4.5 – Gráfico da diferença entre a função K homogênea observada e o esperado pelo processo de Poisson homogêneo.



A análise utilizando a função K não homogênea apresentou resultado diferente do exposto na Figura 4.4.

Na Figura 4.6 está apresentado o resultado da análise utilizando a estatística de teste D baseada na função K não homogênea. O valor-p obtido pelo procedimento de Monte Carlo foi igual a 0,924, portanto, não foi rejeitada ao nível de 5% de significância, a hipótese de que os casos de dengue apresentam uma configuração completamente aleatória no tempo e no espaço.

Figura 4.6 – Histograma da distribuição empírica das estatísticas calculadas D a partir de um procedimento de Monte Carlo com  $N = 999$  repetições utilizando a função K não homogênea e a localização da estatística observada.

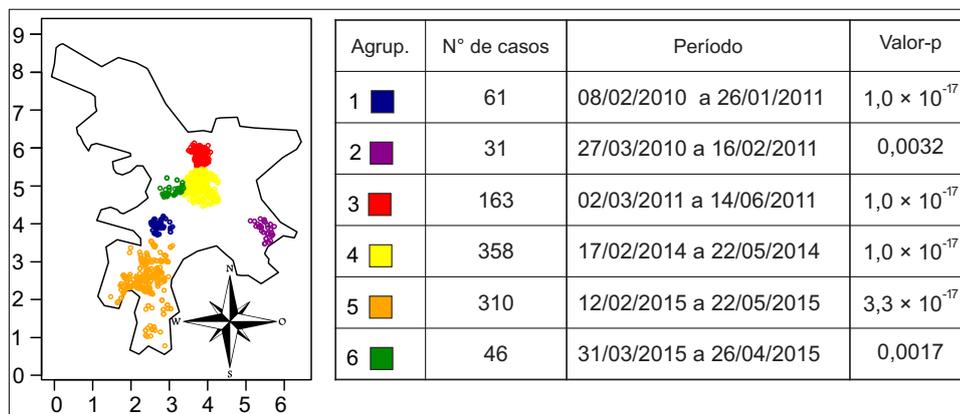


Isso indica então, que as ocorrências de casos de dengue em Três Corações no período de estudo estavam distribuídas de acordo com um processo de Poisson não homogêneo. Assim, os resultados obtidos pela função K homogênea, que indicaram um padrão de agrupamento espaço-temporal devido a um efeito de dependência espacial, na verdade foi devido aos efeitos de primeira ordem (tendência), ou seja, as intensidades espacial (provavelmente devido a densidade populacional) e temporal (provavelmente devido aos períodos de chuva) variaram na região dentro do período em estudo.

Uma vantagem na utilização da função K em relação aos testes Knox, Mantel e Jacquez, é que a função K permite verificar se um agrupamento espaço-temporal é devido a um efeito de primeira ordem (tendência) ou de segunda ordem (dependência). Não foram encontrados trabalhos comparando o poder dos testes baseados na função K com os testes de Knox, Mantel e Jacquez. Neste trabalho não foi executado um procedimento para verificar e comparar o poder, ou a taxa de erro tipo I desses testes, em função do extensivo tempo necessário para os procedimentos computacionais.

A última análise, utilizando a estatística Scan, detectou seis agrupamentos locais estatisticamente significativos, cujo resumo se encontra na Figura 4.7:

Figura 4.7 – Resumo dos resultados obtidos pela análise do métodos de permutação espaço-temporal com o *software* SaTScan™ (unidades do mapa dadas em quilômetros).



Os bairros que possuem coordenadas situadas em cada um dos agrupamentos encontrados são respectivamente:

- 1 – São Sebastião, Cotia, Jardim Rio Verde, Repartição da Rede, Vila Tereza e Vila João Pinto;
- 2 – Vila Emílio, Parque Jussara e São Gerônimo;
- 3 – Vilas Boas, Rio do Peixe, Vila Lima e Jardim Primavera;
- 4 – Centro, Jardim Gonzaga, Rio do Peixe, Vila Amorelli, Bela Vista, Vila Jessé, Vila Mariana, Monte Alegre, Jardim Universo, Boa Ventura e Vila São Domingos;
- 5 – Alto da Boa Vista, Vila Fernão Dias, Monte Verde, Vila Izabel, Santo Afonso, Vila Faria, Barro Branco, Jardim Fabiana, Espriado, Jardim Califórnia, Vila Emília e Parque das Colinas;
- 6 – Triângulo, Chácara das Rosas, Vila Amorelli, Centro e Monte Alegre

Pode-se perceber que os focos dos agrupamentos estão em locais distintos e com intervalos de tempo com diferentes comprimentos. Os agrupamentos 1 e 2, com 61 e 31 casos respectivamente foram detectados em intervalos de quase um ano, indicando que na região onde os casos ocorreram, havia um foco em que as infecções pela dengue persistiram durante todo o período em questão.

Nos anos de 2012 e 2013, em que a quantidade de casos foi relativamente pequena se comparada com os outros anos, não foi detectado nenhum agrupamento espaço-temporal. Os dois maiores agrupamentos – 4 e 5 – foram registrados em intervalos de tempo semelhantes,

ocorrendo no período de fevereiro a maio nos anos de 2014 e 2015, respectivamente. Este período, como se pode ver na Figura 4.1 é onde está concentrado a grande maioria dos casos nesses anos. Diferentemente dos agrupamentos 1 e 2, os casos de dengue dos agrupamentos 3, 4, 5 e 6 ocorreram em um curto intervalo de tempo, caracterizando assim, um surto de dengue para essas regiões naquele período.

A partir dessas análises pode-se notar também, assim como já discutido anteriormente com os mapas de kernel, que a maioria das ocorrências dos agrupamentos ocorreram em localizações nas proximidades da região central da cidade.

As análises utilizando a estatística scan de varredura se diferenciam dos testes anteriores pelo fato de conseguir detectar a localização e o tamanho dos agrupamentos e testar a significância de cada um deles. Assim, de acordo com Kulldorff (2015), como a estatística de varredura e os testes servem a diferentes propósitos, eles não se excluem mas sim, se complementam.

## 5 CONCLUSÕES

Este trabalho mostrou a importância de se utilizar uma combinação de diversos métodos para que se possa realizar uma análise espaço-temporal efetiva dos mecanismos envolvidos nas doenças infecciosas como a dengue.

Foram expostos cinco testes para detecção global de agrupamentos (Knox, Mantel, Jacquez, Função K homogênea e Função K não homogênea) e um método de detecção local de agrupamentos no espaço-tempo.

Através das análises desenvolvidas neste trabalho foi possível verificar a partir dos testes Knox, Mantel, Jacquez e a estatística de teste D baseada na Função K homogênea, que o padrão das ocorrências da dengue na cidade de Três Corações durante o período de 01/01/2010 a 31/12/2015 exibiram um padrão global de agrupamento espaço-temporal. Esse resultado, no entanto, através da análise com a estatística de teste D baseada na Função K não homogênea, se mostrou não significativo quanto a presença de agrupamento espaço-temporal, indicando que o resultado obtido pelos testes anteriores não foi devido a uma estrutura de correlação espaço-temporal entre as ocorrências, mas sim devido a variações no valor médio do processo pontual. Ou seja, as ocorrências estavam distribuídas no espaço-tempo de acordo com um processo de Poisson não homogêneo e os agrupamentos ocorreram em função de um efeito de tendência espaço-temporal, como por exemplo variações na densidade populacional e de precipitação de chuvas dentro da região de estudo.

O método baseado na estatística Scan encontrou seis agrupamentos significativos com diferentes áreas de abrangência e períodos de tempo. É importante salientar que apesar das análises detectarem agrupamentos espaço-temporais, esta foi uma análise retrospectiva, indicando apenas um possível padrão de comportamento espaço-temporal da dengue. Esses resultados não podem ser utilizados para prever o número de casos em locais e/ou tempos futuros. Esses resultados podem ser utilizados como uma ferramenta para que os agentes da saúde/vigilância epidemiológica possam investigar o que existe nessas localidades e períodos de tempos que poderia potencializar a proliferação do mosquito transmissor e traçar estratégias que auxiliem na prevenção e combate à dengue.

## REFERÊNCIAS

- BADDELEY, A. Spatial Point Processes and their Applications. In: BADDELEY, A.; BÁRÁNY, I.; SCHNEIDER, R. Spatial point processes and their applications. **Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004**, p. 1-75, 2007.
- BADDELEY, A. J.; MØLLER J.; WAAGEPETERSEN R. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. **Statistica Neerlandica**, v. 54, n.3, p.329-350, 2000.
- BADDELEY, A; TURNER, R. spatstat: An R Package for Analyzing Spatial Point Patterns. **Journal of Statistical Software**, [S.l.], v. 12, Issue 6, p. 1 - 42, jan. 2005. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v012i06>>. Acesso em: 08 mar. 2016.
- BARTON D.E.; DAVID F.N. The random intersection of two graphs. In: David F.N. (Eds.), **Research Papers in Statistics**, 45-59. New York: Wiley, 1966.
- BESAG, J; DIGGLE, P. J. Simple Monte Carlo tests for spatial pattern. **Applied statistics**, p. 327-333, 1977.
- BIVAND, R.S.; PEBESMA, E.; RUBIO, V.G. **Applied Spatial Data Analysis with R**. New York: Springer, 2008.
- BONITA, R.; BEAGLEHOLE, R.; KJELLSTRÖM, T. Basic epidemiology. 2. ed.: **World Health Organization**, 2006. Disponível em: <[http://apps.who.int/iris/bitstream/10665/43541/1/9241547073\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/43541/1/9241547073_eng.pdf)>. Acesso em: 10 fev. 2016.
- BRASIL. Resolução RE Nº 3.519, de 23 de dezembro de 2015. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 28 dez. 2015. Suplemento - Seção 1, p. 4.
- BRASIL, Ministério da Saúde. **Situação Epidemiológica / Dados**. Disponível em: <<http://portalsaude.saude.gov.br/index.php/situacao-epidemiologica-dados-dengue>>. Acesso em: 07 mar. 2016.
- BRILLINGER, D. R., GUTTORP, P. M.; SCHOENBERG, F. P. 2006. Point Processes, Temporal. In: EL-SHAARAWI, A. H.; PIEGORSCH, W. W. **Encyclopedia of environmetrics**. John Wiley & Sons, 2001.
- BUTANTAN. **Dengue: desafios para uma vacina**. Disponível em: <<http://www.butantan.gov.br/butantan/dengue/Paginas/default.aspx>>. Acesso em: 16 nov. 2016.
- CÂMARA, G.; CARVALHO, M. S. Análise espacial de eventos. In: FUKS, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. São José dos Campos, INPE, 2003 - on-line (3a. edição, revista e ampliada). Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/>>. Acesso em: 10 jan. 2016.

CRESSIE, N. **Statistics for Spatial Data**. New York: J. Wiley, 1993, 900 p.

COX, V.; ISHAM, D. **Point processes**. 1. ed. [S.l.]: Chapman and Hall, 1980. (Monographs on applied probability and statistics). ISBN 0412219107,9780412219108.

DALEY, D. J.; VERE-JONES, D. **An introduction to the theory of point processes: volume I: general theory and structure**. Springer Science & Business Media, 2003, 500 p.

DICK, O. B.; SAN MARTÍN, J. L.; MONTOYA, R. H.; DEL DIEGO, J.; ZAMBRANO, B.; DAYAN, G. H. The history of dengue outbreaks in the Americas. **The American journal of tropical medicine and hygiene**, v. 87, n. 4, p. 584-593, 2012.

DIGGLE, P. J. **Statistical Analysis of Spatial and Spatio-Temporal Point Patterns**. 3. ed. Boca Raton: Chapman & Hall/CRC Monographs on Statistics and Applied Probability, 2013. 297 p.

DIGGLE, P. J.; CHETWYND, A. G.; HAGGKIVIST, R., MORRIS, S. E. Second order analysis of space-time clustering. **Statistical Methods in Medical Research**, v 4,n. 2, p. 124-136, 1995.

DIGGLE, P. J.; GABRIEL, E. Spatio-temporal point processes. In: GELFAND, A.; DIGGLE P. J.; FUENTES, M.; GUTTORP, P. (Orgs.) **Handbook of spatial statistics**. Chapman and Hall/CRC, Londres, 2010. P. 449–461.

FUNDAÇÃO INSTITUTO OSWALDO CRUZ . **Aedes e dengue: vetor e doença**. Disponível em: <<http://www.ioc.fiocruz.br/dengue/textos/aedesvetoredoenca.html>>. Acesso em: 14 nov. 2016.

GABRIEL, E.; DIGGLE P. J. Second-order analysis of inhomogeneous spatio-temporal point process data. **Statistica Neerlandica**, v. 63, n.1, p.43-51, 2009.

GABRIEL, E; ROWLINGSON, B; DIGGLE, P. stpp: an R package for plotting, simulating and analyzing Spatio-Temporal Point Patterns. **Journal of Statistical Software**, v. 53, n. 2, p. 1-29, 2013.

GAUNT, M. W.; SALL, A. A.; DE LAMBALLERIE, X.; FALCONAR, A. K.; DZHIVANIAN, T. I.; GOULD, E. A. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. **Journal of General Virology**, v. 82, n. 8, p. 1867-1876, 2001.

GNEITING, T.; GUTTORP, P. Continuous Parameter Spatio-Temporal Processes. In: GELFAND, A. E.; DIGGLE, P.; GUTTORP, P.; FUENTES, M. (Ed.). **Handbook of spatial statistics**. CRC press, 2010.

GOOGLE INC. Google Earth [software]. **Version 7.1.5.1557**. 2016. Disponível em: <<https://www.google.com.br/intl/pt-BR/earth/>>.

HOLMES, E. C.; TWIDDY, S. S. The origin, emergence and evolutionary genetics of dengue virus. **Infection, genetics and evolution**, v. 3, n. 1, p. 19-28, 2003.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Cidades**, 2016. Disponível em: <<http://www.cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=3169307>>. Acesso em: 20 fev. 2016.

JACQUEZ, G. M. A k nearest neighbour test for space-time interaction. **Statistics in medicine**, v. 15, n. 18, p. 1935-1949, 1996.

KNOX, G. Detection of Low Intensity Epidemicity Application to Cleft Lip and Palate. **British journal of preventive & social medicine**, v. 17, n. 3, p. 121-127, 1963.

KULLDORFF, M. INFORMATION MANAGEMENT SERVICES INC.(2009). SaTScan™ v8. 0: Software for the spatial and space–time scan statistics. 2010.

KULLDORFF, M. **SaTScan User Manual v 9.4** . 2015. Disponível em: <<https://www.satscan.org/>>. Acesso em: 10 dez. 2016.

KULLDORFF, M.; HEFFERNAN, R.; HARTMAN, J.; ASSUNÇÃO, R.; MOSTASHARI, F. A space–time permutation scan statistic for disease outbreak detection. **Plos med**, v. 2, n. 3, p. e59, 2005.

LLOYD, Christopher D. **Local models for spatial analysis**. CRC Press, 2010.

LOECHER, M.; ROPKINS, K. RgoogleMaps and loa: unleashing R graphics power on map tiles. **Journal of Statistical Software**, v. 63, n. 4, 2015.

MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer research**, v. 27, n. 2 Part 1, p. 209-220, 1967.

MENDONÇA, F.A.; SOUZA, A. V.; DUTRA, D. A. Saúde pública, urbanização e dengue no Brasil. **Soc. nat. (Online)**, Uberlândia , v. 21, n. 3, p. 257-269, Dec. 2009 . Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1982-45132009000300003&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-45132009000300003&lng=en&nrm=iso)>. Acesso em: 07 Mar. 2016

MEYER, S.; WARNKE, I.; RÖSSLER, W.; HELD, L. Model-based testing for space–time interaction using point processes: An application to psychiatric hospital admissions in an urban area. **Spatial and spatio-temporal epidemiology**, v. 17, p. 15-25, 2016.

PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. **R news**, v. 5, n. 2, p. 9-13, 2005.

QGIS DEVELOPMENT TEAM. QGIS Geographic Information System. **Open Source Geospatial Foundation Project**. 2016. Disponível em: <<http://www.qgis.org/>>.

R CORE TEAM. R: A Language and Environment for Statistical Computing. Vienna, **Austria : the R Foundation for Statistical Computing**. 2016 . Disponível em: <<http://www.R-project.org/>>. Acesso em: 08 mar. 2016.

RIGAU-PÉREZ, J. G. The early use of break-bone fever (Quebranta huesos, 1771) and dengue (1801) in Spanish. **The American journal of tropical medicine and hygiene**, v. 59, n. 2, p. 272-274, 1998.

RIPLEY, B. D. The second-order analysis of stationary point processes. **Journal of applied probability**, JSTOR, v. 13, n. 2, p. 255–266, jun. 1976.

RIPLEY, B. D. Modelling spatial patterns. **Journal of the Royal Statistical Society. Series B** (Methodological), JSTOR, p. 172–212, 1977.

ROWLINGSON, B. S.; DIGGLE, P. J. Splancs: spatial point pattern analysis code in S-Plus. **Computers & Geosciences**, v. 19, n. 5, p. 627-655, 1993.

SCALON, J.D.; ALVES, G. F.; AVELAR, M.B.L.; ZACARIAS, M.S. Spatial distribution of coffee-leaf-miner (*Leucoptera coffeella*) in an organic coffee (*Coffea arabica* L.) field in formation. **Coffee Science**, v. 6, p. 226-232, 2011.

SCHABENBERGER, O.; GOTWAY, C. A. **Statistical methods for spatial data analysis**. Boca Raton, FL: Chapman & Hall/CRC, c 2005. xvii, 488 p. (Texts in statistical science ;). ISBN 1584883227.

SIMPSON, J.; WEINER, E. S. Oxford English dictionary. **Oxford: Clarendon Press. Retrieved March**, v. 6, p. 2008, 1989.

SNYDER, D. L.; MILLER, M. I. **Random point processes in time and space**. Springer Science & Business Media, 2012.

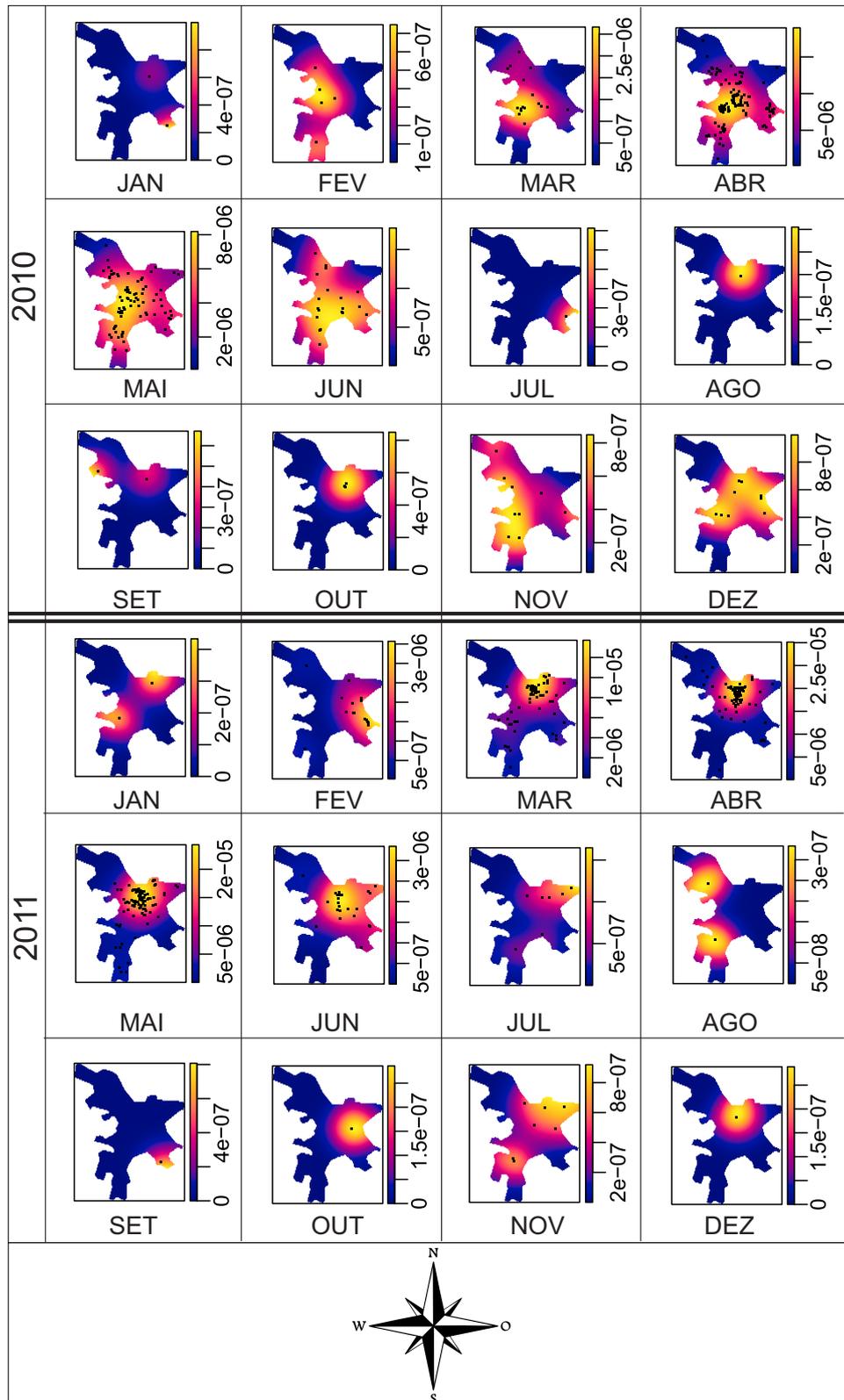
TANGO, T. **Statistical methods for disease clustering**. Springer Science & Business Media, 2010.

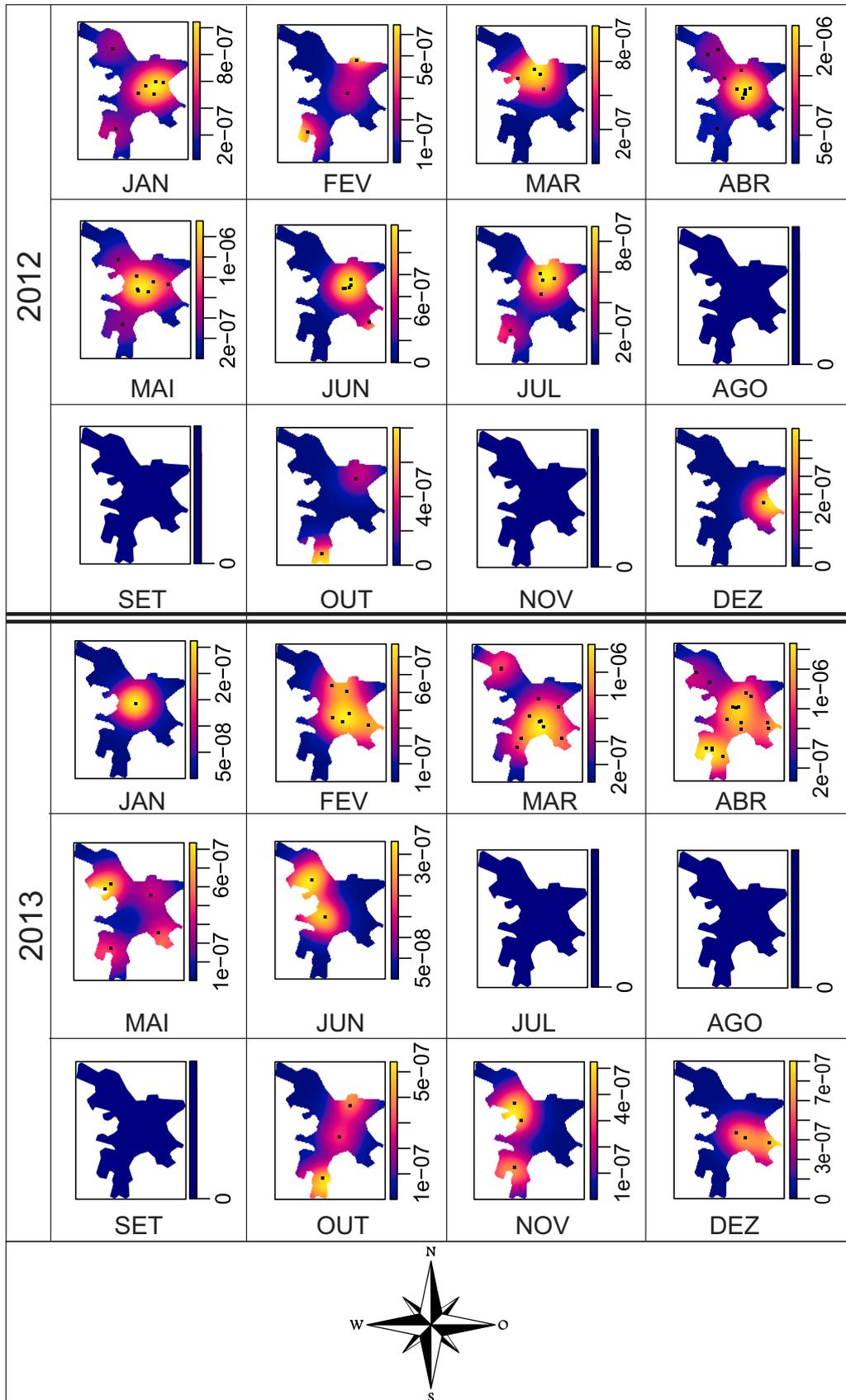
VIANA, D. V.; IGNOTTI, E. A ocorrência da dengue e variações meteorológicas no Brasil: revisão sistemática. **Revista Brasileira de Epidemiologia**, v. 16, n. 2, p. 240-256, 2013.

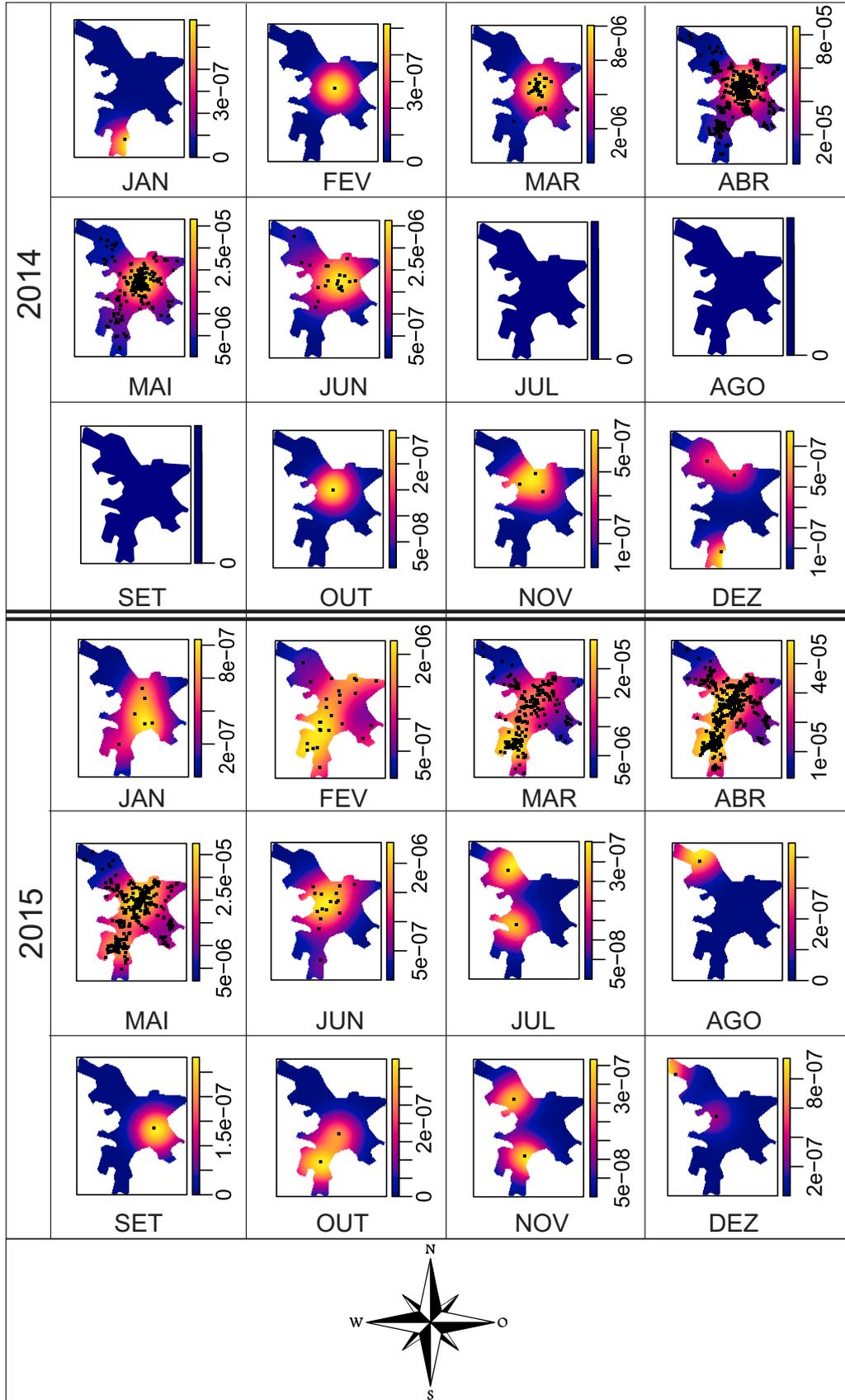
WORLD HEALTH ORGANIZATION et al. **Dengue: guidelines for diagnosis, treatment, prevention and control**. World Health Organization, 2009.

WORLD HEATH ORGANIZATION. **Frequently asked questions on dengue**. Disponível em: <[http://www.wpro.who.int/mediacentre/releases/2010/20100916\\_FAQs/en/](http://www.wpro.who.int/mediacentre/releases/2010/20100916_FAQs/en/)>. Acesso em: 16 nov 2016.

**APÊNDICE A – Estimação das intensidades das ocorrências de dengue por mês no período de 2010 a 2015 utilizando uma kernel gaussiana. Escala dada em número de casos por unidade de área ( $\text{km}^2$ ) e unidade de tempo (dia)**







### ANEXO A – Estimação da intensidade por kernel

O método de alisamento por kernel é um método não paramétrico que estima a densidade de uma variável aleatória. A função kernel tem sido bastante utilizada para se estimar a intensidade em uma determinada região de estudo, de forma a obter uma superfície ou curva contínua. Existe uma relação entre a densidade e a intensidade, que Schabenberger e Gotway (2005) abordam do seguinte modo:

Considere uma amostra aleatória  $Y_1, \dots, Y_n$  da distribuição da variável aleatória  $Y$ . Uma estimativa da função  $f(y)$  em  $y_0$  pode ser encontrada a partir do número de realizações da amostra dentro de uma certa distância  $h$  de  $y_0$ ,

$$\hat{f}(y_0) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{y_i - y_0}{h}\right), \quad (1)$$

onde  $k(t)$  é a densidade uniforme em  $-1 \leq t \leq 1$ ,

$$k(t) = \begin{cases} 0, & \text{se } |y_i - y_0| > h \\ 1, & \text{caso contrário.} \end{cases} \quad (2)$$

De acordo com os autores se a vizinhança  $h$  é pequena,  $\hat{f}(y_0)$  é um estimador quase não viesado de  $f(y_0)$ , mas sofre de grande variabilidade. Com o aumento da largura da janela  $h$ , a estimativa torna-se mais suave, menos variável e mais viesada. Assim, um procedimento baseado em erro quadrático médio, como validação cruzada, pode ser usado para determinar um valor apropriado para o parâmetro  $h$ .

Ao invés de escolher uma função kernel uniforme que dá pesos iguais para todos os pontos dentro da janela  $y_0 \pm h$ , Schabenberger e Gotway (2005) sugerem a utilização de funções kernel modais. Algumas funções kernel populares são a Gaussiana

$$k(t) = \frac{1}{\sqrt{2\pi}} e^{(-t^2/2)}, \quad (3)$$

a kernel quadrática

$$k(t) = \begin{cases} \frac{3}{4} (1 - t^2), & \text{se } |t| \leq 1 \\ 0, & \text{caso contrário.} \end{cases} \quad (4)$$

a kernel de variância mínima

$$k(t) = \begin{cases} \frac{3}{8}(3 - 5t^2), & \text{se } |t| \leq 1 \\ 0, & \text{caso contrário.} \end{cases} \quad (5)$$

ou a kernel quártica

$$k(t) = \begin{cases} \frac{15}{16}(1 - t^2)^2, & \text{se } |t| \leq 1 \\ 0, & \text{caso contrário.} \end{cases} \quad (6)$$

Schabenberger e Gotway (2005) afirmam que na prática a escolha da kernel é menos importante do que a escolha da largura de banda. A exigência para que uma função  $k(u)$  possa servir como um kernel é que  $\int k(u)du = 1$  e  $\int uk(u)du = 0$ .

Para o caso de um processo pontual espacial, a obtenção da densidade produz uma estimativa da probabilidade de se observar um evento na localização  $\mathbf{s}$ . Segundo os autores, a relação entre a densidade  $f_S(\mathbf{s})$  em  $S$  e a intensidade  $\lambda(\mathbf{s})$  é

$$\lambda(\mathbf{s}) = f_S(\mathbf{s}) \int_S \lambda(\mathbf{u})d\mathbf{u} = f_S(\mathbf{s})\mu(S). \quad (7)$$

Para a construção do estimador kernel da intensidade de primeira ordem, seja  $\nu_d(D)$  a medida de Lebesgue de um conjunto arbitrário  $D \in \mathbb{R}^d$ . Se  $d = 1$  então  $\nu_d(D)$  é a medida do comprimento de um intervalo, no  $\mathbb{R}^2$  é a medida da área de  $D$  e no  $\mathbb{R}^3$  é a medida do volume de  $D$ . A partir disso, para um processo em  $\mathbb{R}^1$ , Schabenberger e Gotway (2005) sugerem uma modificação do estimador da densidade em (1) como

$$\hat{\lambda}(s_0) = \frac{1}{\nu(S)h} \sum_{i=1}^n k\left(\frac{s_i - s_0}{h}\right), \quad (8)$$

para obter o estimador kernel da intensidade de primeira ordem.

No caso de um processo em  $\mathbb{R}^2$ , os autores afirmam que a função kernel univariada necessita ser trocada por uma função que possa acomodar duas coordenadas. Schabenberger e Gotway (2005) afirmam que por conveniência se faz uma kernel-produto obtida pela multiplicação de duas funções kernel univariadas. No entanto, essa escolha implica na ausência de interação entre as coordenadas. As bandas de largura podem ser escolhidas diferentemente para as duas dimensões, mas, segundo os autores, funções kernel com contornos esféricos tem sido comumente utilizadas. Se  $x_i$  e  $y_i$  denotam as coordenadas de  $\mathbf{s}_i$  a kernel-produto leva-nos ao

estimador da intensidade dado por

$$\hat{\lambda}(\mathbf{s}_0) = \frac{1}{v(S)h_x h_y} \sum_{i=1}^n k\left(\frac{x_i - x_0}{h_x}\right) k\left(\frac{y_i - y_0}{h_y}\right), \quad (9)$$

onde  $h_x$  e  $h_y$  são bandas de largura nas direções do sistema de coordenadas.

De acordo com Schabenberger e Gotway (2005) a independência das coordenadas pode ser contornada utilizando funções kernel bivariadas. Contornos elípticos, por exemplo, podem ser obtidos com uma função kernel gaussiana bivariada com variâncias diferentes.

As expressões acima no entanto, não consideram os efeitos de borda. Assim Diggle (1985) sugere um estimador kernel para a intensidade corrigido para o efeito de bordas, dado por

$$\hat{\lambda}(\mathbf{s}) = \frac{1}{p_h(\mathbf{s})} \sum_{i=1}^n \frac{1}{h^2} k\left(\frac{s_i - s_0}{h}\right), \quad (10)$$

em que  $p_h(\mathbf{s}) = \int_S h^{-2} k\left(\frac{\mathbf{s} - \mathbf{u}}{h}\right) du$  serve como um fator de correção do efeito de bordas.

## ANEXO B – Códigos do R para os testes

```

#####
#                                #
#                                #
#####

loc<-read.csv("coordenadasUTM.csv",header = TRUE,sep = ";")
trescor<-read.csv("contorno.csv",header = TRUE,sep=";")

#####
#                                #
#                                #
#####

# Dividindo por 1000 para colocar a unidade em km

loc$x<-loc$x/1000
loc$y<-loc$y/1000

trescor$x<-trescor$x/1000
trescor$y<-trescor$y/1000

# Deslocando para a origem

mx<-min(loc$x)
my<-min(loc$y)

loc$x<-loc$x-min(loc$x)
loc$y<-loc$y-min(loc$y)

trescor$x<-trescor$x-mx
trescor$y<-trescor$y-my

# Atribuindo as coordenadas e o tempo
xcoord<-loc[,1]
xcoord
ycoord<-loc[,2]
ycoord
time<-loc[,3]
time

```

## TESTE KNOX

```

KnoxM.test<-function(x,y,time,dell,del2,Nrep) {
# -----
# An example of execution
#   dat<-scan("d:/demo/McHardy.dat",list(x=0,y=0,time=0))
#   dell<-2; del2<-0; Nrep<-999
#   out<-KnoxM.test(dat$x,dat$y,dat$time,dell,del2,Nrep)
# -----
# ARGUEMENTS
#   x      : a vector of x-coordinates of case
#   y      : a vector of y-coordinates of case
#   time   : a vector of observed times
#   dell   : a measure of closeness in space
#   del2   : a measure of closeness in time
#   Nrep   : The number of Monte Carlo replications, e.g., 999, 9999
#
# VALUES
#   Knox.T      : test statistic
#   Freq        : a vector of simulated test statistics under the
#                 null hypothesis
#   Simulated.p.value : simulated p-value
# -----
n<-length(x)
ktmon<-1:(Nrep+1)
sdis<-matrix(0,n,n)
tdis<-matrix(0,n,n)
as<-sdis
at<-tdis
for (i in 1:n){
for (j in 1:n){
sdis[i,j]<-sqrt( (x[i]-x[j])^2 + (y[i]-y[j])^2 )
tdis[i,j]<- abs( time[i] - time[j])
}
}
for(i in 1:n){
for (j in 1:n){
as[i,j] <- ifelse(sdis[i,j]<= dell, 1, 0)
at[i,j] <- ifelse(tdis[i,j]<= del2, 1, 0)
}
}
as[i,i]<-0
at[i,i]<-0
}
s1<-0
for(i in 1:n){
for (j in 1:n){

```

```

s1<-s1+as[i,j] * at[i,j]
}
}
obst <- s1/2
#
# Monte Carlo Hypothesis testing
#
for(k in 1:Nrep){
timeR<-sample(time)

for (i in 1:n){
for (j in 1:n){
tdis[i,j]<- abs( timeR[i] - timeR[j])
}
}
for(i in 1:n){
for (j in 1:n){
at[i,j] <- ifelse(tdis[i,j]<= del2, 1, 0)
}
at[i,i]<-0
}
s1<-0
for(i in 1:n){
for (j in 1:n){
s1<-s1+as[i,j] * at[i,j]
}
}
ktmon[k] <- s1/2
}
ktmon[Nrep+1]<-obst
r<-length(ktmon[ktmon>=obst])
p<-r/(Nrep+1)
list(Knox.T=obst , Freq=ktmon, Simulated.p.value=p)
}
Knoxstat<-KnoxM.test(xcoord,ycoord,time,0.2,30,999)
Knoxstat

hist(Knoxstat$Freq,freq = F, ylab = "Densidade",xlab="",main = "")
lines(x=Knoxstat$Knox.T,type = "h",col="red",y=0.0001)

```

## TESTE MANTEL

```

Mantel.test<-function(x,y,time,c1,c2,Nrep){
# -----
# An example of execution
#
#   dat<-scan("d:/demo/McHardy.dat",list(x=0,y=0,time=0))
#   c1<-2; c2<-0; Nrep<-999
#   out<-KnoxM.test(dat$x,dat$y,dat$time,c1,c2,Nrep)
#
# -----
# ARGUMENTS
#   x      : a vector of x-coordinates of case
#   y      : a vector of y-coordinates of case
#   time   : a vector of observed times
#   c1     : a constant for Mantel's measure of closeness in space
#   c2     : a constant for Mantel's measure of closeness in time
#   Nrep   : The number of Monte Carlo replications, e.g., 999, 9999
#
# VALUES
#   Mantel.T      : test statistic
#   Expected      : expected value of test statistic
#   Simulated.p.value : simulated p-value
#   Freq          : a vector of simulated test statistics under the
#   null hypothesis
#
# -----

n<-length(x)
ktmon<-1:(Nrep+1)
sdis<-matrix(0,n,n)
tdis<-matrix(0,n,n)
as<-sdis
at<-tdis
for (i in 1:n){
for (j in 1:n){
sdis[i,j]<-sqrt( (x[i]-x[j])^2 + (y[i]-y[j])^2 )
tdis[i,j]<- abs( time[i] - time[j])
}
}
for(i in 1:n){
for (j in 1:n){
as[i,j] <- 1/(sdis[i,j]+c1)
at[i,j] <- 1/(tdis[i,j]+c2)
}
}
as[i,i]<-0

```

```

at[i,i]<-0
}
s1<-0; s2<-0; s3<-0
for(i in 1:n){
for (j in 1:n){
s1<-s1+as[i,j] * at[i,j]
s2<-s2+as[i,j]
s3<-s3+at[i,j]
}
}
obst <- s1/2
Ms <- s2/2
Mt <- s3/2
ee <- Ms*Mt/(n*(n-1)/2)
#
# Monte Carlo Hypothesis testing
#
for(k in 1:Nrep){
timeR<-sample(time)

for (i in 1:n){
for (j in 1:n){
tdis[i,j]<- abs( timeR[i] - timeR[j])
}
}
for(i in 1:n){
for (j in 1:n){
at[i,j] <- 1/(tdis[i,j]+c2)
}
}
at[i,i]<-0
}
s1<-0
for(i in 1:n){
for (j in 1:n){
s1<-s1+as[i,j] * at[i,j]
}
}
ktmon[k] <- s1/2
}
ktmon[Nrep+1]<-obst
r<-length(ktmon[ktmon>=obst])
p<-r/(Nrep+1)
list(Mantel.T=obst, Expected=ee, Simulated.p.value=p,Freq=ktmon)
}
Mantelstat<-Mantel.test(xcoord,ycoord,time,0.1,15,999)
Mantelstat
hist(Mantelstat$Freq,freq = F, ylab = "Densidade",xlab="",main = "")
lines(x=Mantelstat$Mantel.T,type = "h",col="red",y=0.0001)

```

## TESTE JACQUEZ

```

Jacquez.test<-function(x,y,time,k,Nrep) {
# -----
# An example of execution
#   dat<-scan("d:/demo/McHardy.dat",list(x=0,y=0,time=0))
#   k<-1; Nrep<-999
#   out<-Jacquez.test(dat$x,dat$y,dat$t,k,Nrep)
# -----
# ARGUMENTS
#   x      : a vector of x-coordinates of case
#   y      : a vector of y-coordinates of case
#   time   : a vector of observed times
#   k      : k of k nearest neighbors
#   Nrep   : The number of Monte Carlo replications, e.g., 999, 9999
# VALUES
#   Jacquez.T      : test statistic
#   Expected       : expected value of test statistic
#   Simulated.p.value : simulated p-value
#   Freq          : a vector of simulated test statistics under
#                   the null hypothesis
# -----
n<-length(x)
time<-time+runif(n)/100
ktmon<-1:(Nrep+1)
sdis<-matrix(0,n,n)
tdis<-matrix(0,n,n)
as<-sdis; at<-tdis
for (i in 1:n){
for (j in 1:n){
sdis[i,j]<-sqrt( (x[i]-x[j])^2 + (y[i]-y[j])^2 )
tdis[i,j]<- abs( time[i] - time[j])
}
}
for(i in 1:n){
ff<-order(sdis[i,])
gg<-order(tdis[i,])
for(v in 1:k){
m1<-ff[v+1]
m2<-gg[v+1]
as[i,m1]<-1
at[i,m2]<-1
}
}
}

```

```

s1<-0; s2<-0; s3<-0
for(i in 1:n){
  for (j in 1:n){
    s1<-s1+as[i,j] * at[i,j]
    s2<-s2+as[i,j]
    s3<-s3+at[i,j]
  }
}
obst <- s1/2
Ms <- s2/2
Mt <- s3/2
ee <- Ms*Mt/(n*(n-1)/2)
#
# Monte Carlo Hypothesis testing
#
for(ij in 1:Nrep){
  timeR<-sample(time)
  for (i in 1:n){
    for (j in 1:n){
      tdis[i,j]<- abs( timeR[i] - timeR[j])
    }
  }
  at[,]<-0
  for(i in 1:n){
    gg<-order(tdis[i,])
    for(v in 1:k){
      m2<-gg[v+1]
      at[i,m2]<-1
    }
  }
  s1<-0
  for(i in 1:n){
    for (j in 1:n){
      s1<-s1+as[i,j] * at[i,j]
    }
  }
  ktmon[ij] <- s1/2
}
ktmon[Nrep+1]<-obst
r<-length(ktmon[ktmon>=obst])
p<-r/(Nrep+1)
list(Jacquez.T=obst, Expected=ee, Simulated.p.value=p, Freq=ktmon)
}
Jacquezstat<-Jacquez.test(xcoord,ycoord,time,20,999)
Jacquezstat

hist(Jacquez.test$Freq,freq = F, ylab = "Densidade",xlab="",main = "")
lines(x=Jacquez.T,type = "h",col="red",y=0.0001)

```

## FUNÇÃO K HOMOGÊNEA

```

library(stpp)
Khom.Test<-function(xyt,poly,Nrep){
#-----#
# Argumentos
# xyt : um arquivo de coordenadas xy e tempos t
# poly : um arquivo de coordenadas xy do contorno
# Nrep : número de repetições para o procedimento
#       de Monte Carlo
#-----#
locTC <- as.3dpoints(xyt)
TC<-as.points(poly)

# estimação da função K homogênea
u <- seq(0, 10, by = 1)
v <- seq(0, 30, by = 1)
stk <- STIKhat(xyt = locTC,s.region = TC, t.region = c(0,2191),
dist = u, times = v, infectious = TRUE)

# Computando a estatística de teste
K1<-sum(((stk$Khat-stk$Ktheo)^2)/(stk$Ktheo/pi))
ktest<-K1
#-----#
# Fazendo as simulações

#simulando sobre o processo de Poisson
for (j in 2:Nrep1){
hpp<-rpp(lambda=0.05845632, s.region=TC,t.region = c(0,2191),
npoints = 2818, discrete.time = F)
# estimação da função K homogênea
SimStk <- STIKhat(xyt = hpp$xyt,s.region = TC, t.region = c(0,2191),
dist = u, times = v, infectious = TRUE)

Ksim<-sum(((SimStk$Khat-SimStk$Ktheo)^2)/(SimStk$Ktheo/pi))
K1<-c(K1,Ksim)
}
Pvalue<-rank(-K1)[1]/(Nrep1)
list(K.teste=ktest, valorP=Pvalue,Freq.D=K1)
}
Kteste<-Khom.Test(loc,trescor,999)
Kteste
hist(Kteste$Freq.D,freq = F, ylab = "Densidade",xlab="",main = "")
lines(x=Kteste$ktest,type = "h",col="red",y=0.0001)

```

## FUNÇÃO K NÃO HOMOGÊNEA

```

library(stpp)
library(stpp)

KIhom.Test<-function(xyt,poly,Nrep){
#-----#
# Argumentos
# xyt : um arquivo de coordenadas xy e tempos t
# poly : um arquivo de coordenadas xy do contorno
# Nrep : número de permutações para o procedimento
#       de Monte Carlo
#-----#
locTC <- as.3dpoints(xyt)
TC<-as.points(poly)

Mt <- density(locTC[,3],kernel = "gaussian", n = 200, bw=10)
mut <- Mt$y[findInterval(locTC[,3], Mt$x)] * dim(locTC)[1]
h <- mse2d(as.points(locTC[,1:2]), TC, nsmse = 50, range = 10)
h <- h$h[which.min(h$mse)]
Ms <- kernel2d(as.points(locTC[,1:2]), TC, h = h, nx = 200,
ny = 200)
atx <- findInterval(x = locTC[,1], vec = Ms$x)
aty <- findInterval(x = locTC[,2], vec = Ms$y)
mhat <- NULL
for(i in 1:length(atx)) mhat <- c(mhat, Ms$z[atx[i],aty[i]])
u <- seq(0, 10, by = 1)
v <- seq(0, 30, by = 1)
stik <- STIKhat(xyt = locTC, s.region = TC, t.region = c(1, 2191),
lambda = mhat * mut / dim(locTC)[1], dist = u, times = v,
infectious = TRUE)

# Computando a estatística de teste
K1<-sum(((stik$Khat-stik$Ktheo)^2)/(stik$Ktheo/pi))
ktest<-K1
#-----#
# Fazendo as simulações

h <- mse2d(as.points(locTC[, 1:2]), TC, nsmse = 500, range = 10)
h <- h$h[which.min(h$mse)]
Ls <- kernel2d(as.points(locTC[, 1:2]), TC, h, nx = 200, ny = 200)
Lt <- dim(locTC)[1] * density(locTC[, 3], n = 400)$y
Lst <- array(0, dim = c(200, 200, 400))
for(k in 1:400) Lst[, ,k] <- Ls$z * Lt[k] / dim(locTC)[1]

#simulando sobre o processo de Poisson

```

```

for (j in 2:Nrep+1){
ipp2 <- rpp(lambda = Lst, s.region = TC, t.region = c(1, 2191),
npoints = 2818, discrete.time = FALSE)

ipp <- as.3dpoints(ipp2$xyt[,1],ipp2$xyt[,2],ipp2$xyt[,3])

### ESTIMANDO A FUNÇÃO

Mt <- density(ipp[,3], kernel="gaussian", n = 200,bw=10)
mut <- Mt$y[findInterval(ipp[,3], Mt$x)] * dim(ipp)[1]
h <- mse2d(as.points(ipp[,1:2]), TC, nsmse = 50, range = 10)
h <- h$h[which.min(h$mse)]
Ms <- kernel2d(as.points(ipp[,1:2]), TC, h = h, nx = 200,
ny = 200)
atx <- findInterval(x = ipp[,1], vec = Ms$x)
aty <- findInterval(x = ipp[,2], vec = Ms$y)
mhat <- NULL
for(i in 1:length(atx)) mhat <- c(mhat, Ms$z[atx[i],aty[i]])
u <- seq(0, 10, by = 1)
v <- seq(0, 30, by = 1)
SimStik <- STIKhat(xyt = ipp, s.region = TC, t.region = c(1, 2191),
lambda = mhat * mut / dim(ipp)[1], dist = u, times = v,
infectious = TRUE)

Ksim<-sum(((SimStik$Khat-SimStik$Ktheo)^2)/(SimStik$Ktheo/pi))
K1<-c(K1,Ksim)
}
Pvalue<-rank(-K1)[1]/(Nrep+1)
list(K.teste=ktest, valorP=Pvalue,Freq.D=K1)
}
KITeste<-KIhom.Test(loc,trescor,999)
KITeste

hist(KITeste$Freq.D,freq = F, ylab = "Densidade",xlab="",main = "")
lines(x=KITeste$ktest,type = "h",col="red",y=0.0001)

```