



**RAFAEL AGOSTINHO FERREIRA**

**MODELAGEM E PREDIÇÃO ESPAÇO-TEMPORAL DOS  
CASOS DE DENGUE UTILIZANDO PROCESSO PONTUAL  
DE COX LOG-GAUSSIANO**

**LAVRAS – MG**

**2017**

**RAFAEL AGOSTINHO FERREIRA**

**MODELAGEM E PREDIÇÃO ESPAÇO-TEMPORAL DOS CASOS DE DENGUE  
UTILIZANDO PROCESSO PONTUAL DE COX LOG-GAUSSIANO**

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.

Dr. João Domingos Scalon  
Orientador

**LAVRAS – MG**

**2017**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Ferreira, Rafael Agostinho

Modelagem e predição espaço-temporal dos casos de dengue utilizando processo pontual de Cox log-Gaussiano / Rafael Agostinho Ferreira. – Lavras : UFLA, 2017.

82 p. : il.

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

Orientador: Dr. João Domingos Scalon.

Bibliografia.

1. Dengue. 2. Processos Pontuais. 3. Modelo espaço-temporal. I. Universidade Federal de Lavras. II. Título.

**RAFAEL AGOSTINHO FERREIRA**

**MODELAGEM E PREDIÇÃO ESPAÇO-TEMPORAL DOS CASOS DE DENGUE  
UTILIZANDO PROCESSO PONTUAL DE COX LOG-GAUSSIANO**

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 10 de Março de 2017.

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

Dr. João Domingos Scalon  
Orientador

**LAVRAS – MG  
2017**

*À DEUS, razão da minha vida.  
Aos meus pais, por todo apoio e amor;  
As minhas amadas irmãs, pelo carinho e amizade.  
**Dedico.***

## AGRADECIMENTOS

À Deus, que é o motivo e razão de todas as coisas! Obrigado Senhor por nunca me desamparar, por me direcionar no caminho certo e me mostrar que posso prosseguir sem medo, mesmo quando tudo parece o contrário. A cada dia, pude ver a Tua fidelidade em minha vida! Ao Senhor toda a glória!

Aos meus pais, Paulo e Márcia, minha maior fonte de inspiração. Obrigado por acreditarem em mim e me apoiarem incondicionalmente. Jamais me esquecerei de todos os sacrifícios que já fizeram por mim. Esta vitória também é de vocês!

Às minhas preciosas irmãs, Caroline e Veridiane, por estarem sempre ao meu lado com todo amor e afeto. À minha prima (e irmã de coração) Camila por todo carinho e amizade. Ao meu cunhado Gustavo por todo apoio. Não há como descrever a alegria que eu sinto em ter vocês ao meu lado.

À minha amada sobrinha, Ana Júlia, pelo simples fato de existir. Obrigado por alegrar os meus dias com o seu sorriso cativante.

À minha querida família de Lavras: Antônio Henrique, Isah, Lauro, Neander, Silvana, Thaís, Thiago e Vanésia. Em especial, minha querida tia Ilma, que se tornou minha segunda mãe. Nada será suficiente para expressar minha eterna gratidão pelo que fizeram e fazem por mim.

Ao professor Dr. João Domingos Scalon por ter aceitado me orientar e ter contribuído, com excelência, para minha formação e para que este trabalho pudesse ser concluído.

Ao meu amigo Rodrigo pelo companheirismo durante esta jornada. Muito obrigado por toda contribuição durante todo o período de desenvolvimento do trabalho.

À Vigilância Epidemiológica da Secretaria de Saúde dos municípios de Três Corações e Lavras por cederem, gentilmente, os dados de dengue para que pudéssemos realizar esta pesquisa.

Aos membros da banca, pelas contribuições para a melhoria deste trabalho.

Aos colegas do Grupo de Pesquisa em Estatística Espacial (GPS) por me proporcionar a oportunidade de conhecer melhor a área da estatística espacial.

Aos meus colegas de mestrado André, Andrezza, Cristian, Eleanderson, Érica, Ernandes, Hercílio, Rodrigo, Rogério, Sérgio e Taís por todo o companheirismo e amizade durante estes dois anos.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Estatística (DES) pela oportunidade de realizar meu mestrado.

Aos excelentes professores do Programa de Pós-graduação em Estatística e Experimentação Agropecuária por todo conhecimento transmitido a mim.

Aos servidores do Departamento de Estatística (DES), em especial a secretária Nádia por toda ajuda.

Aos professores do Instituto de Ciências Sociais Aplicadas da Universidade Federal de Alfenas, campus Varginha, o qual serei eternamente grato por terem sido os responsáveis por toda a base do meu conhecimento em estatística, adquirido através do curso de Ciências Atuariais.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pela concessão da bolsa de estudos.

À todos os amigos e parentes, não citados anteriormente, mas não menos importantes e que sempre torcem pelo meu sucesso.

**Muito Obrigado!**

*“Corramos com perseverança a corrida  
que nos é proposta, tendo os olhos fitos  
em Jesus, autor e consumador da nossa fé.”  
(Hebreus 12:1-2)*

## RESUMO

A dengue é uma doença viral infecciosa que tem causado grandes preocupações para a saúde pública no Brasil nos últimos anos. Dentre os estados brasileiros que mais sofrem com esta doença, destaca-se o estado de Minas Gerais. Conhecer o comportamento do vírus da dengue em relação a sua forma de propagação em locais e épocas de grande incidência é de suma importância, a fim de que se possa reduzir o número de ocorrências. A Estatística pode ser uma ferramenta importante para o combate da doença, principalmente com a utilização de métodos e técnicas que consideram informações de tempo e localização espacial relevantes na análise. Assim, o objetivo deste trabalho foi descrever o comportamento da dengue utilizando o modelo de Cox log-Gaussiano espaço-temporal para identificar regiões onde o risco da doença é alto. A análise considerou notificações dos casos ocorridos no município de Três Corações no período de 2010 a 2015. Constatou-se, através de análises descritivas, de modelagem e de predição, que o período com grande número de ocorrências acontece entre os meses de Fevereiro e Junho. As regiões que se mostraram com alto risco da doença abrangeram os bairros Però Um, Però Dois, Santana, Parque São José, Vila Lima, Loteamento Bela Vista, Odilon Rezende, Vila Gesse, Monte Alegre, Centro, Jardim Santa Tereza, Cotia, Vila Fernão Dias, Vila Santo Afonso, Jardim Califórnia, Jardim Paraíso, São Jerônimo e Cinturão Verde. Com estes resultados, observou-se uma certa proximidade de ocorrências entre os dias e, também, entre os bairros, o que reflete a presença de aglomerados de casos de dengue tanto no tempo quanto no espaço, característica típica da doença.

**Palavras-chave:** Dengue. Processos pontuais. Processo de Cox. Espaço-tempo.

## ABSTRACT

Dengue fever is an infectious viral disease that has caused great concerns for public health in Brazil in recent years. Among the Brazilian states that suffer the most from this disease, the Minas Gerais' state stands out. Knowing the behavior of the dengue virus in relation to its way of propagation in places and times of great incidence is of paramount importance in order to reduce the number of these occurrences. The statistic can be an important tool for combating the disease, mainly using methods and techniques that consider time and spatial location relevant informations in the analysis. In this work, it describes the behavior of dengue using the Cox log-Gaussian model for space-time in order to identify regions where the risk of disease is high. Analysis considered notifications of the cases occurred in the Três Corações city during the years 2010 to 2015. It was verified through descriptive analysis, modeling and prediction that the period with large number of occurrences happens between February and June. Areas that were at high risk of the disease included the following neighborhoods: Però Um, Però Dois, Santana, Parque São José, Vila Lima, Loteamento Bela Vista, Odilon Rezende, Vila Gesse, Monte Alegre, Centro, Jardim Santa Tereza, Cotia, Vila Fernão Dias, Vila Santo Afonso, Jardim Califórnia, Jardim Paraíso, São Jerônimo and Cinturão Verde. With these results, we observed a certain proximity among days and these neighborhoods too, which reflects the presence of clusters of dengue's cases both in time and space, typical feature of this disease.

**Keywords:** Dengue fever. Point processes. Cox process. Space-time.

## LISTA DE FIGURAS

Figura 2.1 – Processo de reinfestação de casos de dengue na América Central e Latina entre 1930 a 2007. . . . .	8
Figura 2.2 – Formas clínicas da doença. . . . .	9
Figura 2.3 – Ilustração gráfica do estimador da intensidade por kernel. . . . .	14
Figura 2.4 – Padrões de interação entre eventos. . . . .	15
Figura 2.5 – Gráfico de análise da Função $g$ espacial comparando sua estimativa (em vermelho) com a função teórica (linha tracejada) para os três cenários possíveis. . . . .	18
Figura 2.6 – Exemplificação do processo de estimação da Função $K$ espacial para três eventos. . . . .	20
Figura 2.7 – Gráfico da Função $K$ espacial comparando sua estimativa (em vermelho) com a função teórica (suposição de completa aleatoriedade espacial) para os três cenários possíveis. . . . .	21
Figura 2.8 – Exemplificação do método de estimação da Função $K$ temporal para um evento $t_i$ . . . . .	23
Figura 2.9 – Exemplificação do método de estimação da Função $K$ espaço-temporal para um evento. . . . .	28
Figura 2.10 – Ilustração gráfica em relação ao método de estimação via Mínimos Contrastes. . . . .	40
Figura 3.1 – Município de Três Corações, localizado no sul de Minas Gerais, com identificação da área urbana no município. . . . .	47
Figura 4.1 – Distribuição dos casos de dengue na área urbana do município de Três Corações - MG, nos anos de 2010 a 2015. . . . .	49
Figura 4.2 – Frequência dos casos de dengue em dias durante os anos de 2010 e 2015. . . . .	50
Figura 4.3 – Estimativa da densidade espacial por alisamento de Kernel. . . . .	51
Figura 4.4 – Estimativa da intensidade temporal ajustada por um modelo de regressão de Poisson (em vermelho). . . . .	53
Figura 4.5 – Gráfico da função de correlação entre pares (Função $g$ ) e da função de autocovariância temporal ajustados pelo método de mínimos contrastes. . . . .	55
Figura 4.6 – Mapa de probabilidade de excedência da predição para os dias 24 de Janeiro à 08 de Fevereiro de 2015. . . . .	57
Figura 4.7 – Mapa de probabilidade de excedência da predição para os dias 09 de Fevereiro à 24 de Fevereiro de 2015. . . . .	58

Figura 4.8 – Mapa de probabilidade de excedência da predição para os dias 25 de Fevereiro à 12 de Março de 2015. . . . .	59
Figura 4.9 – Mapa de probabilidade de excedência da predição para os dias 13 de Março à 28 de Março de 2015. . . . .	60
Figura 4.10 – Mapa de probabilidade de excedência da predição para os dias 29 de Março à 13 de Abril de 2015. . . . .	61
Figura 4.11 – Mapa de probabilidade de excedência da predição para os dias 14 de Abril à 29 de Abril de 2015. . . . .	62
Figura 4.12 – Mapa de probabilidade de excedência da predição para os dias 30 de Abril à 15 de Maio de 2015. . . . .	63
Figura 4.13 – Mapa de probabilidade de excedência da predição para os dias 16 de Maio à 31 de Maio de 2015. . . . .	64
Figura 4.14 – Mapa de probabilidade de excedência da predição para os dias 01 de Junho à 16 de Junho de 2015. . . . .	65
Figura 4.15 – Áreas de risco da doença observadas na cidade de Três Corações - MG durante os dias 05 de março à 30 de maio de 2015. . . . .	67
Figura 1 – Erro padrão de predição para os dias 24 de Janeiro à 08 de Fevereiro. . . . .	74
Figura 2 – Erro padrão de predição para os dias 09 de Fevereiro à 24 de Fevereiro. . . . .	75
Figura 3 – Erro padrão de predição para os dias 25 de Fevereiro à 12 de Março. . . . .	76
Figura 4 – Erro padrão de predição para os dias 13 de Março à 28 de Março. . . . .	77
Figura 5 – Erro padrão de predição para os dias 29 de Março à 13 de Abril. . . . .	78
Figura 6 – Erro padrão de predição para os dias 14 de Abril à 29 de Abril. . . . .	79
Figura 7 – Erro padrão de predição para os dias 30 de Abril à 15 de Maio. . . . .	80
Figura 8 – Erro padrão de predição para os dias 16 de Maio à 31 de Maio. . . . .	81
Figura 9 – Erro padrão de predição para os dias 01 de Junho à 16 de Junho. . . . .	82

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>5</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>7</b>
<b>2.1</b>	<b>Dengue</b>	<b>7</b>
<b>2.1.1</b>	<b>Aspectos históricos</b>	<b>7</b>
<b>2.1.2</b>	<b>Aspectos epidemiológicos da dengue</b>	<b>8</b>
<b>2.2</b>	<b>Estatística Espacial</b>	<b>10</b>
<b>2.3</b>	<b>Processos pontuais espaciais</b>	<b>11</b>
<b>2.3.1</b>	<b>Propriedades de primeira ordem</b>	<b>12</b>
<b>2.3.1.1</b>	<b>Estimador da intensidade por kernel</b>	<b>12</b>
<b>2.3.2</b>	<b>Propriedades de segunda ordem</b>	<b>14</b>
<b>2.3.2.1</b>	<b>Estacionariedade e isotropia</b>	<b>16</b>
<b>2.3.2.2</b>	<b>Função de correlação entre pares espacial</b>	<b>16</b>
<b>2.3.2.3</b>	<b>Função <math>K</math> espacial</b>	<b>18</b>
<b>2.4</b>	<b>Processos pontuais temporais</b>	<b>21</b>
<b>2.4.1</b>	<b>Função <math>K</math> temporal</b>	<b>22</b>
<b>2.5</b>	<b>Processos pontuais espaço-temporal</b>	<b>23</b>
<b>2.5.1</b>	<b>Propriedades de primeira ordem</b>	<b>24</b>
<b>2.5.2</b>	<b>Propriedades de segunda ordem</b>	<b>24</b>
<b>2.5.2.1</b>	<b>Função de correlação entre pares espaço-temporal</b>	<b>25</b>
<b>2.5.2.2</b>	<b>Função <math>K</math> espaço-temporal</b>	<b>26</b>
<b>2.6</b>	<b>Campos Aleatórios Gaussianos</b>	<b>29</b>
<b>2.7</b>	<b>Modelos espaço-temporais</b>	<b>32</b>
<b>2.7.1</b>	<b>Processo de Poisson</b>	<b>32</b>
<b>2.7.1.1</b>	<b>Processo de Poisson homogêneo</b>	<b>32</b>
<b>2.7.1.2</b>	<b>Processo de Poisson não-homogêneo</b>	<b>33</b>
<b>2.7.2</b>	<b>Processo de Cox log-Gaussiano</b>	<b>34</b>
<b>2.7.2.1</b>	<b>Estimação do modelo</b>	<b>39</b>
<b>2.7.2.2</b>	<b>Método de mínimos contrastes</b>	<b>39</b>
<b>2.8</b>	<b>Predição espaço-temporal</b>	<b>43</b>
<b>2.8.1</b>	<b>Simulação Monte Carlo via Cadeias de Markov</b>	<b>43</b>
<b>2.8.2</b>	<b>Probabilidade de excedência</b>	<b>45</b>

<b>3</b>	<b>MATERIAIS E MÉTODOS</b> . . . . .	<b>47</b>
<b>3.1</b>	<b>Descrição dos dados</b> . . . . .	<b>47</b>
<b>3.2</b>	<b>Softwares</b> . . . . .	<b>48</b>
<b>4</b>	<b>RESULTADOS</b> . . . . .	<b>49</b>
<b>4.1</b>	<b>Análises preliminares</b> . . . . .	<b>49</b>
<b>4.2</b>	<b>Ajuste do modelo</b> . . . . .	<b>51</b>
<b>4.3</b>	<b>Predição espaço-temporal</b> . . . . .	<b>56</b>
<b>5</b>	<b>CONCLUSÕES</b> . . . . .	<b>69</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>70</b>
	<b>APENDICE A – Mapas para o erro de predição</b> . . . . .	<b>74</b>

## 1 INTRODUÇÃO

Nos últimos anos, a dengue tem se tornado um grande problema de saúde pública mundial e tem causado forte impacto no campo epidemiológico. A transmissão da doença ocorre, principalmente, pela picada do mosquito *Aedes aegypti*, principal transmissor da doença, que se alimenta de sangue humano e que é comumente encontrado em áreas urbanas (DIAS et al., 2010).

Segundo Braga e Valle (2007), esse problema está relacionado às mudanças demográficas ocorridas durante a segunda metade do século XX. Um grande fluxo migratório das zonas rurais para as áreas urbanas, atrelado ao crescimento populacional desenfreado e à falta de saneamento básico, impulsionaram o crescimento de tais ocorrências.

Dados divulgados pela Organização Mundial de Saúde (OMS) estimam que, cerca de 50 milhões de casos de dengue ocorrem, em média, durante o ano e que aproximadamente 2,5 bilhões de pessoas, cerca 1/3 da população mundial, vivem em países endêmicos (WORLD HEALTH ORGANIZATION et al., 2009).

No Brasil a situação não é diferente. Por ser um país tropical, esses mosquitos se proliferaram devido ao clima quente e úmido. Segundo boletins divulgados pelo Ministério da Saúde, no período de 2014 foram notificados 589.107 casos de dengue. Em 2015, esse número aumentou para 1.649.008 casos, o que revela um aumento de aproximadamente 179,9% de casos em relação ao ano anterior (MINISTÉRIO DA SAÚDE, 2016).

A dengue é um problema que tem afetado a saúde de diversos brasileiros, contribuindo para um aumento significativo nas taxas de mortalidade. Dentre as regiões brasileiras, a que mais se destacou em relação ao número de casos de dengue, em 2015, foi a região Sudeste, com 62,2% dos casos. Os estados de São Paulo e Minas Gerais foram os principais responsáveis.

Assim, é necessário que medidas severas de prevenção sejam tomadas pelos órgãos competentes, a fim de que tais números possam ser reduzidos. Para isso, é importante detectar precocemente possíveis epidemias em locais de risco da doença.

Técnicas e métodos de análise estatística têm sido utilizados para essa finalidade, principalmente aquelas que consideram a informação temporal e espacial relevantes para entender o fenômeno. Assim, com as devidas informações, os órgãos públicos voltados para a saúde poderão gerenciar, de forma eficaz, os programas de prevenção e conscientização da sociedade, reduzindo o impacto dessa doença nos níveis de mortalidade da população.

Diante desse contexto, o presente estudo tem por objetivo realizar uma análise descritiva e a modelagem dos casos de dengue notificados no município de Três Corações, situado na região sul do estado de Minas Gerais, no período de 2010 a 2015. A análise será realizada com enfoque em processos pontuais espaço-temporal, campo de interesse da Estatística Espacial.

Como objetivos específicos, tem-se:

- Ajustar o modelo de Cox log-Gaussiano espaço-temporal ao conjunto de dados;
- Realizar previsões espaço-temporal dos casos de dengue;
- Identificar regiões e períodos de tempo cujo o risco de epidemias da doença seja alto.

Com este estudo, espera-se conhecer melhor o padrão da doença no município em estudo através do modelo a ser utilizado, a fim de identificar locais e momentos de grande risco epidêmico.

Como hipótese, acredita-se que casos de dengue tendem a ocorrer próximos uns dos outros, apresentando um padrão de agrupamento, tanto no espaço quanto no tempo, que pode ser descrito, indiretamente, através de um modelo de Cox log-Gaussiano.

## 2 REFERENCIAL TEÓRICO

Nesta seção serão apresentadas informações referentes aos assuntos abordados neste trabalho. Primeiramente, será exposta uma breve contextualização referente à dengue. Abordagens relacionadas à teoria espacial de dados serão brevemente apresentadas. Posteriormente, será introduzida a abordagem de processos pontuais sob o ponto de vista espacial, temporal e espaço-temporal. Por fim, modelos espaço-temporais serão apresentados.

### 2.1 Dengue

#### 2.1.1 Aspectos históricos

A origem da palavra "dengue" se remete à Espanha no século XIX, que vem da expressão *swahili ka dinga pepo*, que significa espasmo muscular causado pelos maus espíritos (CASTELLANOS; BELLO; VELANDIA-ROMERO, 2014). Anteriormente, várias nomenclaturas foram utilizadas para caracterizar a doença, variando de região para região, tais como: "febre da China", na Ásia; "bouhou", na Oceania; "polca", no Rio de Janeiro; "patuléia", na Bahia; entre outras (REZENDE, 1997). Porém, somente em 1839 é que a palavra "dengue" passou a ser usada para se referir à doença como hoje é conhecida.

Os primeiros registros encontrados na história, referente a casos que, possivelmente, poderiam estar relacionados à dengue, foram escritos em uma enciclopédia chinesa de sintomas e remédios de doenças, publicada entre 265 e 420 d.C. na China. A doença era chamada pelos chineses de "veneno da água", pois acreditavam que a doença se dava por insetos voadores ligados à água (GUBLER, 1998).

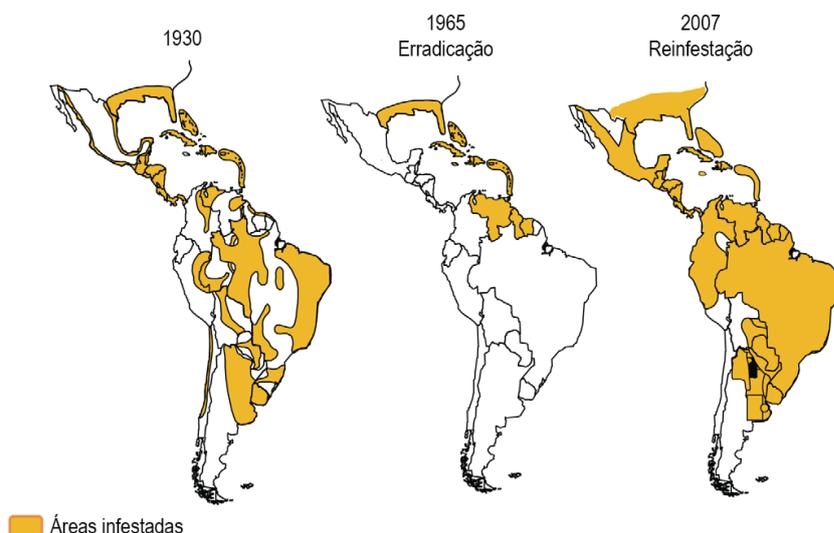
De acordo com Dick et al. (2012), as primeiras evidências de epidemias de dengue no continente americano aconteceram em Martinica e em Guadalupe, no Panamá, no ano de 1635. No entanto, segundo Rezende (1997), os primeiros casos da doença relatados na literatura médica, tais como é conhecida hoje, foram descritos por David Bylon, em 1780, e Benjamin Rush, em 1789, que descreveram um surto epidêmico ocorrido em Java e na Filadélfia, nos Estados Unidos, respectivamente.

No século seguinte, a região do Caribe e o sul dos Estados Unidos foram afetados por novos surtos epidêmicos de dengue (PINHEIRO; NELSON, 1997). Entre o final do século XIX e início do século XX, uma das maiores incidências de casos se alastraram por diversos países

da América, que iam desde o norte dos Estados Unidos até o sul do Chile e Argentina (DICK et al., 2012).

No Brasil, os primeiros registros de epidemias de dengue ocorreram nos períodos de 1846 à 1848 e 1851 à 1853. Novos casos foram notificados no início do século XX, nos anos de 1916 e 1923 (VASCONCELOS et al., 1999). Nos anos seguintes, especificamente em 1947, a Organização Pan-Americana de Saúde (OPAS) começou a elaborar um plano de erradicação do mosquito *Aedes aegypti* em países da América Latina e Central, incluindo o Brasil (TAPIA-CONYER; MENDEZ-GALVAN; GALLARDO-RINCON, 2009). Entre 1952 e 1965, 19 países foram certificados como livres do mosquito. No entanto, em 1967 o mosquito voltou a proliferar nessas áreas, perpetuando-se até o ano de 2007, como pode ser visto na Figura 2.1.

Figura 2.1 – Processo de reinfestação de casos de dengue na América Central e Latina entre 1930 a 2007.



Fonte: Adaptado de Tapia-Conyer, Mendez-Galvan e Gallardo-Rincon (2009)

No início do século XXI, o Brasil passou a ocupar a primeira posição no ranking mundial de casos reportados de dengue e está entre os dez países de maior risco dessa doença, segundo Teixeira (2012).

### 2.1.2 Aspectos epidemiológicos da dengue

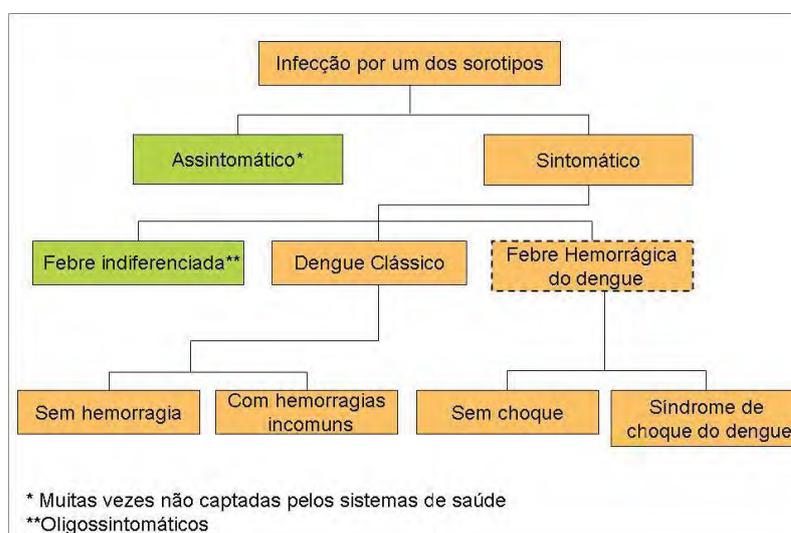
A dengue é uma doença viral, transmitida por mosquitos do gênero *Aedes*, no qual os principais agentes transmissores são os mosquitos *Aedes aegypti* e *Aedes albopictus* (GUBLER, 1998). Segundo Bicalho, Safadi e Charret (2014), o *Aedes Aegypti* é um mosquito da família *Culicidae*, comumente encontrado em áreas urbanas de regiões tropicais e subtropicais, que

utiliza, preferencialmente, depósitos de água para se proliferarem. Já o *Aedes albopictus* pode ser encontrado em áreas rurais e florestais. Alimenta-se tanto de sangue humano quanto de outros animais e podem se adaptar em ambientes de temperaturas baixas, diferentemente do *Aedes Aegypti* (CATÃO, 2011). A transmissão do vírus da dengue ao ser humano é feita pela picada desses mosquitos.

O vírus da dengue pertence ao gênero *Flavivirus* e à família *Flaviviridae* que, segundo Gubler (1998), contém, aproximadamente, 70 tipos de vírus, entre eles o que transmite a febre amarela. Segundo Tauil (2001), o vírus da dengue pode ser classificado de acordo com quatro sorotipos: DENV-1, DENV-2, DENV-3 e DENV-4, os quais possuem características genéticas distintas. A infecção por um desses sorotipos garante imunidade permanente, ou duradoura, ao paciente de acordo com o sorotipo no qual fora infectado. No entanto, não é garantida imunidade em relação aos outros. Assim, uma pessoa pode contrair o vírus da dengue por quatro vezes durante toda a vida.

Uma vez contraído o vírus, seja por qualquer um dos sorotipos, a pessoa pode apresentar sintomas de uma doença febril, conhecida como a Dengue Clássica. Casos mais extremos, como hemorragia e/ou choque hipovolêmico, podem ocorrer. Esses casos são chamados de Febre Hemorrágica de Dengue ou Dengue Hemorrágica (DALBEM et al., 2014). Entretanto, a doença pode apresentar poucos sintomas, como uma febre indiferenciada, ou, em alguns casos, nenhuma manifestação conhecida (CATÃO, 2011). Na Figura 2.2, é apresentada, sistematicamente, as principais formas de sintomas clínicos que a doença apresenta.

Figura 2.2 – Formas clínicas da doença.



Fonte: Adaptado de Catão (2011)

De acordo com Catão (2011), a Dengue Clássica pode apresentar diversos sintomas, tais como febre, dores de cabeça, nas articulações, nos músculos, atrás dos olhos, vômitos, coceira na pele e, até mesmo, algumas hemorragias, porém não tão comuns. Já a Febre Hemorrágica da Dengue (FHD) pode apresentar manifestações hemorrágicas, que variam dentro de uma escala de quatro graus, sendo que, a FHD Grau III e a FHD Grau IV são as mais graves, podendo levar a pessoa infectada a óbito.

Conhecer o comportamento da doença em relação a sua dinâmica espacial é de grande importância, a fim de que se possa avaliar e identificar as áreas em que o risco populacional seja elevado. Uma forma de se obter tais informações é através de ferramentas e técnicas de geoprocessamento e análise estatística voltada para o estudo espacial de dados. Desta forma, informações desta natureza podem ser utilizadas, de forma eficaz, pelos serviços de saúde para o combate e controle do mosquito (SCANDAR, 2007).

## **2.2 Estatística Espacial**

A Estatística é uma ciência bastante importante em diversas áreas do conhecimento. Sua principal ferramenta são os dados. Ela é uma ciência que lida, basicamente, com a coleta, análise e interpretação de dados, cuja finalidade é fornecer informações relevantes para que se possa tomar decisões diante de alguma situação. Seus métodos e técnicas permitem com que outras áreas do conhecimento possam compreender o comportamento de determinados fenômenos presentes no cotidiano.

Todo dado é gerado em um determinado momento no tempo e em uma certa localização no espaço, onde o fenômeno de interesse atua. Assim, no processo de coleta desses dados, as informações referentes à posição temporal e espacial estão atrelados aos dados. Em algumas situações, tais informações podem ser importantes para a análise, seja ela puramente espacial ou temporal ou, até mesmo, ambas as informações.

Uma das áreas da Estatística que integram informações de referências espaciais à análise de dados é a Estatística Espacial, cujo seu desenvolvimento passou a ganhar força durante a segunda metade do século XX (DIGGLE et al., 2013). Ela pode ser compreendida como um conjunto de métodos, conceitos e técnicas que considera, explicitamente, as coordenadas espaciais, por exemplo latitude e longitude, relevantes para a análise, podendo, assim, ser modelada como um processo estocástico (CRESSIE, 1993).

Um processo estocástico pode ser definido como uma coleção de variáveis aleatórias, indexadas a um conjunto (ROSS, 1996). Esse conjunto, também chamado de "conjunto de indexação", pode ser unidimensional (geralmente relacionado ao tempo) ou  $d$ -dimensional (relacionado às coordenadas espaciais).

Segundo Gelfand et al. (2010), a Estatística Espacial pode ser definida como uma coleção de variáveis aleatórias indexadas por um conjunto  $\mathbf{s} \in \mathbf{S}$  correspondente à coordenadas no espaço  $d$ -dimensional, ou seja, é um processo estocástico espacial, definido como

$$\{Y(\mathbf{s}) : \mathbf{s} \in \mathbf{S} \subseteq \mathbb{R}^d\}. \quad (2.1)$$

Seu principal objetivo se encontra em descrever o comportamento de algum fenômeno, como, por exemplo, a presença (ou ausência) de dependência espacial. Através dos dados, esse comportamento pode ser caracterizado pelos efeitos de primeira ordem (relacionado ao valor médio do processo) e de segunda ordem (relacionado à dependência espacial).

Com o advento da computação e o grande desenvolvimento de métodos, alguns autores como Cressie (1993) sugerem que a Estatística Espacial pode ser dividida em diversas áreas, tais como geostatística, análise de dados de áreas, processos pontuais, etc. Com isso, a estatística espacial passou a ser uma das áreas mais importante da estatística com aplicação em diversos campos do conhecimento científico, tais como a Biologia, Economia, Agronomia, Epidemiologia, dentre outras.

Neste trabalho, será abordada uma análise sob a perspectiva de processos pontuais. Sua teoria será explicitada na próxima seção. Primeiramente, será exposta uma abordagem puramente espacial e temporal. Posteriormente, a teoria será estendida para o espaço-tempo, interesse de estudo neste trabalho.

### 2.3 Processos pontuais espaciais

Os processos pontuais espaciais focam, basicamente, no estudo da distribuição espacial de eventos observados em uma determinada região de estudo. Segundo Waller e Gotway (2004), o processo pontual é descrito por um processo estocástico no qual cada variável aleatória representa a localização de um determinado evento no espaço  $d$ -dimensional, ou seja, é uma coleção de eventos aleatórios  $\{(\mathbf{s}_i) : i = 1, 2, \dots\}$ , em que  $\mathbf{s}_i \in \mathbf{S} \subseteq \mathbb{R}^d$ . A realização desse processo gera um conjunto finito de eventos de localizações em uma região de estudo  $A \subset \mathbf{S} \subseteq \mathbb{R}^d$ , resultantes

de um modelo particular. Para o caso bidimensional, isto é,  $d = 2$ , tem-se que  $\mathbf{s}_i$  representa a localização de um evento definido no plano cartesiano  $\mathbb{R}^2$ .

A finalidade de um estudo relacionado à processos pontuais está em descrever o comportamento do processo estocástico no qual gerou tais eventos. Assim como uma variável aleatória pode ser descrita por meio de duas medidas estatísticas (esperança e variância), os processos estocásticos também podem ser descritos, de forma análoga, por meio de duas propriedades. Deste modo, inferências acerca do processo estocástico gerador são realizadas, a fim de caracteriza-lo em função das propriedades de primeira e segunda ordem (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2008).

### 2.3.1 Propriedades de primeira ordem

As propriedades de primeira ordem se baseiam na mensuração do valor médio do processo em uma região espacial e são descritas pela *intensidade de primeira ordem* do processo pontual, definida por:

$$\lambda(\mathbf{s}) = \lim_{|ds| \rightarrow 0} \frac{\mathbb{E}[N(ds)]}{|ds|}, \quad (2.2)$$

em que  $N(ds)$  representa o número de eventos em uma pequena região contendo o evento  $\mathbf{s}$  e  $|ds|$  representa a área dessa região.

A intensidade de primeira ordem pode ser interpretada como o número médio de eventos por unidade de área (DIGGLE, 2013). Sua intensidade pode ser constante (também chamada de homogênea ou uniforme) ou pode variar sobre o espaço (também chamada de não homogênea ou não uniforme). Quando um processo é considerado não homogêneo, a intensidade de primeira ordem é denominada *função intensidade*.

Assim, é necessário analisar o comportamento da intensidade de uma configuração pontual. Existem várias formas de se estimar esta intensidade, seja via métodos paramétricos (um modelo de regressão, por exemplo) ou não. O método mais utilizado na literatura, utiliza um estimador não paramétrico baseado em densidade de kernel bivariada, que será apresentada a seguir.

#### 2.3.1.1 Estimador da intensidade por kernel

Um método não paramétrico bastante utilizado na prática para se estimar a função intensidade,  $\lambda(\mathbf{s})$ , é o método de alisamento de kernel, proposto por Diggle (1985). Este método

se baseia em estimar a intensidade em toda a região espacial em estudo, obtendo-se, assim, uma superfície contínua. Seu estimador é dado pela seguinte expressão:

$$\hat{\lambda}(\mathbf{s}_0) = \frac{1}{\delta(\mathbf{s}_0)} \sum_{i=1}^n \frac{1}{h^2} k\left(\frac{\mathbf{s}_0 - \mathbf{s}_i}{h}\right), \quad (2.3)$$

em que  $\mathbf{s}_0$  é um ponto espacial o qual deseja-se estimar a intensidade,  $\mathbf{s}_i$  representa a localização espacial dos  $n$  eventos observados,  $k(\cdot)$  é a função kernel,  $h$  é descrito como uma largura de banda que controla o alisamento da kernel na superfície espacial e  $\delta(\mathbf{s}_0)$  é um fator para correção de efeito de borda.

Para a escolha de possíveis pontos  $\mathbf{s}_0$  para a estimação da intensidade espacial, busca-se estabelecer uma malha regular sobre a região em estudo, com pontos igualmente espaçados entre si. Para uma análise visual gráfica, quanto mais fina for a malha, melhor será a resolução da superfície estimada.

Em relação à função kernel,  $k(\cdot)$ , ela assume a forma de uma função densidade de probabilidade para a construção da intensidade. De acordo com Schabenberger e Gotway (2004), esta função pode apresentar várias formas, tais como a kernel Gaussiana

$$k(p) = \frac{1}{\sqrt{2\pi}} \exp\{-p^2/2\},$$

a kernel quadrática

$$k(p) = \frac{3}{4}(1 - p^2) \quad \text{para} \quad |p| \leq 1,$$

ou a kernel de variância mínima

$$k(p) = \frac{3}{8}(3 - 5p^2) \quad \text{para} \quad |p| \leq 1.$$

No entanto, segundo Diggle (2013), a escolha da função kernel pouco influencia nas estimativas, se comparado com a escolha da largura de banda,  $h$ . Valores altos para  $h$  geram estimativas de intensidade mais homogêneas, ao passo que valores mais baixos para  $h$  apresentam estimativas mais heterogêneas. Segundo Waller e Gotway (2004), o analista pode determinar, com alguma subjetividade, o valor de  $h$  que melhor representa o comportamento da intensidade do processo.

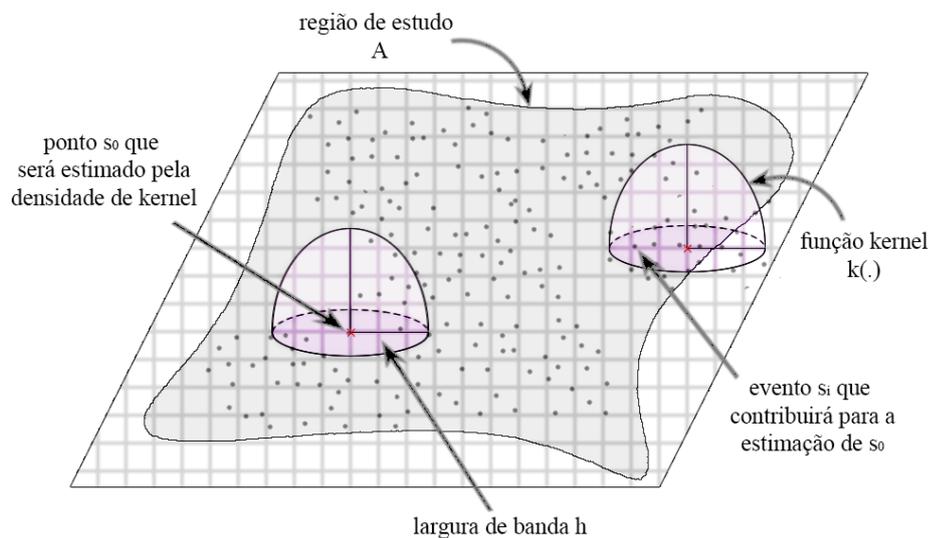
Em relação aos efeitos de borda, a estimativa da intensidade em localizações próximas às delimitações da região em estudo podem ser viesadas, devido ao fato de não considerar

eventos que possam ter ocorrido fora desta região. Uma forma de contornar esse problema, é estabelecer um fator que corrige esse viés. Um fator bastante utilizado na literatura, proposto por Diggle (1985), é aquele que considera a proporção do volume da kernel que se encontra dentro da região espacial, que é expresso por

$$\delta(s_0) = \int_A \frac{1}{h^2} k\left(\frac{s_0 - \mathbf{u}}{h}\right) d\mathbf{u}.$$

A Figura 2.3 apresenta uma esquematização sobre o funcionamento do estimador de kernel. No entanto, a figura ilustra a estimação da kernel em apenas dois pontos na região espacial.

Figura 2.3 – Ilustração gráfica do estimador da intensidade por kernel.



Fonte: Do autor (2017).

### 2.3.2 Propriedades de segunda ordem

As propriedades de segunda ordem se preocupam em identificar a existência de interações entre eventos e como elas ocorrem, provenientes de uma estrutura de correlação espacial. Basicamente, uma configuração pontual pode apresentar interações de agrupamento ou regularidade.

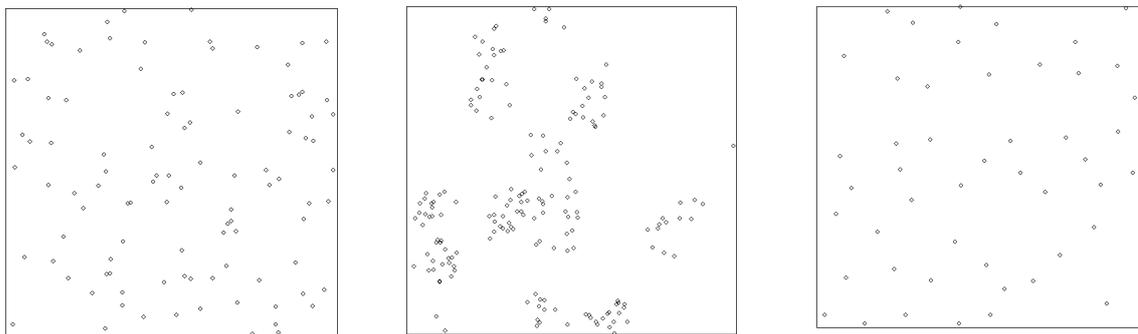
No entanto, um processo pontual pode não apresentar nenhum grau de relação entre eventos. Nesse caso, a configuração pontual apresenta uma estrutura de *completa aleatoriedade espacial*. Este comportamento define uma situação em que os eventos são equiprováveis e independentes, ou seja, a realização de um evento gerado em alguma localização não aumenta

e nem diminui a probabilidade de outro evento ser gerado próximo a ele. Esse modelo pode ser utilizado como forma de comparação entre eventos que exibem algum tipo de interação espacial (WALLER; GOTWAY, 2004).

Em interações que exibem agrupamentos, os eventos tendem a estar mais próximos uns dos outros em relação ao que é esperado por um processo sem dependência espacial (completa aleatoriedade espacial). Neste caso, a realização de um evento aumenta a chance de ocorrer outros eventos próximos a ele.

Já em interações que apresentam regularidade, esses eventos tendem a estar mais distantes uns dos outros em relação ao que é esperado segundo uma configuração que exibe uma completa aleatoriedade espacial. Assim, a ocorrência de um evento na região espacial diminui a probabilidade de ocorrer outros eventos próximos a ele. Na Figura 2.4, são apresentados possíveis cenários de configurações que apresentam as interações descritas anteriormente.

Figura 2.4 – Padrões de interação entre eventos.



(a) Completa aleatoriedade espacial

(b) Agrupamento

(c) Regularidade

Fonte: Do autor (2017).

As propriedades de segunda ordem são, formalmente, descritas pela intensidade de segunda ordem do processo pontual, definida por

$$\lambda_2(s_i, s_j) = \lim_{|ds_i| \rightarrow 0, |ds_j| \rightarrow 0} \frac{\mathbb{E}[N(ds_i)N(ds_j)]}{|ds_i| |ds_j|}, \quad (2.4)$$

em que  $\mathbb{E}[N(ds_i)N(ds_j)]$  representa o número esperado de eventos dentro de pequenos círculos de área  $|ds_i|$  e  $|ds_j|$ , respectivamente.

Segundo Sherman (2011), a intensidade de segunda ordem pode ser informalmente interpretada como a chance de se observar um evento em ambas as localizações  $s_i$  e  $s_j$ .

### 2.3.2.1 Estacionariedade e isotropia

Em propriedades de segunda ordem, algumas suposições são impostas a fim de facilitar o processo de análise. Tais suposições são de estacionariedade e isotropia.

O conceito de estacionariedade está atrelado à ideia de que o processo é invariante sobre translação, ou seja, sua intensidade de primeira ordem não varia sobre a região no espaço  $d$ -dimensional e que a relação entre dois eventos depende somente da distância relativa entre eles. Assim,  $\lambda(\mathbf{s}) = \lambda$  e  $\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\|\mathbf{s}_i - \mathbf{s}_j\|)$ , sendo  $\|\cdot\|$  a distância euclidiana entre dois eventos no plano.

Além de estacionário, a configuração pontual também será isotrópica quando a estrutura de dependência espacial entre os eventos for invariante sobre rotação (WALLER; GOTWAY, 2004). Nesse caso, a relação entre eventos dependerá somente de sua distância, independente de suas localizações e direções no espaço.

Assim, sob a suposição de isotropia e estacionariedade, tem-se que:

$$\lambda(\mathbf{s}) = \lambda \quad (2.5)$$

e

$$\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\|\mathbf{s}_i - \mathbf{s}_j\|) = \lambda_2(u). \quad (2.6)$$

### 2.3.2.2 Função de correlação entre pares espacial

Outras medidas estatísticas podem ser utilizadas, de forma equivalente, para descrever as propriedades de segunda ordem, tal como a *função de correlação entre pares de eventos* (ou Função  $g$ ), definida como

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

Se os eventos são gerados por um processo completamente aleatório (independente), a intensidade de segunda ordem é fatorada no produto de duas intensidades de primeira ordem, avaliadas nas localizações  $\mathbf{s}_i$  e  $\mathbf{s}_j$  respectivamente. Desta forma, sob a suposição de completa aleatoriedade espacial, a função de correlação entre pares de eventos se reduz a

$$\begin{aligned}
g(\mathbf{s}_i, \mathbf{s}_j) &= \frac{\lambda_2(\mathbf{s}_i, \mathbf{s}_j)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \\
&= \frac{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \\
&= 1.
\end{aligned} \tag{2.8}$$

Considerando o processo pontual estacionário e isotrópico, a equação (2.7) pode ser reescrita como

$$g(u) = \frac{\lambda_2(u)}{\lambda^2}, \tag{2.9}$$

em que  $u = \|\mathbf{s}_i - \mathbf{s}_j\|$  é a distância euclidiana entre os dois eventos.

De acordo com Baddeley, Rubak e Turner (2015), a função de correlação entre pares (ou Função  $g$ ) pode ser informalmente interpretada como a probabilidade de se observar um par de eventos separados por uma distância  $u$ , dividido pela probabilidade correspondente ao caso de completa aleatoriedade espacial.

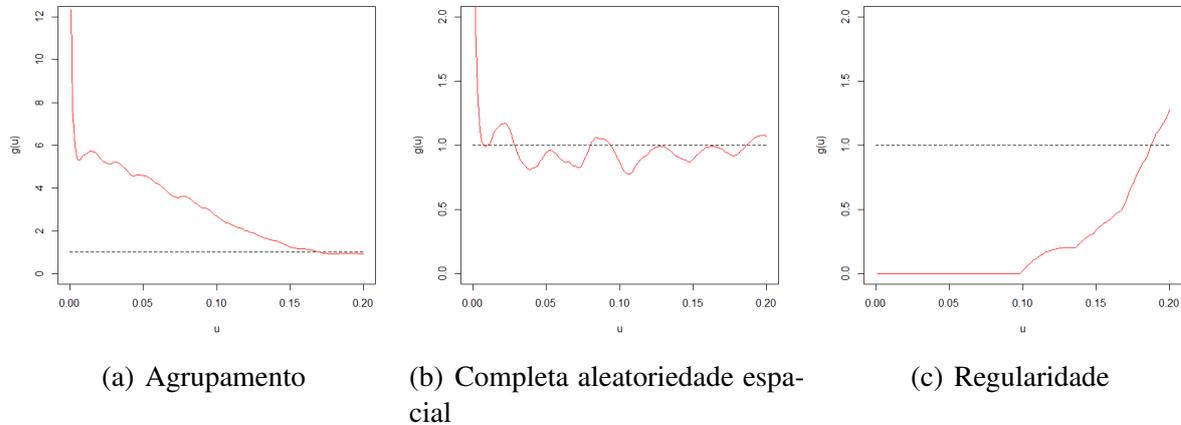
Um estimador para a Função  $g$ , proposto por Stoyan e Stoyan (1994), sob a suposição de estacionariedade e isotropia, é dado por

$$\hat{g}(u) = \frac{|W|}{2\pi un^2} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} K_h(u - \|\mathbf{s}_i - \mathbf{s}_j\|), \tag{2.10}$$

em que  $|W|$  é a área da região espacial,  $n$  é o número de eventos observados,  $K_h(\cdot)$  é um estimador de kernel e  $w_{ij}$  é o efeito de correção de borda espacial.

O método de análise consiste em confrontar a estimativa da Função  $g$ , obtida pela equação (2.10) através dos dados observados, com sua respectiva função teórica (para o caso de completa aleatoriedade espacial), obtida pela equação (2.8). De acordo com Baddeley, Rubak e Turner (2015), se  $\hat{g}(u) > 1$ , há indícios de que a configuração pontual observada exibe um padrão de agrupamentos de eventos. Por outro lado, se  $\hat{g}(u) < 1$ , uma repulsão entre eventos pode ser o padrão de interação observado. No entanto, se a configuração pontual foi gerada de forma completamente aleatória sobre o espaço, espera-se que a estimativa da Função  $g$  seja aproximadamente igual a 1. A Figura 2.5 apresenta uma representação gráfica do método de análise da Função  $g$  para os três cenários possíveis.

Figura 2.5 – Gráfico de análise da Função  $g$  espacial comparando sua estimativa (em vermelho) com a função teórica (linha tracejada) para os três cenários possíveis.



Fonte: Do autor (2017).

No entanto, alguns fenômenos não obedecem a suposição de estacionariedade de primeira ordem, ou seja, de que o processo pontual possui intensidade homogênea. Desta forma, Baddeley, Møller e Waagepetersen (2000) apresenta um estimador para a Função  $g$  para um processo estacionário com intensidade de segunda ordem reponderada. Este estimador permite que o processo pontual tenha uma intensidade não homogênea. Assim, o estimador para a Função  $g$ , com intensidade de primeira ordem não homogênea, é dado por

$$\hat{g}(u) = \frac{1}{|W|} \frac{1}{2\pi u} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{K(u - \|\mathbf{s}_i - \mathbf{s}_j\|)}{\lambda(\mathbf{s}_i) \lambda(\mathbf{s}_j)}, \quad (2.11)$$

em que  $\lambda(\cdot)$  é a estimativa da intensidade avaliada em algum ponto no espaço.

Neste caso, o processo de análise da Função  $g$ , definido em (2.11), é o mesmo em relação ao caso homogêneo. Se  $\hat{g}(u) > 1$ , o processo apresenta um padrão de agrupamentos ao passo que, se  $\hat{g}(u) < 1$  os eventos apresentarão interações de regularidade.

### 2.3.2.3 Função $K$ espacial

Uma alternativa, proposta por Ripley (1976, 1977), para descrever propriedades de segunda ordem do processo é a Função  $K$ . Segundo Cressie (1993), sua finalidade está em capturar a dependência espacial a partir de diferentes escalas de distâncias entre eventos de uma configuração pontual. Para tanto, os conceitos de estacionariedade e isotropia serão impostos como suposições para análise dessa função. O objetivo é que a Função  $K$  explique a dependência

especial somente pelas distâncias entre eventos e não dependa de suas localizações relativas e sua direção. Assim, a Função  $K$  é definida por:

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

em que  $\lambda$  denota a intensidade homogênea do processo pontual e  $\mathbb{E}[N(u)]$  o valor esperado do número de eventos dentro de uma distância  $u$  em torno de um evento arbitrário.

Outra forma de definir a Função  $K$ , segundo Diggle (2013), é em termos da função de correlação entre pares de eventos, definida anteriormente em (2.7). Logo,

$$K(u) = 2\pi \int_0^u g(x) x dx. \quad (2.13)$$

Sobre a suposição de independência e estacionariedade, o que representa um caso de completa aleatoriedade espacial, tem-se que a Função  $K$  é dada por:

$$\begin{aligned} K(u) &= 2\pi \int_0^u g(s_i, s_j) x dx \\ &= 2\pi \int_0^u \frac{\lambda_2(s_i, s_j)}{\lambda(s_i)\lambda(s_j)} x dx && \text{(independência)} \\ &= 2\pi \int_0^u \frac{\lambda(s_i)\lambda(s_j)}{\lambda(s_i)\lambda(s_j)} x dx && \text{(estacionariedade)} \\ &= 2\pi \int_0^u \frac{\lambda^2}{\lambda^2} x dx \\ &= 2\pi \int_0^u x dx \\ &= \pi u^2. \end{aligned} \quad (2.14)$$

Um estimador para a Função  $K$  é dado por:

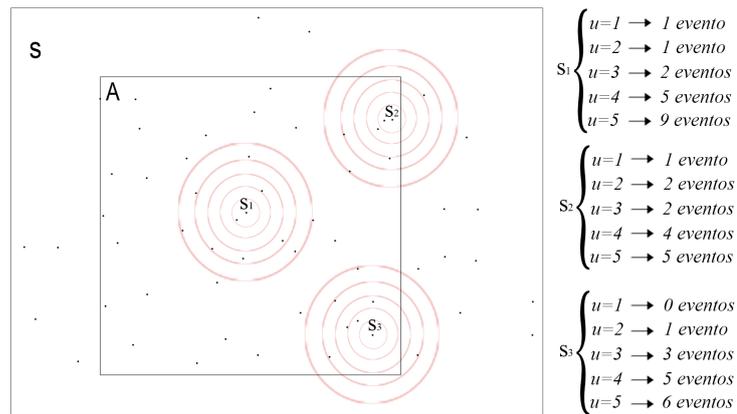
$$\hat{K}(u) = \hat{\lambda}^{-1} \sum_{i=1}^n \sum_{j \neq i} \frac{I(\|\mathbf{s}_i - \mathbf{s}_j\| \leq u)}{n-1}, \quad (2.15)$$

em que  $\hat{\lambda}$  é um estimador não viesado da intensidade homogênea  $\lambda$ , definido por  $\hat{\lambda} = \frac{n}{|A|}$ ;  $n$  é o número de eventos em uma região com área  $|A|$  e  $I(\cdot)$  é a função que indica se há, ou não, evento com distância menor ou igual a  $u$ .

A ideia do estimador da Função  $K$  é estimar o número médio de eventos dentro de um círculo de raio  $u$  centrado no evento  $\mathbf{s}_i$  e dividi-lo pelo estimador da intensidade homogênea

do processo. A Figura 2.6 apresenta a forma como a Função  $K$  é estimada para diferentes distâncias  $u$ . Para fins didáticos, o exemplo mostra a forma de estimação somente para três eventos. Porém, a ideia é que o procedimento seja feito para cada  $s_i \in A$ . As circunferências em torno dos eventos  $s_i$  representam as diversas distâncias espaciais a serem utilizadas a fim de que a dependência espacial possa ser obtida em diferentes escalas.

Figura 2.6 – Exemplificação do processo de estimação da Função  $K$  espacial para três eventos.



Fonte: Do autor (2017).

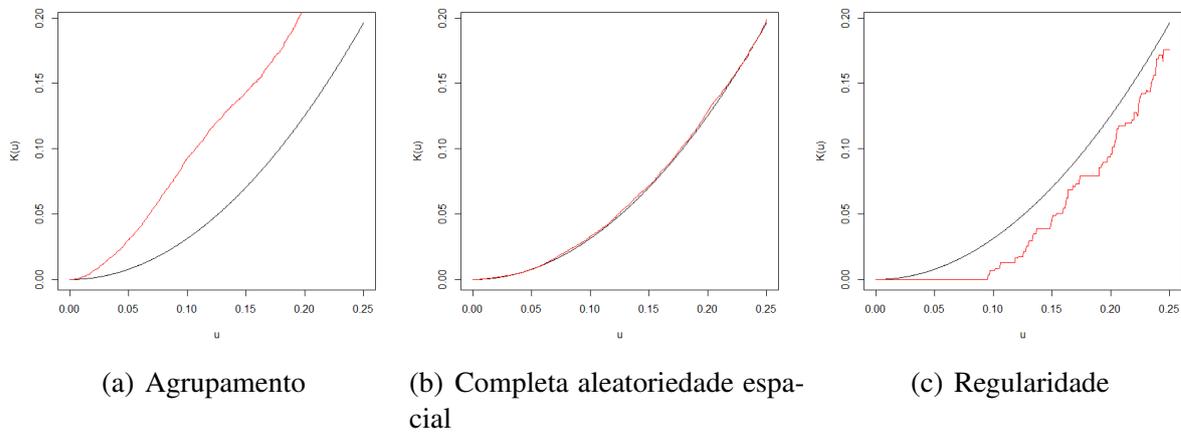
No entanto, o estimador apresentado em (2.15) é viesado devido ao problema de efeito de borda, uma vez que existem eventos próximos às delimitações da região de estudo, tais como os eventos  $s_2$  e  $s_3$  apresentados no exemplo da Figura 2.6. Segundo Waller e Gotway (2004), para uma distância  $u$  maior que a distância entre a borda e o evento mais próximo a ela, a contagem do número de eventos dentro de um raio  $u$  será menor devido à não contabilização dos possíveis eventos que estão fora da região de análise. De acordo com Cressie (1993), existem pelo menos três propostas para correção desses efeitos de borda. A mais utilizada é o efeito de correção de borda proposto por Ripley, que consiste em ponderar o estimador pela proporção da circunferência, de raio  $u$ , que se encontra dentro da região de estudo  $A$ . Logo, o estimador da Função  $K$  corrigindo este efeito de borda é dado por:

$$\hat{K}(u) = \hat{\lambda}^{-1} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{I(\|s_i - s_j\| \leq u)}{n-1}, \quad (2.16)$$

em que  $w_{ij}$  é a proporção da área da circunferência que se encontra dentro da região  $A$ , centrado em um evento arbitrário.

A forma de análise da Função  $K$  está em confrontar a função estimada aos dados a partir do estimador apresentado em (2.16),  $\hat{K}(u)$ , com a função teórica,  $K(u)$ , dado pela equação (2.14). Portanto, conclui-se que, se  $\hat{K}(u) > K(u)$  a configuração pontual tende a formar grupos de eventos, apresentando uma interação de agrupamentos entre eventos. No entanto, se  $\hat{K}(u) < K(u)$  os eventos desta configuração tendem a estar mais distantes uns dos outros, exibindo, assim, um comportamento de regularidade. A Figura 2.7 apresenta o gráfico das Funções  $K$  empírica (curva em vermelho) e teórica (curva em preto) para os três possíveis cenários.

Figura 2.7 – Gráfico da Função  $K$  espacial comparando sua estimativa (em vermelho) com a função teórica (suposição de completa aleatoriedade espacial) para os três cenários possíveis.



Fonte: Do autor (2017).

## 2.4 Processos pontuais temporais

Um processo pontual no tempo pode ser definido como um processo estocástico, cuja a realização consiste em uma coleção finita de eventos indexados em algum instante de tempo,  $\{(t_i) : i = 1, 2, \dots, n\}$ , tal que  $t_i \in \mathbb{R}^+$  (BRILLINGER; GUTTORP; SCHOENBERG, 2002).

Assim como no caso espacial, um processo pontual indexado no tempo pode ser caracterizado pelas propriedades de primeira e segunda ordem. Assim, a intensidade do processo, descrito como o número médio de eventos por unidade de tempo, caracteriza as propriedades de primeira ordem do processo pontual e, segundo Cox e Isham (1980), é definida por:

$$\lambda(t) = \lim_{|dt| \rightarrow 0} \frac{\mathbb{E}[N(dt)]}{|dt|}, \quad (2.17)$$

em que  $\mathbb{E}[N(dt)]$  representa o número esperado de eventos dentro de um pequeno intervalo de tempo  $dt$ .

De forma análoga, considera-se a intensidade de segunda ordem como a estatística que caracteriza as propriedades de segunda ordem de um processo no tempo, definida por

$$\lambda_2(t_i, t_j) = \lim_{|dt_i| \rightarrow 0, |dt_j| \rightarrow 0} \frac{\mathbb{E}[N(dt_i)N(dt_j)]}{|dt_i| |dt_j|}, \quad (2.18)$$

em que  $N(dt_i)$  e  $N(dt_j)$  representam o número de eventos dentro de pequenos intervalos de tempo  $|dt_i|$  e  $|dt_j|$ , respectivamente.

### 2.4.1 Função $K$ temporal

A fim de verificar o comportamento da dependência temporal entre eventos, Diggle et al. (1995) apresentam uma adaptação da Função  $K$ , proposto por Ripley(1976,1977), para o caso temporal, definindo-a como

$$K(v) = \lambda^{-1} E[N(v)], \quad (2.19)$$

em que  $v$  representa um intervalo temporal e  $E[N(v)]$  o número de eventos esperados a partir de uma distância  $v$  em torno de um evento arbitrário.

Sobre a suposição de completa aleatoriedade temporal, ou seja, assumindo estacionariedade e independência, tem-se que a Função  $K$  no tempo é dada por

$$K(v) = 2v. \quad (2.20)$$

Um estimador para a Função  $K$  no tempo, proposto por Diggle et al. (1995), é dado pela seguinte expressão:

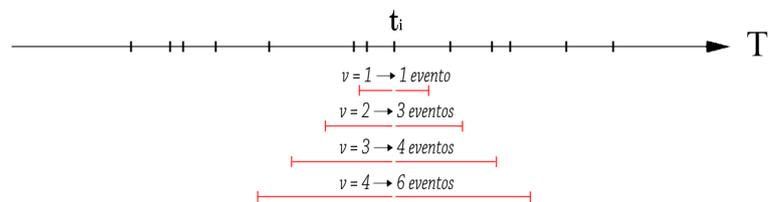
$$\hat{K}(v) = \hat{\lambda}^{-1} \sum_{i=1}^n \sum_{i \neq j} \frac{1}{r_{ij}} \frac{I(|t_i - t_j| \leq v)}{n-1}, \quad (2.21)$$

em que  $\hat{\lambda}$  é um estimador não viesado considerando uma intensidade temporal homogênea, definido por  $\hat{\lambda} = \frac{n}{T}$ ;  $n$  é o número de eventos gerados no intervalo  $T$ ,  $I(\cdot)$  é a função que indica se há, ou não, evento com distância menor ou igual a  $v$ , e  $r_{ij}$  denota um fator de correção de borda unidimensional, análogo a correção proposta por Ripley. Assim,  $r_{ij} = 1$  se ambas

as extremidades do intervalo, de comprimento  $2v$ , centrado em  $t_i$  estiver dentro do intervalo temporal em estudo, caso contrário  $r_{ij} = 1/2$ .

A forma de estimação da Função  $K$ , para o caso temporal, é apresentada na Figura 2.8. Basicamente, o estimador calcula o número esperado de eventos dentro de uma distância  $v$  centrada em um evento arbitrário,  $t_i$ , e divide pela intensidade homogênea do processo. A Figura 2.8 apresenta a ideia de estimação apenas para um único evento. No entanto, o estimador realiza o mesmo procedimento para todos os eventos contidos dentro do intervalo temporal considerado na análise.

Figura 2.8 – Exemplificação do método de estimação da Função  $K$  temporal para um evento  $t_i$ .



Fonte: Do autor (2017).

A forma de análise da Função  $K$  consiste em comparar, em diferentes escalas de distâncias  $v$ , sua estimativa,  $\hat{K}(v)$ , com sua função teórica, expressa em (2.20). Logo, se  $\hat{K}(v) > 2v$ , então existem evidências de um padrão de agrupamentos de eventos no tempo, ao passo que, se  $\hat{K}(v) < 2v$ , há indícios de que os eventos exibem uma interação de regularidade temporal.

## 2.5 Processos pontuais espaço-temporal

Os métodos estudados em processos pontuais espaciais e temporais levam em consideração somente análises de fenômenos cujo o interesse é, essencialmente, descrever sua distribuição espacial e temporal, respectivamente. Em algumas situações, esses fenômenos podem ser melhor estudados considerando, conjuntamente, ambas as componentes. Para entender melhor o comportamento de alguns desses fenômenos é necessário analisar simultaneamente sua variação no espaço e evolução no tempo. Esses métodos são denominados de processos pontuais espaço-temporais.

Um processo pontual espaço-temporal descreve um processo estocástico no qual cada variável aleatória representa a localização de um evento em uma região no espaço e em algum momento no tempo. A realização do processo estocástico consiste em um conjunto contável de pontos,  $\{(s_i, t_i), i = 1, 2, \dots, n\}$ , no qual  $s_i \in \mathbf{S} \subset \mathbb{R}^2$  é a localização do  $i$ -ésimo evento em uma região espacial  $d$ -dimensional e  $t_i \in T \subset \mathbb{R}^+$  é o tempo no qual ocorreu este evento (DIGGLE, 2013; GABRIEL; ROWLINGSON; DIGGLE, 2013).

Assim como em processos pontuais indexados no espaço ou no tempo, a análise espaço-temporal pode ser caracterizada pelas propriedades de primeira e segunda ordem.

### 2.5.1 Propriedades de primeira ordem

As propriedades de primeira ordem descrevem a intensidade do processo, definida como o número esperado de eventos por unidade de área e por unidade de tempo. Assim,

$$\lambda(\mathbf{s}, t) = \lim_{|ds| \rightarrow 0, |dt| \rightarrow 0} \frac{\mathbb{E}[N(ds, dt)]}{|ds| |dt|}, \quad (2.22)$$

em que  $\mathbb{E}[N(ds, dt)]$  representa o número esperado de eventos em um cilindro infinitesimal com base  $ds$  e altura  $dt$ .

Segundo Gabriel, Rowlingson e Diggle (2013), o processo pontual pode ser estacionário:

- no espaço, se a intensidade variar somente no tempo:  $\lambda(\mathbf{s}, t) = \lambda(t)$ ;
- no tempo, se a intensidade variar somente no espaço:  $\lambda(\mathbf{s}, t) = \lambda(\mathbf{s})$ ;
- no tempo e no espaço, se a intensidade for constante:  $\lambda(\mathbf{s}, t) = \lambda$ .

### 2.5.2 Propriedades de segunda ordem

As propriedades de segunda ordem também podem ser descritas em termos da intensidade de segunda ordem do processo pontual estendida para o espaço-tempo, de tal forma que

$$\lambda_2((s_i, t_i)(s_j, t_j)) = \lim_{|D_i|, |D_j| \rightarrow 0} \frac{\mathbb{E}[N(D_i)N(D_j)]}{|D_i| |D_j|}, \quad (2.23)$$

em que  $D_i = ds_i \times dt_i$  e  $D_j = ds_j \times dt_j$  são pequenos cilindros contendo os eventos  $(s_i, t_i)$  e  $(s_j, t_j)$ , respectivamente.

### 2.5.2.1 Função de correlação entre pares espaço-temporal

A função de correlação entre pares de eventos (ou Função  $g$ ) pode, também, ser entendida para o espaço-tempo como uma das medidas descritoras das propriedades de segunda ordem. De forma equivalente a (2.7), esta medida é definida, respectivamente, por

$$g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \frac{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}. \quad (2.24)$$

Sob a suposição de completa aleatoriedade no espaço e no tempo, tem-se que a Função  $g$  espaço-temporal se reduz a:

$$\begin{aligned} g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \frac{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \\ &= \frac{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \\ &= 1. \end{aligned} \quad (2.25)$$

Supondo um processo estacionário e isotrópico, um estimador proposto para esta medida, definida na equação (2.24), é dado por

$$\hat{g}(u, v) = \frac{|\mathbf{S} \times T|}{4 \pi u n^2} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij} r_{ij}} K_s(u - \|\mathbf{s}_i - \mathbf{s}_j\|) K_t(v - |t_i - t_j|), \quad (2.26)$$

em que  $K_s(\cdot)$  e  $K_t(\cdot)$  são funções kernel com larguras de banda  $h_s$  e  $h_t$ , respectivamente.  $w_{ij}$  e  $r_{ij}$  descrevem os efeitos de correção de bordas espacial e temporal, respectivamente.

Se o processo for não homogêneo, um estimador para a Função  $g$ , apresentado por Gabriel, Rowlingson e Diggle (2013), é dado por

$$\hat{g}(u, v) = \frac{1}{|\mathbf{S} \times T|} \frac{1}{4 \pi u} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij} r_{ij}} \frac{K_s(u - \|\mathbf{s}_i - \mathbf{s}_j\|) K_t(v - |t_i - t_j|)}{\lambda(\mathbf{s}_i, t_i) \lambda(\mathbf{s}_j, t_j)}, \quad (2.27)$$

em que  $\lambda(\cdot, \cdot)$  é a estimativa da intensidade de primeira ordem avaliada em algum ponto da região espaço-temporal.

De forma análoga ao caso espacial, o método de análise da Função  $g$  consiste em confrontar sua estimativa, obtida por um estimador empírico através dos eventos observados na região espaço-temporal, com o caso em que o processo é completamente aleatório, obtida

pela função teórica apresentado em (2.25). Desta forma, se  $\hat{g}(u, v) > 1$ , há indícios de que a configuração pontual apresenta agrupamentos de eventos no espaço-tempo ao passo que, se  $0 < \hat{g}(u, v) < 1$ , os eventos podem estar apresentando um padrão de repulsão, tanto no tempo quanto no espaço.

Outro estimador, apresentado por Diggle, Rowlingson e Su (2005), para a Função  $g$  espacial com média no tempo, considerando um comportamento não homogêneo na intensidade, é dado por

$$\hat{g}(u) = \frac{1}{|\mathbf{S}|} \frac{1}{2\pi u} \sum_{t=1}^T \frac{1}{T} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{K_s(u - \|\mathbf{s}_i - \mathbf{s}_j\|)}{\lambda_t(\mathbf{s}_i) \lambda_t(\mathbf{s}_j)}, \quad (2.28)$$

em que  $\lambda_t(\mathbf{s}_i)$  é definido como um estimador para a intensidade espacial em um tempo  $t$  fixo,  $|\mathbf{S}|$  descreve a área da região espacial em estudo,  $K_s(\cdot)$  é um estimador de Kernel espacial e  $w_{ij}$  é o efeito de correção de borda espacial.

### 2.5.2.2 Função $K$ espaço-temporal

Diggle et al. (1995) propôs uma extensão da Função  $K$  clássica para um processo pontual no espaço-tempo, definida como

$$K(u, v) = \lambda^{-1} \mathbb{E}[N(u, v)], \quad (2.29)$$

em que  $u = \|\mathbf{s}_i - \mathbf{s}_j\|$  representa a distância euclidiana entre dois eventos na região espacial,  $v = |t_i - t_j|$  representa um intervalo temporal e  $\mathbb{E}[N(u, v)]$  representa o número de eventos esperados em uma distância  $u$  e  $v$  a partir de um evento arbitrário.

Definindo a Função  $K$  espaço-tempo em termos da função de correlação entre pares de eventos, tem-se que

$$K(u, v) = 2\pi \int_{-v}^v \int_0^u g(x, y) dx dy. \quad (2.30)$$

Sobre a suposição de completa aleatoriedade espaço-temporal, tem-se que a função teórica para a Função  $K$  espaço-temporal é

$$\begin{aligned}
K(u, v) &= 2\pi \int_{-v}^v \int_0^u g(s_i, s_j) x dx dy \\
&= 2\pi \int_{-v}^v \int_0^u \frac{\lambda_2((s_i, t_i), (s_j, t_j))}{\lambda(s_i, t_i) \lambda(s_j, t_j)} x dx dy && \text{(independência)} \\
&= 2\pi \int_{-v}^v \int_0^u \frac{\lambda(s_i, t_i) \lambda(s_j, t_j)}{\lambda(s_i, t_i) \lambda(s_j, t_j)} x dx dy && \text{(estacionariedade)} \\
&= 2\pi \int_{-v}^v \int_0^u \frac{\lambda^2}{\lambda^2} x dx dy && (2.31) \\
&= 2\pi \int_{-v}^v \int_0^u x dx dy \\
&= \pi u^2 \int_{-v}^v dy \\
&= 2\pi u^2 v.
\end{aligned}$$

Segundo Diggle et al. (1995), um estimador para a Função  $K$  espaço-temporal homogênea, corrigindo o efeito de borda no tempo e no espaço, é dado por:

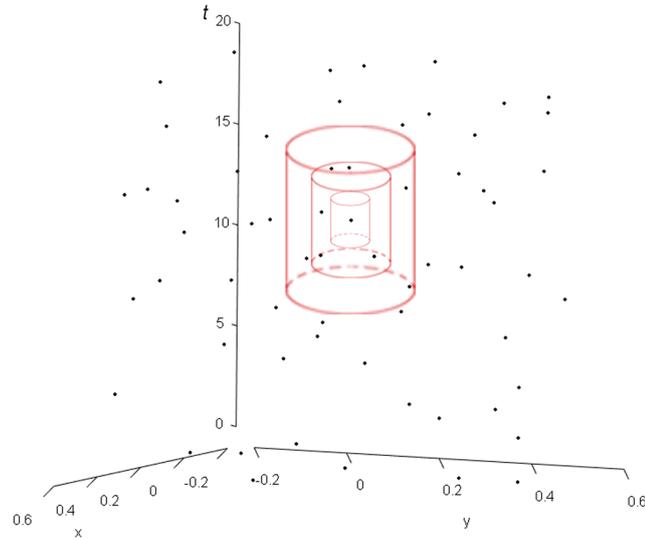
$$\hat{K}(u, v) = \hat{\lambda}^{-1} \sum_{i=1}^n \sum_{i \neq j} \frac{1}{w_{ij}} \frac{1}{r_{ij}} \frac{I(\| \mathbf{s}_i - \mathbf{s}_j \| \leq u) I(| t_i - t_j | \leq v)}{n-1}, \quad (2.32)$$

em que  $\hat{\lambda}$  é um estimador não viesado da intensidade homogênea  $\lambda$ , definido por  $\hat{\lambda} = \frac{n}{|\mathbf{S} \times \mathcal{T}|}$ ;  $n$  é o número de eventos gerados na região  $\mathbf{S} \times T$ ,  $I(\cdot)$  é a função que indica se há, ou não, evento com distância menor ou igual a  $u$  e  $v$ , e  $w_{ij}$  e  $r_{ij}$  denotam a correção de borda no espaço e no tempo, respectivamente.

O processo de estimação da Função  $K$  espaço-temporal é análogo ao caso espacial e temporal, que consiste em mensurar o número médio de eventos dentro de uma distância espacial  $u$  e uma distância temporal  $v$ , construindo, assim, um cilindro em torno de um evento arbitrário. Assim, por meio dessas estimativas é possível identificar a interação existente entre eventos no espaço-tempo a partir de diferentes escalas espaciais e temporais.

A Figura 2.9 apresenta, esquematicamente, a forma de estimação da Função  $K$ . No exemplo, a ideia é apresentada apenas para um evento em  $\mathbf{S} \times T$ , definido em  $\mathbb{R}^2 \times \mathbb{R}^+$ . No entanto, o procedimento deve ser realizado a todos os eventos gerados dentro da região espaço-temporal em estudo. Nesta ilustração, os diferentes cilindros em torno do evento representam o aumento das distâncias espaciais,  $u$ , e temporais,  $v$ , a fim de capturar a estrutura de dependência em diferentes escalas.

Figura 2.9 – Exemplificação do método de estimação da Função  $K$  espaço-temporal para um evento.



Fonte: Do autor (2017).

Assim, se a diferença entre os valores de  $\hat{K}(u, v)$  e a função teórica, em (2.31), for maior que zero, então há evidências que indicam um comportamento de agrupamento de eventos no tempo e no espaço, ao passo que, se  $\hat{K}(u, v) - 2\pi u^2 v$  apresentar valores negativos, há indícios de que a configuração pontual exibe uma interação de regularidade no espaço-tempo (GABRIEL; ROWLINGSON; DIGGLE, 2013).

Um estimador para a Função  $K$  para o espaço-tempo, proposto por Gabriel e Diggle (2009), considerando que o processo pontual seja não estacionário de primeira ordem, isto é, que sua intensidade de primeira ordem apresente não homogeneidade, é dado por

$$\hat{K}(u, v) = \frac{1}{|\mathbf{S} \times T|} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{1}{r_{ij}} \frac{I(\|\mathbf{s}_i - \mathbf{s}_j\| \leq u) I(|t_i - t_j| \leq v)}{\lambda(\mathbf{s}_i, t_i) \lambda(\mathbf{s}_j, t_j)}, \quad (2.33)$$

em que  $\lambda(.,.)$  é a estimativa da intensidade de primeira ordem avaliada em algum ponto da região espaço-temporal  $|\mathbf{S} \times T|$ .

Um caso particular ao seu estimador espaço-temporal é uma proposta para o caso espacial com tempo médio, considerando, também, não homogeneidade na intensidade de primeira ordem. Esse estimador para a Função  $K$ , apresentado por Davies e Hazelton (2013), é definido como

$$\hat{K}(u) = \frac{1}{|S|} \sum_{t=1}^T \frac{1}{T} \sum_{i=1}^n \sum_{i \neq j} \frac{1}{w_{ij}} \frac{I(\|\mathbf{s}_i - \mathbf{s}_j\| \leq u)}{\lambda_t(\mathbf{s}_i) \lambda_t(\mathbf{s}_j)} \quad (2.34)$$

em que  $\lambda_t(\cdot)$  é a estimativa da intensidade espacial de primeira ordem avaliada no tempo  $t$ .

## 2.6 Campos Aleatórios Gaussianos

A abordagem geoestatística foi desenvolvida para descrever fenômenos o qual variam sobre espaço e tempo contínuos, baseados em um conjunto de observações dispersos sobre uma determinada região espaço-temporal  $\mathbf{S} \times T$ . Esta abordagem é baseada na teoria de processos estocásticos sendo que, no contexto geoestatístico, é também denominado de campos aleatórios ou funções aleatórias (FINKENSTÄDT; HELD; ISHAM, 2006).

De acordo com Silva, Ribeiro Jr e Elmatzoglou (2007), um campo aleatório espaço-temporal é definido como um vetor aleatório, isto é, um conjunto de variáveis aleatórias, no qual cada uma delas está definida em uma localização espacial e em um instante de tempo. Deste modo, define-se um campo aleatório como

$$\mathbf{Z} = \{Z(\mathbf{s}, t), \mathbf{s} \in \mathbf{S} \subseteq \mathbb{R}^2, t \in \mathbf{T} \subseteq \mathbb{R}^+\}, \quad (2.35)$$

em que  $Z(\mathbf{s}, t)$  é a medida da variável aleatória  $Z$  indexada na localização espacial  $\mathbf{s}$  e no instante de tempo  $t$ .

Segundo Chilès e Delfiner (2009), um campo aleatório espaço-temporal será Gaussiano se, para qualquer subconjunto finito do vetor aleatório  $\mathbf{Z}$ , definido em (2.35), a distribuição conjunta for uma Gaussiana multivariada.

Uma vez que o processo estocástico o qual atua na região espaço-temporal é Gaussiano, o fenômeno pode ser completamente caracterizado pelos momentos de primeira e segunda ordem, isto é, pela média e pela estrutura de covariância (CHILÈS; DELFINER, 2009). Assim, define-se as funções média, variância e covariância como:

$$\mathbb{E}[Z(\mathbf{s}, t)] = \mu(\mathbf{s}, t), \quad (2.36)$$

$$\text{Var}[Z(\mathbf{s}, t)] = \sigma_Z^2(\mathbf{s}, t) \quad (2.37)$$

$$\begin{aligned} \text{Cov}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)] &= \text{Cov}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j) \\ &= \sqrt{\sigma_Z^2(\mathbf{s}_i, t_i)} \sqrt{\sigma_Z^2(\mathbf{s}_j, t_j)} \text{Corr}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)], \end{aligned} \quad (2.38)$$

em que  $\text{Corr}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)]$  descreve a função de correlação entre  $Z(\mathbf{s}_i, t_i)$  e  $Z(\mathbf{s}_j, t_j)$ .

Ainda, segundo Chilès e Delfiner (2009), quando o campo aleatório apresenta apenas uma única realização, como é comum na prática, sua distribuição não pode ser determinada sem assumir algumas suposições, tais como a estacionariedade e isotropia.

O processo estocástico espaço-temporal é considerado estritamente estacionário se a sua distribuição de dimensão finita for invariante sob translação, isto é, dado um vetor de distâncias espaciais e temporais  $(\mathbf{u}, v)$ , os processos  $\mathbf{Z} = \{Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n)\}$  e  $\mathbf{Z}_{\mathbf{u}, v} = \{Z(\mathbf{s}_1 + \mathbf{u}, t_1 + v), \dots, Z(\mathbf{s}_n + \mathbf{u}, t_n + v)\}$  possuem a mesma distribuição (CHILÈS; DELFINER, 2009).

De acordo com Chilès e Delfiner (2009), uma suposição mais fraca de estacionariedade impõe que o processo será fracamente estacionário (ou estacionário de segunda ordem) se os momentos de primeira e segunda ordem forem constantes, finitos e se a covariância (ou correlação) entre  $\mathbf{Z}$  e  $\mathbf{Z}_{\mathbf{u}, v}$  depender somente das distâncias espaciais e temporais,  $\mathbf{u}$  e  $v$ , respectivamente.

Além de estacionário, o processo será isotrópico sobre o espaço se sua estrutura de dependência for invariante sobre rotação, isto é, para qualquer direção no espaço a estrutura de dependência espacial não se altera (CHILÈS; DELFINER, 2009).

O processo Gaussiano pode assumir, além das suposições de estacionariedade e isotropia, uma estrutura de separabilidade na covariância espaço-temporal. Desta forma, a função que descreve o comportamento da dependência espaço-temporal do processo pode ser fatorada como um produto de duas covariâncias, uma puramente espacial e outra temporal. Deste modo, sob a suposição de estacionariedade, isotropia e separabilidade entre espaço e tempo, as funções (2.36), (2.37) e (2.38) podem ser reescritas como

$$\mathbb{E}[Z(\mathbf{s}, t)] = \mu, \quad (2.39)$$

$$\text{Var}[Z(\mathbf{s}, t)] = \sigma_Z^2 \quad (2.40)$$

$$\begin{aligned} \text{Cov}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)] &= \sigma_Z^2 \text{Corr}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)] \\ &= \sigma_Z^2 c_1(u; \phi) c_2(v; \theta) \end{aligned} \quad (2.41)$$

em que  $c_1(\cdot)$  e  $c_2(\cdot)$  descrevem as funções de correlação estacionárias e isotrópicas para o caso espacial e temporal, controladas pelos parâmetros  $\phi$  e  $\theta$ , respectivamente.

De acordo com Finkenstädt, Held e Isham (2006), uma exigência estabelecida para uma função de covariância de um campo aleatório é que ela seja positiva definida, de tal forma que

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)] \geq 0, \quad (2.42)$$

para todos os  $n$  inteiros positivos, de qualquer conjunto de pontos  $(\mathbf{s}_i, t_i) \in \mathbb{R}^2 \times \mathbb{R}^+$  (MØLLER; SYVERSVEN; WAAGEPETERSEN, 1998). Considerar uma estrutura de covariância espaço-temporal separável para  $Z(\mathbf{s}, t)$  não é natural, uma vez que a informação de interação entre tempo e espaço é perdida (DIGGLE, 2013). No entanto, validar uma função de covariância como positiva definida não é uma tarefa fácil. Assim, a suposição de separabilidade é utilizada uma vez que o produto entre duas funções de covariâncias válidas para o caso temporal e espacial irá gerar uma função válida de covariância espaço-temporal. Na prática, são escolhidas funções que são conhecidas por serem válidas (DIGGLE et al., 2013).

Uma das classes de funções bastante utilizadas para modelar correlações espaciais isotrópicas, o qual são consideradas positivas definidas, é a classe de Matérn. Assim, define-se esta função de correlação como

$$c(d) = \frac{(\phi d)^\nu}{2^{\nu-1} \Gamma(\nu)} K_\nu(\phi d), \quad \nu > 0, \phi > 0, \quad (2.43)$$

em que  $d$  descreve a distância entre os pontos,  $K_\nu$  representa a função Bessel modificada do segundo tipo e de ordem  $\nu$ ,  $\Gamma$  é a função gama e  $\phi$  é o parâmetro do modelo que descreve o alcance da dependência, isto é, a distância máxima em que se observa uma estrutura de dependência entre pontos.

A equação definida em (2.43) é uma forma generalizada das funções Matérn e podem tomar diferentes formas a medida que  $\nu$  é modificada (SCHABENBERGER; GOTWAY, 2004). Desta forma, se  $\nu \rightarrow \infty$ , a função de correlação resultante assume a forma

$$c(d) = \exp\{-\phi d^2\}, \quad (2.44)$$

o qual é conhecido como modelo gaussiano.

De acordo com Schabenberger e Gotway (2004), outra relação importante e que assume a forma de outro modelo de correlação bastante utilizado é quando  $\nu$  é igual a  $1/2$ . Assumindo os seguintes resultados da função Bessel

$$K_\nu(r) = \frac{\pi I_{-\nu}(r) - I_\nu(r)}{2 \sin(\pi\nu)},$$

em que  $I_{-1/2}(r) = \sqrt{\frac{2}{\pi r}} \cosh(r)$ ,  $I_{1/2}(r) = \sqrt{\frac{2}{\pi r}} \sinh(r)$ ,  $\sinh(r) = \frac{1}{2}(e^r - e^{-r})$  e  $\cosh(r) = e^{-r} + \sinh(r)$ , a função Bessel modificada assume a forma

$$K_{1/2}(r) = \sqrt{\frac{\pi}{2r}} e^{-r}.$$

Assim, quando  $\nu = 1/2$ , a equação (2.43) apresenta a seguinte forma

$$c(d) = \exp\{-\phi d\}, \quad (2.45)$$

o qual é descrito como modelo exponencial.

Segundo Chilès e Delfiner (2009), a classe de funções Matérn está definida em  $\mathbb{R}^n$ , para qualquer  $n \in \mathbb{N}$ . Assim, tais modelos de correlação podem ser utilizados para modelar tanto estruturas de dependência espacial quanto temporal. Em particular, o modelo de correlação exponencial, definido em  $\mathbb{R}^1$ , apresenta uma estrutura de correlação autorregressiva de primeira ordem em uma série temporal (SCHABENBERGER; GOTWAY, 2004).

## 2.7 Modelos espaço-temporais

### 2.7.1 Processo de Poisson

#### 2.7.1.1 Processo de Poisson homogêneo

O processo de Poisson homogêneo é considerado o modelo mais simples para processos pontuais estacionários no espaço-tempo. Segundo Gabriel, Rowlingson e Diggle (2013), é um modelo utilizado como forma de explicar uma configuração de eventos que apresenta completa aleatoriedade espacial e temporal, ou seja, o processo de Poisson homogêneo é um modelo no

qual não apresenta efeitos de primeira e segunda ordem. Assim, sua intensidade é considerada invariante em toda a região espaço-temporal  $\mathbf{S} \times T$  e os eventos não apresentam nenhuma estrutura de correlação espaço-temporal. Apesar de ser considerado o mais simples, ele é bastante utilizado como modelo de referência para distinção entre processos que apresentam dependência espaço-temporal. Esse modelo pode ser formalmente definido pelas seguintes propriedades:

1. O número de eventos  $N(\mathbf{S} \times T)$  na região  $\mathbf{S} \times T$  segue uma distribuição de Poisson com média  $\lambda |\mathbf{S} \times T|$ , em que  $\lambda$  denota a intensidade do processo e  $|\cdot|$  é o volume espaço-temporal da região  $\mathbf{S} \times T$ .
2. O número de eventos  $N(\mathbf{S} \times T) = n$  gerados em  $\mathbf{S} \times T$  formam uma amostra aleatória independente e uniformemente distribuído em  $\mathbf{S} \times T$ .

Utilizando o método de máxima verossimilhança para estimar a intensidade de um processo de Poisson homogêneo, tem-se que o logaritmo da função de verossimilhança, baseado em uma única realização do processo com  $n$  eventos observados, é dado por:

$$\ell(\lambda; n) = -\lambda |\mathbf{S} \times T| + n \ln(\lambda |\mathbf{S} \times T|) - \ln(n!). \quad (2.46)$$

A Função Score,  $S(\lambda)$ , é obtida derivando (2.46) em relação à  $\lambda$ :

$$\begin{aligned} S(\lambda) &= \frac{d \ell(\lambda; n)}{d \lambda} \\ &= -|\mathbf{S} \times T| + \frac{n |\mathbf{S} \times T|}{\lambda |\mathbf{S} \times T|} \\ &= \frac{n}{\lambda} - |\mathbf{S} \times T|. \end{aligned} \quad (2.47)$$

Assim, o estimador de máxima verossimilhança para  $\lambda$  é obtido igualando a Função Score, definida em (2.47), a zero. Logo,

$$\hat{\lambda} = \frac{n}{|\mathbf{S} \times T|}. \quad (2.48)$$

### 2.7.1.2 Processo de Poisson não-homogêneo

O processo de Poisson não homogêneo é considerado o modelo mais simples entre todos aqueles que modelam processos pontuais não estacionários no espaço-tempo. Diferentemente

do modelo apresentado anteriormente, o processo de Poisson não homogêneo considera a intensidade do processo variante no tempo e/ou no espaço,  $\lambda(\mathbf{s}, t)$ . No entanto, assim como no caso homogêneo, esse modelo não considera uma estrutura de correlação espacial e/ou temporal entre os eventos gerados pelo processo estocástico atuante na região. De acordo com Diggle (2013), esse modelo pode ser definido pelos seguintes postulados:

1. O número de eventos  $N(\mathbf{S} \times T)$  em uma região  $\mathbf{S} \times T$  segue uma distribuição de Poisson com média  $\int_{\mathbf{S}} \int_T \lambda(\mathbf{s}, t) dt ds$ .
2. O número de eventos gerados em  $\mathbf{S} \times T$ ,  $N(\mathbf{S} \times T) = n$ , formam uma amostra aleatória independente de uma distribuição em  $\mathbf{S} \times T$  com função densidade de probabilidade proporcional à  $\lambda(\mathbf{s}, t)$ .

### 2.7.2 Processo de Cox log-Gaussiano

O processo de Cox espaço-temporal é uma classe de modelos o qual é considerado uma generalização dos modelos de Poisson, particularmente os não homogêneos. Na literatura, esses modelos são denominados por processos duplamente estocásticos devido a sua principal característica que é considerar a intensidade, também, resultante de um mecanismo estocástico, representado por  $\Lambda(\mathbf{s}, t)$  (DIGGLE, 2013). Formalmente, o processo de Cox pode ser definido pelas seguintes propriedades:

1.  $\{\Lambda(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}^+\}$  é uma intensidade estocástica espaço-temporal de valores não negativos.
2. Condicionado à realização do processo,  $\{\Lambda(\mathbf{s}, t) = \lambda(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}^+\}$ , os eventos formam um processo de Poisson não homogêneo com função intensidade  $\lambda(\mathbf{s}, t)$ .

Um processo de Cox log-Gaussiano é um caso particular do modelo de Cox e possui as mesmas propriedades descritas acima. A diferença deste modelo se encontra na forma como a intensidade estocástica é apresentada. Assim, define-se  $\Lambda(\mathbf{s}, t)$  como

$$\Lambda(\mathbf{s}, t) = \exp\{Z(\mathbf{s}, t)\}, \quad (2.49)$$

em que  $Z = \{Z(\mathbf{s}, t); \mathbf{s} \in \mathbf{S}, t \in T\}$  é um campo aleatório Gaussiano com média  $\mu(\mathbf{s}, t) = \mathbb{E}[Z(\mathbf{s}, t)]$ , variância  $\sigma_Z^2(\mathbf{s}, t) = \text{Var}[Z(\mathbf{s}, t)]$  e função de covariância  $\text{Cov}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)]$  (GABRIEL; ROWLINGSON; DIGGLE, 2013).

Assumindo uma estrutura de separabilidade espaço-temporal para a função de correlação de  $Z(\mathbf{s}, t)$ , temos que

$$\begin{aligned} \text{Corr}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)] &= \text{Corr}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] \text{Corr}[Z(t_i), Z(t_j)] \\ &= c_1(\mathbf{s}_i, \mathbf{s}_j; \phi) c_2(t_i, t_j; \theta). \end{aligned} \quad (2.50)$$

em que  $c_1(\cdot)$  e  $c_2(\cdot)$  descrevem as correlações espaciais e temporais, respectivamente, controladas pelos parâmetros  $\phi$  e  $\theta$ .

Considerar uma estrutura de covariância (ou correlação) espaço-temporal separável não é natural, uma vez que a informação de interação entre tempo e espaço é perdida (DIGGLE, 2013). No entanto, esta suposição será essencial para fins de estimação dos parâmetros do modelo via mínimos contrastes.

Trabalhar com um modelo que seja contínuo tanto no tempo quanto no espaço, pode não ser possível uma vez que, em diversas aplicações, os eventos são observados somente em tempo discreto (dias, meses, anos, etc.), como discutido por Davies e Hazelton (2013). Assim, a intensidade estocástica espacial do processo de Cox log-Gaussiano, em um determinado tempo  $t$ , pode ser formalmente expressa por

$$\begin{aligned} \Lambda_t(\mathbf{s}) &= \int_{t_0}^t \exp\{Z(\mathbf{s}, y)\} dy \\ &\approx (t - t_0) \exp\{Z(\mathbf{s}, t)\} \end{aligned} \quad (2.51)$$

em que  $(t - t_0)$  é um pequeno intervalo de tempo.

De acordo com Møller, Syversveen e Waagepetersen (1998), as propriedades do processo pontual de Cox log-Gaussiano estão diretamente ligadas às propriedades do processo Gaussiano  $Z(\mathbf{s}, t)$ . Utilizando  $\zeta_t^{(k)}$  para denotar a  $k$ -ésima ordem da intensidade espacial do processo de Cox log-Gaussiano, no tempo  $t$ , segundo Davies e Hazelton (2013), tem-se que

$$\begin{aligned} \zeta_t^{(k)}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k) &= \Delta t \exp \sum_{i=1}^k \left[ \mu(\mathbf{s}_i, t) + \frac{\sigma_Z^2(\mathbf{s}_i, t)}{2} \right] \\ &+ \mathbb{I}(k > 1) \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sigma_Z(\mathbf{s}_i, t) \sigma_Z(\mathbf{s}_j, t) c_1(\mathbf{s}_i, \mathbf{s}_j; \phi) \end{aligned} \quad (2.52)$$

Assim, a partir da equação (2.52), é possível obter a intensidade de primeira ordem do processo de Cox, dada por

$$\zeta_t^{(1)}(\mathbf{s}) = \Delta t \exp \left\{ \mu(\mathbf{s}, t) + \frac{\sigma_Z^2(\mathbf{s}, t)}{2} \right\}, \quad (2.53)$$

e, também, a intensidade de segunda ordem, dada por

$$\zeta_t^{(2)}(\mathbf{s}_i, \mathbf{s}_j) = \Delta t \exp \left\{ \mu(\mathbf{s}_i, t) + \mu(\mathbf{s}_j, t) + \frac{\sigma_Z^2(\mathbf{s}_i, t) + \sigma_Z^2(\mathbf{s}_j, t)}{2} + \sigma_Z(\mathbf{s}_i, t) \sigma_Z(\mathbf{s}_j, t) c_1(\mathbf{s}_i, \mathbf{s}_j; \phi) \right\}. \quad (2.54)$$

Assumindo que  $Z(\mathbf{s}, t)$  seja estacionário e isotrópico, isto é, o processo é invariante sobre translação e rotação na região espaço-temporal, tem-se que  $Z$  será descrito por uma média e variância constantes e uma função de covariância separável e que dependerá somente das distâncias espaciais e temporais. Assim,

$$\mathbb{E}[Z(\mathbf{s}, t)] = \mu,$$

$$\text{Var}[Z(\mathbf{s}, t)] = \sigma_Z^2$$

e

$$\text{Cov}[Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)] = \sigma_Z^2 c(\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|) = \sigma_Z^2 c(u, v) = \sigma_Z^2 c_1(u; \phi) c_2(v; \theta).$$

Segundo Davies e Hazelton (2013), impor condições de estacionariedade e isotropia ao processo Gaussiano  $Z(\mathbf{s}, t)$ , implica que o processo pontual de Cox log-Gaussiano também será estacionário e isotrópico. Logo, sobre essas condições e assumindo um intervalo de tempo unitário,  $\Delta t = 1$ , as equações (2.53) e (2.54) podem ser reescritas como

$$\zeta_t^{(1)} = \exp \left\{ \mu + \frac{\sigma_Z^2}{2} \right\}, \quad (2.55)$$

e

$$\begin{aligned}
\zeta_t^{(2)}(u) &= \exp\{2\mu + \sigma_Z^2 + \sigma_Z^2 c_1(u; \phi)\} \\
&= \exp\{2\mu + \sigma_Z^2\} \exp\{\sigma_Z^2 c_1(u; \phi)\} \\
&= (\zeta_t^{(1)})^2 \exp\{\sigma_Z^2 c_1(u; \phi)\}.
\end{aligned} \tag{2.56}$$

Portanto, utilizando a relação apresentada na equação (2.7), a função de correlação entre pares de eventos no espaço para o processo de Cox log-Gaussiano é definido por

$$\begin{aligned}
g_t(u) &= \frac{\zeta_t^{(2)}(u)}{(\zeta_t^{(1)})^2} \\
&= \frac{(\zeta_t^{(1)})^2 \exp\{\sigma_Z^2 c_1(u; \phi)\}}{(\zeta_t^{(1)})^2} \\
&= \exp\{\sigma_Z^2 c_1(u; \phi)\},
\end{aligned} \tag{2.57}$$

A equação (2.49), considerando estruturas de estacionariedade e isotropia para o processo Gaussiano,  $Z(\mathbf{s}, t)$ , apresenta uma estrutura de intensidade para o processo de Cox log-Gaussiano cuja distribuição dos eventos é homogênea sobre toda a região espaço-temporal (DAVIES; HAZELTON, 2013). No entanto, em diversas aplicações, como na epidemiologia, esse tipo de comportamento pode não ocorrer. Uma forma de contornar este problema, é incluir na estrutura do modelo componentes fixas que expressam a não homogeneidade contida no fenômeno.

De acordo com Diggle (2013), uma parametrização conveniente para expressar a intensidade estocástica do processo de Cox é dada por

$$\Lambda_t(\mathbf{s}) = \lambda_t(\mathbf{s}) R_t(\mathbf{s}), \tag{2.58}$$

em que  $\lambda_t(\mathbf{s})$  é a componente fixa do modelo, que descreve as propriedades de primeira ordem da intensidade estocástica, ou seja, a média do processo  $\Lambda_t(\mathbf{s})$ . Já  $R_t(\mathbf{s})$  é denominada componente residual que descreve um processo estocástico estacionário com esperança igual à 1 e função de covariância estacionária e isotrópica.

A intensidade estocástica pode, também, apresentar uma estrutura de separabilidade espaço-temporal. Logo,  $\Lambda_t(\mathbf{s})$  é dita ser separável de primeira ordem se

$$\lambda_t(\mathbf{s}) = \rho(\mathbf{s}) m(t), \quad (2.59)$$

em que  $\rho(\mathbf{s})$  descreve uma densidade do processo pontual puramente espacial, de tal forma que  $\int_A \rho(\mathbf{s}) d\mathbf{s} = 1$ . Já,  $m(t)$  representa a intensidade do processo no tempo  $t$ .

Um modelo proposto por Diggle, Rowlingson e Su (2005) assume que a intensidade estocástica do processo de Cox espaço-temporal pode ser expressa como a equação em (2.58). No entanto, o modelo assume separabilidade entre espaço e tempo de primeira ordem, o qual  $\lambda_t(\mathbf{s})$  é descrito como na equação (2.59). Logo, o modelo assume a forma:

$$\Lambda_t(\mathbf{s}) = \rho(\mathbf{s}) m(t) R_t(\mathbf{s}), \quad (2.60)$$

em que  $\rho(\mathbf{s})$  descreve a variação espacial na intensidade,  $m(t)$  apresenta a variação no número médio de eventos por unidade de tempo  $t$ .

O termo  $R_t(\mathbf{s})$  será modelado como  $\exp\{Z_t(\mathbf{s})\}$ , em que  $Z_t(\mathbf{s})$  um processo estocástico Gaussiano estacionário, isotrópico e com estrutura de correlação espaço-temporal separável, o qual representa uma variação residual em  $\Lambda$  não explicada por  $\lambda(\mathbf{s})$  e  $m(t)$ . No entanto, a média de  $Z$  será expressa em termos de sua variância, tal que  $\mathbb{E}[Z_t(\mathbf{s})] = \mu = \frac{-\sigma_Z^2}{2}$ , a fim de que  $\mathbb{E}[R_t(\mathbf{s})] = 1$ . Assim, de acordo com Davies e Hazelton (2013), para cada tempo  $t \in \{0, 1, \dots, T\}$ , as equações (2.55) e (2.56), que descrevem as propriedades de primeira e segunda ordem, respectivamente, podem ser reescritas como

$$\zeta_t^{(1)}(\mathbf{s}) = \mathbb{E}[\Lambda_t(\mathbf{s})] = \lambda_t(\mathbf{s}) = \rho(\mathbf{s}) m(t), \quad (2.61)$$

e

$$\zeta_t^{(2)}(\mathbf{s}_i, \mathbf{s}_j) = \rho(\mathbf{s}_i) \rho(\mathbf{s}_j) m(t)^2 \exp\{\sigma_Z^2 c_1(u; \phi)\}. \quad (2.62)$$

Logo, a função de correlação entre pares de eventos no espaço, para um determinado tempo  $t$ , pode ser expressa como

$$g_t(u) = \exp\{\sigma_Z^2 c_1(u; \phi)\}, \quad (2.63)$$

o qual depende do modelo de correlação espacial do processo Gaussiano,  $c_1(u; \phi)$ .

### 2.7.2.1 Estimação do modelo

Existem várias formas de se estimar as componentes espacial e temporal do modelo, seja ela por modelos paramétricos ou não. Por exemplo, Diggle, Rowlingson e Su (2005) utilizaram o método de alisamento de kernel, como apresentado anteriormente na subseção 2.3.1.1, para obter uma estimativa da componente fixa espacial,  $\rho(\mathbf{s})$ .

Já a estimação da componente fixa temporal do modelo pode ser realizada utilizando métodos paramétricos. Um método é considerar a teoria de modelos lineares generalizados (MLG) utilizando uma regressão de Poisson, de tal forma que

$$\log(m(t)) = f(t) \beta, \quad (2.64)$$

em que  $m(t)$  é considerada a média da variável resposta do modelo com distribuição de Poisson,  $t$  representa um vetor de tempos,  $\beta$  é um vetor de parâmetros,  $\log(\cdot)$  é a função de ligação do modelo o qual garante que  $m(t) > 0$  e  $f(t)$  é uma função qualquer utilizada para modelar possíveis comportamentos de sazonalidade ao longo do tempo.

Em relação a componente residual, o objetivo é obter estimativas dos parâmetros de covariância espaço-temporal de  $Z(\mathbf{s}, t)$ , a partir de um modelo paramétrico. Porém, essas funções devem garantir a suposição de serem positivas-definidas. Alguns desses modelos foram apresentados na seção sobre campos aleatórios Gaussianos. No entanto, o processo de estimação, para o caso espacial, é feito por meio da Função  $g$ , definida em (2.63), que está diretamente ligada a estrutura de covariância do processo Gaussiano  $Z$ .

### 2.7.2.2 Método de mínimos contrastes

Um dos métodos utilizados na literatura para estimar os parâmetros dos modelos de covariância de  $Z(\mathbf{s}, t)$  é o método de mínimos contrastes (DAVIES; HAZELTON, 2013). Essa técnica, apresentada por Brix e Diggle (2001), tem por objetivo minimizar a diferença entre descritores empíricos e teóricos das propriedades de segunda ordem do processo de Cox log-Gaussiano.

Para a estimação dos parâmetros espaciais do processo, será considerada uma função  $J$  qualquer, o qual descreve as propriedades de segunda ordem do processo de Cox log-Gaussiano. Esta função  $J$  poderá ser representada pela função de correlação entre pares,  $g(u)$ , ou pela Função  $K$ ,  $K(u)$ . Seja também  $\hat{J}$  uma estimativa não paramétrica de  $J$ . Assim, para o caso

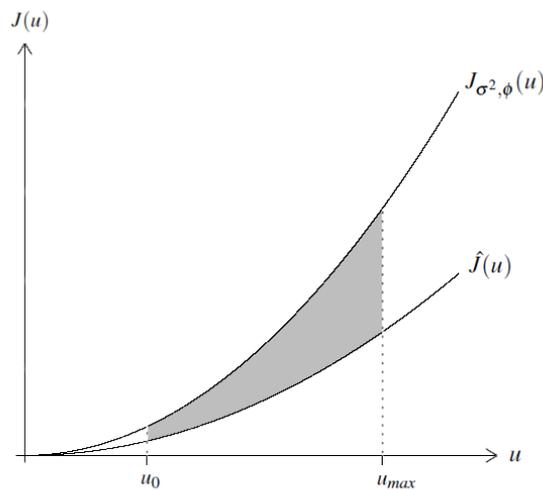
especial, as estimativas para  $\{\sigma^2, \phi\}$  serão encontradas através da minimização do critério de contraste

$$D_J(\sigma^2, \phi) = \int_{u_0}^{u_{max}} w(u) \{v[\hat{J}(u)] - v[J_{\sigma^2, \phi}(u)]\}^2 du, \quad (2.65)$$

em que  $u_0$  e  $u_{max}$  são os limites inferior e superior do critério de contraste,  $w(u)$  representa alguma ponderação associada a cada distância espacial  $u$  e  $v[.]$  é alguma transformação feita em  $J$  e  $\hat{J}$ .

A Figura 2.10 apresenta uma ilustração sobre a forma de estimação do método de mínimos contrastes. A ideia é encontrar estimativas para  $\{\sigma^2, \phi\}$  que minimizem a área ao quadrado entre as curvas das duas funções descritoras das propriedades de segunda ordem do processo de Cox log-Gaussiano, limitado ao intervalo  $[u_0, u_{max}]$ .

Figura 2.10 – Ilustração gráfica em relação ao método de estimação via Mínimos Contrastes.



Fonte: Do autor (2017).

Segundo Taylor et al. (2015), a integral apresentada na equação (2.65) pode ser aproximada, na prática, por

$$D_J(\sigma^2, \phi) \approx u_{diff}^{-1} \sum_{u \in U} w(u) \{v[\hat{J}(u)] - v[J_{\sigma^2, \phi}(u)]\}^2 \quad (2.66)$$

em que  $u_{diff}$  representa a diferença entre dois consecutivos termos em  $U = \{u_0, u_1, \dots, u_{max}\}$ .

O estimador da função de correlação entre pares de eventos para o caso espacial com tempo médio, considerando não homogeneidade na intensidade de primeira ordem, é definido na equação (2.28).

De forma alternativa, a Função  $K$  pode ser utilizada para descrever a dependência espacial do processo de Cox log-Gaussiano. Seu estimador para o caso espacial com tempo médio, considerando, também, não homogeneidade na intensidade de primeira ordem, é definido na equação (2.34).

Segundo Davies e Hazelton (2013), a estrutura de dependência no tempo do processo de Cox log-Gaussiano pode ser entendida como uma medida de autocovariância, que expressa a relação entre a frequência esperada de eventos em dois momentos no tempo. Esta medida pode ser definida como

$$\begin{aligned} C(t, v; \theta) &= Cov[N_t, N_{t-v}] \\ &= \mathbb{E}[N_t, N_{t-v}] - \mathbb{E}[N_t] \mathbb{E}[N_{t-v}], \end{aligned} \quad (2.67)$$

sendo  $\mathbb{E}[N_t, N_{t-v}]$  descrito como a intensidade de segunda ordem do processo pontual no tempo e  $\mathbb{E}[N_t] \mathbb{E}[N_{t-v}]$  descrito como o produto das intensidades de primeira ordem avaliadas no tempo  $t$  e  $t - v$  respectivamente.

Definindo a intensidade de segunda ordem espaço-temporal do processo de Cox log-Gaussiano como

$$\zeta^{(2)}((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda(\mathbf{s}_i, t_i) \lambda(\mathbf{s}_j, t_j) \exp\{\sigma_Z^2 c_1(u; \phi) c_2(v; \theta)\}, \quad (2.68)$$

é possível obter sua respectiva intensidade de segunda ordem para o tempo como

$$\begin{aligned} \mathbb{E}[N_t, N_{t-v}] &= \int_S \int_S \zeta^{(2)}((\mathbf{s}, t), (\mathbf{s} - u, t - v)) d\mathbf{s} d\mathbf{s} - u \\ &= \int_S \int_S \lambda(\mathbf{s}, t) \lambda(\mathbf{s} - u, t - v) \exp\{\sigma_Z^2 c_1(u; \phi) c_2(v; \theta)\} d\mathbf{s} d\mathbf{s} - u \\ &= \int_S \int_S \rho(\mathbf{s}) m(t) \rho(\mathbf{s} - u) m(t - v) \exp\{\sigma_Z^2 c_1(u; \phi) c_2(v; \theta)\} d\mathbf{s} d\mathbf{s} - u \\ &= m(t) m(t - v) \int_S \int_S \rho(\mathbf{s}) \rho(\mathbf{s} - u) \exp\{\sigma_Z^2 c_1(u; \phi) c_2(v; \theta)\} d\mathbf{s} d\mathbf{s} - u \end{aligned} \quad (2.69)$$

em que  $m(\cdot)$  expressa a intensidade temporal,  $\rho(\cdot)$  é a densidade espacial do processo de Cox e  $c_1(\cdot)$  e  $c_2(\cdot)$  denotam as funções de correlação espacial e temporal do processo Gaussiano, respectivamente.

Já o produto das intensidades de primeira ordem no tempo pode ser definido como

$$\mathbb{E}[N_t] \mathbb{E}[N_{t-v}] = m(t) m(t-v). \quad (2.70)$$

Desta forma, a equação (2.67), que descreve a função teórica da medida de autocovariância no tempo do processo pontual, pode ser reescrita como

$$C(t, v; \theta) = m(t)m(t-v) \times \left\{ \left[ \int_S \int_S \rho(\mathbf{s})\rho(\mathbf{s}-u) \exp[\sigma^2 c_1(u; \phi) c_2(v; \theta)] ds ds - u \right] - 1 \right\}. \quad (2.71)$$

Seu estimador empírico é dado por

$$\hat{C}(t, v) = N_t N_{t-v} - m(t) m(t-v). \quad (2.72)$$

em que  $N_t$  representa o número total de observações do processo pontual no tempo  $t$  e  $m(t)$  representa a intensidade temporal estimada no tempo  $t$ .

De forma análoga ao caso espacial, a estimação do parâmetro do modelo de covariância temporal pode ser feita minimizando a discrepância quadrática entre as medidas de covariância teórica e empírica, definidas em (2.71) e (2.72) respectivamente, a partir de uma sequência finita de distâncias no tempo. Portanto, o valor de  $\hat{\theta}$  será aquele que minimiza o seguinte critério:

$$D_C(\theta) = \sum_{v=1}^{v_{max}} \sum_{t=v+1}^n \{\hat{C}(t, v) - C(t, v; \theta)\}^2, \quad (2.73)$$

em que  $v_{max}$  é algum valor arbitrário estabelecido para a distância máxima temporal.

É possível observar que a função teórica da covariância temporal é totalmente dependente do parâmetro relacionado a medida de covariância espacial,  $c_1(u; \phi)$ . Portanto, a estimativa do parâmetro espacial deve ser feita primeiramente para, enfim, poder ser incorporado ao procedimento de estimação no tempo via mínimos contrastes (DAVIES; HAZELTON, 2013).

## 2.8 Predição espaço-temporal

Uma vez estimado os parâmetros do processo de Cox log-Gaussiano, o pesquisador talvez possa estar interessado em realizar predições do processo pontual no espaço-tempo. Entende-se por predição, neste contexto, como a identificação de ocorrências localizadas no espaço e no tempo resultantes de altas incidências que podem ser consideradas não usuais (DIGGLE; ROWLINGSON; SU, 2005).

Uma vez que um processo de Cox log-Gaussiano, condicionado à uma realização de sua intensidade estocástica, resulta em um processo de Poisson, este modelo não considera, diretamente, a existência de agrupamentos espaço-temporal entre eventos. No entanto, o objetivo do pesquisador se encontra em descrever a presença de possíveis aglomerações ocorrendo no espaço-tempo. De acordo com Diggle (2013), os eventos podem ser agrupados devido a alguma estrutura de correlação espaço-temporal existente entre eles. Entretanto, uma intensidade não homogênea pode, também, ser considerada uma fonte de agregação de eventos. Assim, segundo Brix e Diggle (2001), o modelo de Cox log-Gaussiano pode responder a existência de aglomerações por meio de regiões espaciais e temporais com alta intensidade.

### 2.8.1 Simulação Monte Carlo via Cadeias de Markov

Para realizar esta predição, é necessário ter conhecimento sobre a distribuição preditiva de  $Z(\mathbf{s}, t)$  e, portanto, de  $R(\mathbf{s}, t)$ , condicionado aos dados observados até um determinado momento  $t$ . No entanto, não há como obter uma fórmula analítica para tal distribuição. Assim, métodos de simulação baseados em Monte Carlo via Cadeias de Markov (MCMC) podem ser utilizados a fim de gerar amostras desta distribuição preditiva.

Uma forma de se obter estas amostras é utilizando o algoritmo de Metropolis-Hastings. Este algoritmo, introduzido através dos estudos de Metropolis et al. (1953) e Hastings (1970), é um dos métodos MCMC mais utilizados na literatura. O método consiste, basicamente, em simular amostras correlacionadas de uma distribuição de interesse baseado em uma cadeia de Markov. À medida que o número de simulações tende ao infinito, a distribuição empírica destas amostras geradas converge para uma distribuição estacionária, de interesse do pesquisador (TAYLOR; DIGGLE, 2014).

Denotando  $X_{t_1:t_n}$  como observações do processo pontual espacial no intervalo de tempo  $t_1$  e  $t_n$  e  $Z_{t_1:t_n}$  o processo Gaussiano no mesmo intervalo de tempo, a distribuição preditiva de  $Z_{t_1:t_n}$  condicionado em  $X_{t_1:t_n}$  é definido como

$$\pi(Z_{t_1:t_n} | X_{t_1:t_n}) \propto \pi(X_{t_1:t_n} | Z_{t_1:t_n}) \pi(Z_{t_1:t_n}). \quad (2.74)$$

em que  $\pi(Z_{t_1:t_n})$  é o campo aleatório Gaussiano, cujos parâmetros devem ser conhecidos ou estimados, como por exemplo, pelo método de mínimos contrastes.

Assim, tomando  $\pi(Z | X)$  como a densidade de interesse e  $Z^{(i)}$  como um estado (ou valor) atual da cadeia na simulação, uma nova amostra candidata,  $Z'$ , será gerada a partir de uma densidade de probabilidade qualquer,  $q(Z^{(i)}, Z') = P(Z' | Z^{(i)})$ , denominada como núcleo de transição, o qual está condicionada ao valor atual da cadeia. Esta nova amostra será aceita com probabilidade

$$\alpha(Z^{(i)}, Z') = \min \left\{ 1, \frac{\pi(Z' | X) q(Z', Z^{(i)})}{\pi(Z^{(i)} | X) q(Z^{(i)}, Z')} \right\}, \quad (2.75)$$

denominada probabilidade de aceitação.

De acordo com Robert e Casella (2009), uma forma de gerar um valor candidato,  $Z'$ , para compor a amostra da distribuição simulada é levar em consideração a amostra atual simulada. A ideia é simular  $Z'$  através de uma abordagem chamada de passeio aleatório, de tal forma que

$$Z' = Z^{(i)} + e \quad (2.76)$$

em que  $e$  é uma perturbação aleatória com uma distribuição  $g$ , de tal forma que  $e$  seja independente de  $Z^{(i)}$ .

Para o caso em que  $e$  segue uma distribuição Normal com média zero e variância  $h^2$ , o núcleo de transição  $q(Z^{(i)}, Z')$  seguirá uma distribuição  $N(Z^{(i)}, h^2)$  (TAYLOR; DIGGLE, 2014).

De acordo com Robert e Casella (2009), o algoritmo de Metropolis-Hastings consiste em realizar os seguintes passos:

- (1) Obtenha um valor inicial para  $Z$ ,  $Z^{(0)}$ ;
- (2) Simule uma amostra  $Z'$  a partir da densidade  $q(Z^{(i)}, Z')$ ;
- (3) Simule um valor  $u$  de uma distribuição uniforme padrão  $U \sim (0, 1)$ ;
- (4) Compare  $u$  com  $\alpha$ :
  - Se  $u \leq \alpha$ , então  $Z^{(i+1)} = Z'$ ;
  - Se  $u > \alpha$ , então  $Z^{(i+1)} = Z^{(i)}$

- (5) Repita os passos (2) ao (4) até que o número de simulações pré-determinadas seja concluída.

Outra forma de se obter estas amostras é utilizando o algoritmo de Langevin-Hastings, também conhecido como algoritmo de Metropolis ajustado por Langevin (utiliza-se a sigla MALA no inglês). Este método, introduzido pelos estudos de Roberts e Rosenthal (1998), é considerado uma modificação do passeio aleatório do algoritmo de Metropolis-Hastings, a fim de se obter uma eficiência maior na simulação das amostras da distribuição empírica (TAYLOR et al., 2013). Deste modo, o passeio aleatório modificado para o algoritmo de Langevin-Hastings é obtido por

$$Z' = Z^{(i)} + \frac{h^2}{2} \nabla \log\{\pi(Z^{(i)} | X)\} + h e, \quad (2.77)$$

em que  $\nabla \log\{\pi(\cdot)\}$  é o gradiente do logaritmo da função  $\pi(\cdot)$  e  $e$  é uma variável aleatória Normal padrão.

Portanto, o núcleo de transição utilizado para este algoritmo é dado por

$$q(Z^{(i)}, Z') = N \left[ Z^{(i)} + \frac{h^2}{2} \nabla \log\{\pi(Z^{(i)} | X)\}, h^2 \right].$$

Para um estudo mais aprofundado sobre os métodos de Metropolis-Hastings e Langevin-Hastings, veja Metropolis et al. (1953), Roberts e Rosenthal (1998), Robert e Casella (2009), Gamerman e Lopes (2006) e Taylor et al. (2013).

### 2.8.2 Probabilidade de excedência

Após feita a simulação, através do algoritmo de Langevin-Hastings, a distribuição preditiva pode ser utilizada para obter uma estimativa aproximada da probabilidade de  $R(\mathbf{s}, t)$ , condicionado aos dados observados do processo pontual, exceder um determinado valor  $\varepsilon$ . Esta aproximação, segundo Davies e Hazelton (2013), pode ser definida como uma proporção dos valores amostrados de  $R(\mathbf{s}, t)$  que excederam um determinado limiar, em alguma localização no espaço. Assim, esta probabilidade pode ser definida como

$$Pr(R(\mathbf{s}, t) > \varepsilon | \{X_{W,1} \cup \dots \cup X_{W,t}\}) \approx \frac{\sum_{i=1}^L I(\hat{R}_i(\mathbf{s}, t) > \varepsilon)}{L}, \quad (2.78)$$

em que  $I(\cdot)$  representa uma função indicadora,  $L$  é o número total de valores amostrados da distribuição preditiva através do algoritmo de Langevin-Hastings.

Segundo Diggle, Rowlingson e Su (2005), o valor do limiar pode ser determinado subjetivamente pelo pesquisador ou relacionando-o com os valores dos parâmetros do modelo ajustado, através de quantis de  $R(\mathbf{s}, t)$ . Assim, para uma determinada probabilidade  $p$ , o valor do limiar pode ser obtido por

$$\varepsilon = \exp\{\mu + \sigma_Z \Phi^{-1}(p)\}, \quad (2.79)$$

em que  $\mu = -0,5 \sigma_Z^2$  e  $\Phi^{-1}(\cdot)$  descreve a função quantil de uma distribuição Normal padrão.

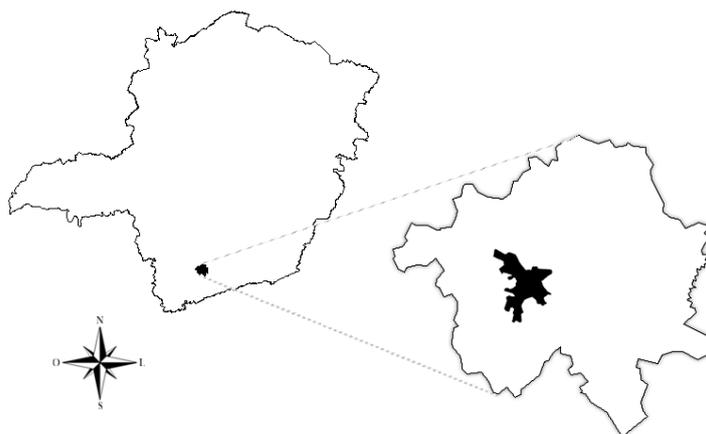
O resultado destas aproximações na probabilidade de excedência pode ser apresentado por meio de mapas espaciais em um determinado tempo  $t$ . De acordo com Davies e Hazelton (2013), estes resultados podem ser utilizados para fins de monitoramento espaço-temporal, permitindo observar o comportamento do fenômeno e, assim, detectar possíveis anomalias que possam se perpetuar ao longo do tempo.

### 3 MATERIAIS E MÉTODOS

#### 3.1 Descrição dos dados

O conjunto de dados a serem analisados neste trabalho são provenientes de casos relacionados às ocorrências de dengue na área urbana do município de Três Corações, situado na região Sul do estado de Minas Gerais, como mostra a Figura 3.1.

Figura 3.1 – Município de Três Corações, localizado no sul de Minas Gerais, com identificação da área urbana no município.



Fonte: Do autor (2017).

O município de Três Corações está localizado a uma latitude de  $21^{\circ}41'33.7''$  sul e longitude  $45^{\circ}15'21.4''$  oeste e possui um território de  $828.038 \text{ km}^2$ . Segundo levantamento do (IBGE, 2016), estima-se que em 2015 sua população era 77.921 habitantes.

As informações referentes aos casos de dengue foram disponibilizadas pela Secretaria de Saúde do município de Três Corações referentes ao período de 01 de janeiro de 2010 à 31 de dezembro 2015.

Para a condução das análises, foram considerados o endereço residencial do paciente e a data de notificação como informações relacionadas à posição espacial e temporal da doença, respectivamente. Tais informações constituem a realização de um evento na região espaço-temporal. Ao todo, foram observados 3.456 casos de dengue no município durante o período de seis anos de análise.

No entanto, alguns casos foram retirados do conjunto de dados devido a falta de informações necessárias para a obtenção de um evento, como, por exemplo, o endereço completo da residência da pessoa que contraiu o vírus ou, então, a data de notificação da doença. Outras

observações também foram desconsideradas, como, por exemplo, casos que ocorreram em zonas rurais do município. A justificativa é que a grande massa populacional se concentra na área urbana e, portanto, poucos casos são observados nessas áreas periféricas. Assim, o conjunto de dados a serem trabalhados nesta dissertação corresponde a um número de 2.818 observações válidas, que se distribuem durante os anos de 2010 a 2015 de acordo com a Tabela 3.1.

Tabela 3.1 – Número de observações de casos de dengue distribuídos entre os anos de 2010 e 2015.

Anos	Número de observações
2010	304
2011	397
2012	51
2013	50
2014	854
2015	1.162

Em relação à informação espacial dos eventos, primeiramente foram obtidas as coordenadas geográficas (latitude e longitude) com auxílio de softwares. No entanto, essas coordenadas representam localizações esféricas e não planas, o que impossibilita cálculos de distâncias euclidianas entre eventos. Para contornar esse problema, foi realizada uma conversão das coordenadas para o sistema UTM (Universal Transversa de Mercator). A unidade de medida espacial a ser considerada neste trabalho será em metros (*m*).

### 3.2 Softwares

Todo o processo de análise e modelagem de dados referentes aos processos pontuais espaço-temporais, discutidos no capítulo anterior, serão desenvolvidos com o auxílio do *software* R (R CORE TEAM, 2016). Para tanto, serão utilizados os seguintes pacotes:

- *lgcp*, desenvolvido por Taylor et al. (2013), cuja finalidade é realizar modelagem de processos pontuais no espaço-tempo utilizando o modelo de Cox log-Gaussiano;
- *RgoogleMaps*, desenvolvido por Loecher e Ropkins (2015), para obter as coordenadas geográficas dos eventos na região espacial em análise;
- *sp*, desenvolvido por Pebesma e Bivand (2005), para a conversão das coordenadas geográficas (latitude e longitude) no formato UTM.

Além do pacote *RgoogleMaps*, também foi utilizado o site Google Maps (GOOGLE, 2016) para a obtenção das coordenadas geográficas.

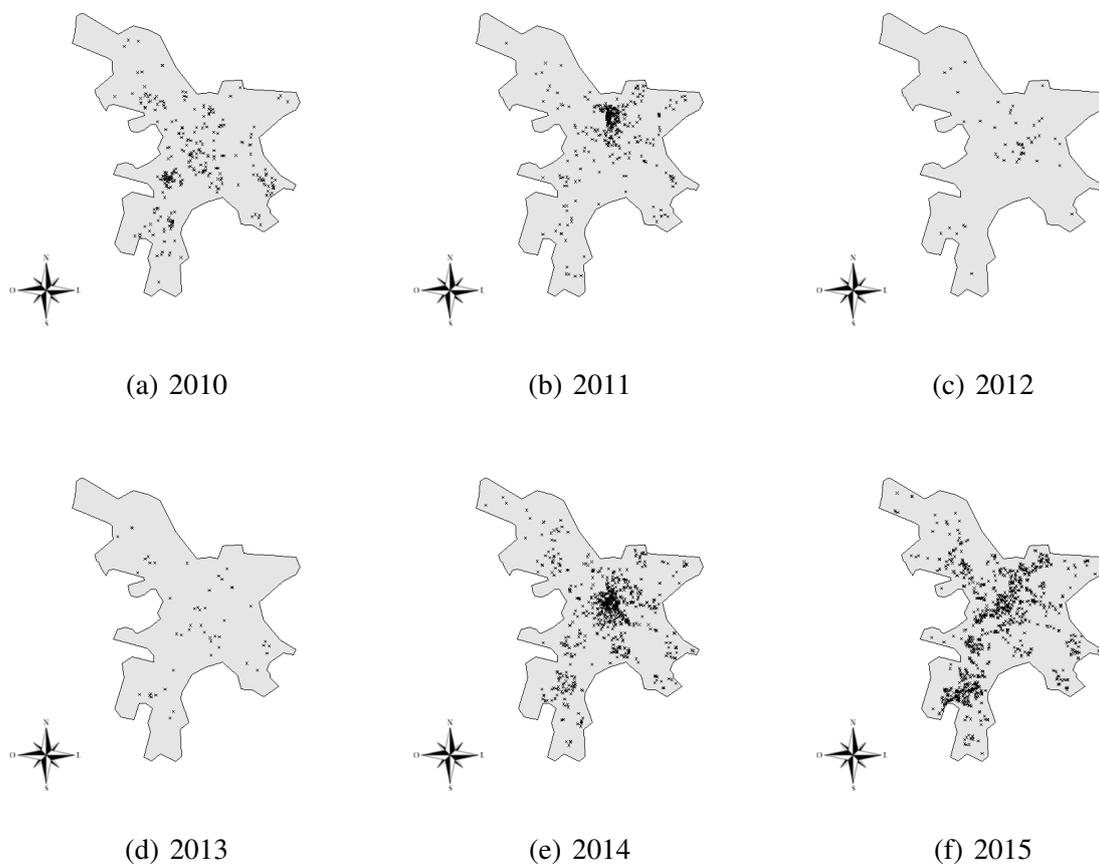
## 4 RESULTADOS

Nesta seção serão apresentados os resultados obtidos neste trabalho. Primeiramente, será apresentada uma análise preliminar do conjunto de dados a serem utilizados. Posteriormente, será exposto o ajuste do modelo de Cox log-Gaussiano e a predição espaço-temporal realizada. Por fim, serão discutidos os resultados encontrados.

### 4.1 Análises preliminares

A Figura 4.1 apresenta a distribuição espacial dos casos de dengue notificados em Três Corações, o qual constituem os eventos. Ao todo, foram observados 2818 casos de dengue, distribuídos sobre o município, por ano, durante o período de 2010 a 2015, como mostram as Figuras 4.1(a) a 4.1(f).

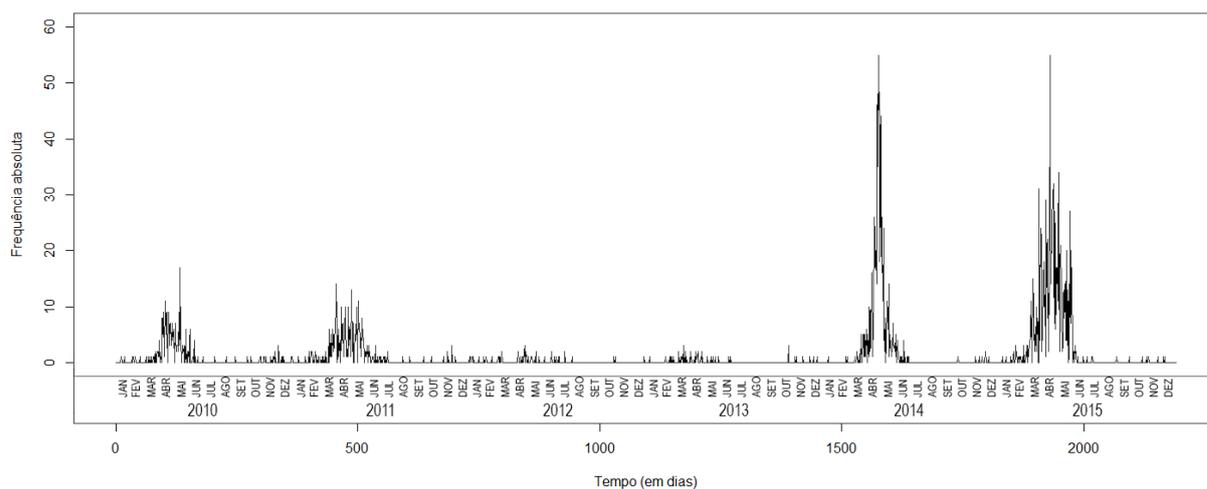
Figura 4.1 – Distribuição dos casos de dengue na área urbana do município de Três Corações - MG, nos anos de 2010 a 2015.



É possível observar nos mapas da Figura 4.1 que os eventos se concentram em torno da região central da cidade. No entanto, é possível visualizar algumas áreas com forte concentração de casos. Em 2010, observa-se que a região sudoeste apresenta uma leve aglomeração de casos. Para o ano de 2011, os casos observados podem ser vistos com uma frequência maior entre a região central e a região norte do município. Já para os anos de 2012 e 2013, observa-se uma frequência menor de eventos em relação aos anos anteriores. Logo, não é possível identificar, visualmente, áreas com grandes quantidades de observações. Contudo, os anos subsequentes, 2014 e 2015, apresentam um aumento de casos de dengue, cuja maior concentração de casos notificados se encontram na região central, para 2014. Para 2015, observa-se uma grande incidência de casos que vai desde o sul do município até a região norte.

Já a Figura 4.2 apresenta o gráfico da frequência absoluta dos 2.818 casos de dengue distribuídos por dia, durante os seis anos considerados nesta análise. Em relação aos valores relacionados aos dias, optou-se por trabalhar com uma medida de tempo de tal forma que, o dia 1 correspondesse a data de 01 de Janeiro de 2010 e o dia 2.191 correspondesse ao dia 31 de Dezembro de 2015. Na Figura 4.2, observa-se a existência de períodos com altas frequências de casos, sendo que, para todos os anos, os meses que apresentaram o maior número de casos foram os meses entre Fevereiro e Junho.

Figura 4.2 – Frequência dos casos de dengue em dias durante os anos de 2010 e 2015.



## 4.2 Ajuste do modelo

Inicialmente, foi realizada a estimação da variação espacial relacionada à componente fixa do modelo, descrita na equação (2.59). Esta componente, basicamente, tem por objetivo estimar a densidade espacial dos casos de dengue em Três Corações.

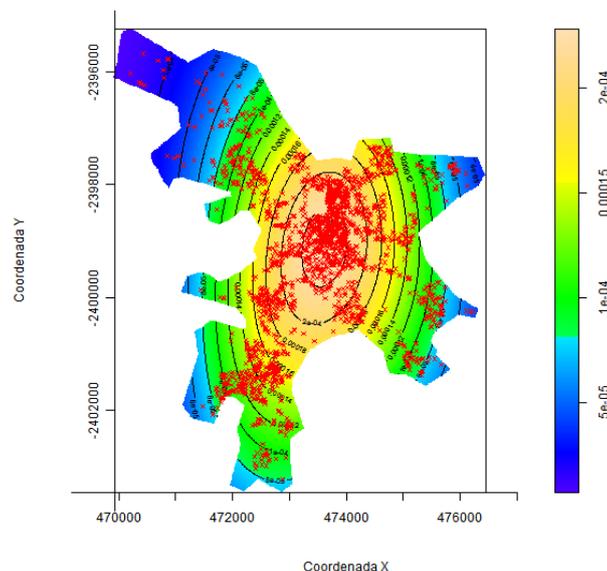
Assim, para estimar  $\rho(\mathbf{s})$ , foi utilizado o método não-paramétrico de alisamento de kernel. Para tanto, foi considerado uma função kernel Gaussiana resultante do produto de duas Gaussianas univariadas, a fim de utilizar medidas diferentes para a largura de banda das coordenadas X e Y. Deste modo, sendo  $x_i$  e  $y_i$  denotadas como coordenadas de  $\mathbf{s}_i$ , a função kernel pode ser fatorada como

$$k\left(\frac{\mathbf{s}_0 - \mathbf{s}_i}{h}\right) = k\left(\frac{x_0 - x_i}{h_x}\right) k\left(\frac{y_0 - y_i}{h_y}\right),$$

em que  $k(\cdot)$  é uma função kernel Gaussiana e  $h_x$  e  $h_y$  denotam as larguras de banda para as coordenadas X e Y respectivamente. Logo, foi considerada uma largura de banda para a coordenada X de 1000 metros e para a coordenada Y, 2300 metros, pois observa-se que a dispersão entre os eventos é maior para a coordenada Y se comparada a coordenada X.

Em relação à correção de efeito de borda, foi considerado o método proposto por Diggle (1985), que considera a proporção do volume kernel que se encontra dentro da região espacial. A Figura 4.3 apresenta o gráfico da estimativa da densidade ajustada.

Figura 4.3 – Estimativa da densidade espacial por alisamento de Kernel.



Em relação a outra componente fixa do modelo,  $m(t)$ , que descreve a intensidade temporal do processo, foi utilizado um modelo de regressão considerando respostas advindas de uma distribuição de Poisson com esperança  $m(t)$ , sendo esta esperança explicada pelo tempo.

Como apresentado anteriormente, o padrão empírico da frequência dos casos de dengue observados por dia, ao longo dos seis anos em análise, mostra alguns períodos com altas incidências da doença. A fim de se obter um modelo que fosse capaz de descrever esses períodos sazonais, foram utilizadas algumas transformações em relação às variáveis explicativas do modelo. Assim, considerando uma combinação de distribuições Gaussianas como transformação, o ajuste que modela o número esperado de casos de dengue por dia é dado por:

$$m(t) = e^{\eta(t)}, \quad (4.1)$$

tal que

$$\begin{aligned} \eta(t) = & \beta_0 + \beta_1 \frac{\exp\left\{-\frac{(t-120)^2}{2 \times 900}\right\}}{\sqrt{2 \pi} 900} + \beta_2 \frac{\exp\left\{-\frac{(t-480)^2}{2 \times 1000}\right\}}{\sqrt{2 \pi} 1000} + \beta_3 \frac{\exp\left\{-\frac{(t-1190)^2}{2 \times 500}\right\}}{\sqrt{2 \pi} 500} \\ & + \beta_4 \frac{\exp\left\{-\frac{(t-1575)^2}{2 \times 300}\right\}}{\sqrt{2 \pi} 300} + \beta_5 \frac{\exp\left\{-\frac{(t-1935)^2}{2 \times 1000}\right\}}{\sqrt{2 \pi} 1000}. \end{aligned}$$

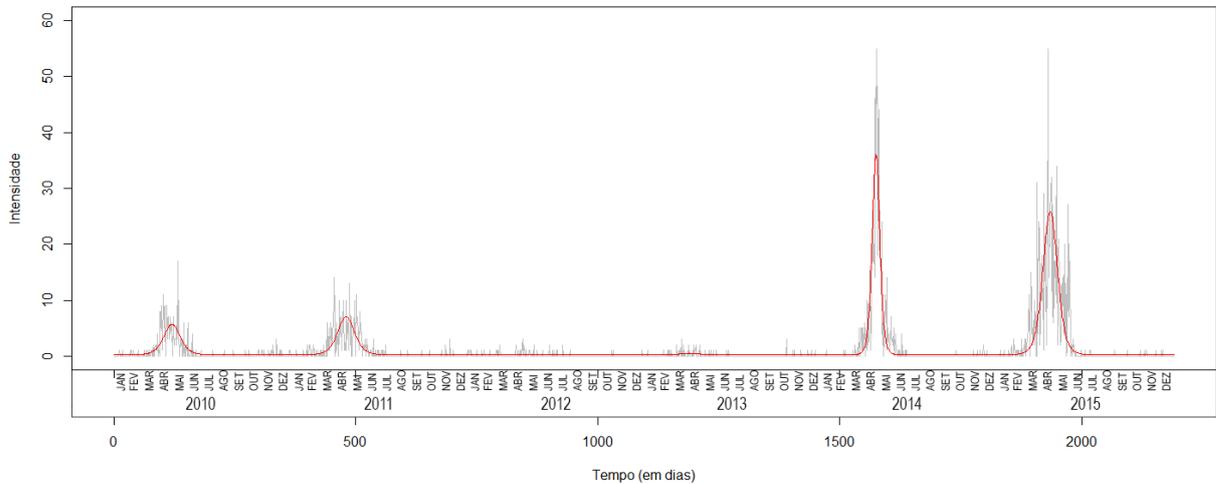
Assim, os valores dos parâmetros estimados, via máxima verossimilhança, é apresentado na Tabela 4.1.

Tabela 4.1 – Estimativa dos parâmetros  $\beta_i$ 's do modelo 4.1

	Estimativas	Erros Padrão	Z	valor-p
$\beta_0$	-1,53	0,05	-33,05	< 0,001
$\beta_1$	245,92	6,72	36,62	< 0,001
$\beta_2$	276,18	6,55	42,19	< 0,001
$\beta_3$	49,45	13,71	3,61	< 0,001
$\beta_4$	222,26	2,83	78,60	< 0,001
$\beta_5$	379,40	4,92	77,07	< 0,001

A partir das estimativas obtidas pelo modelo de regressão de Poisson, o ajuste final é apresentado na Figura 4.4.

Figura 4.4 – Estimativa da intensidade temporal ajustada por um modelo de regressão de Poisson (em vermelho).



Uma vez estimadas as componentes fixas do modelo,  $\rho(\mathbf{s})$  e  $m(t)$ , o próximo passo foi estimar os parâmetros relacionados à covariância espacial e temporal do campo aleatório Gaussiano,  $Z(\mathbf{s}, t)$ .

Para tanto, foi considerado uma função de correlação exponencial, tanto para a componente espacial quanto para a componente temporal. Utilizando a função de correlação entre pares (Função  $g$ ), expresso em (2.63), para a estimação dos parâmetros espaciais, foi considerada uma transformação logarítmica na função, ou seja,  $\log g(u) = \sigma_Z^2 \exp\{-\phi u\}$ . Já na utilização da Função  $K$  utilizou-se uma raiz quártica como transformação, ou seja,  $v[K(u)] = (K(u))^{0,25}$ . Em relação aos parâmetros de distância mínima,  $u_0$ , e distância máxima,  $u_{max}$ , foram utilizados os valores 0 e 1500 metros, respectivamente. Para o caso temporal, as distâncias mínimas e máximas utilizadas foram de 0 e 30 dias.

Primeiramente, foi realizado um ajuste "a sentimento" para se obter estimativas dos parâmetros das correlações espaciais e temporais. O resultado dessas estimativas foram utilizados como valores iniciais para propor um ajuste pelo método de mínimos contrastes.

Logo, os valores estimados para  $\sigma_Z$  e  $\phi$  foram 2,02 e 200 metros, respectivamente, tanto para a Função  $g$  quanto para a Função  $K$ . Já para  $\theta$ , o valor estimado foi 0,68 dias.

O próximo passo foi utilizar estes valores como escolhas iniciais para realizar uma estimação via mínimos contrastes. Desta forma, o critério de contraste para o caso espacial, considerando as funções  $g$  e  $K$ , respectivamente, utilizados neste trabalho foram

$$D_g(\sigma^2, \phi) = \int_0^{1500} \{\log(\hat{g}(u)) - \log(g_{\sigma^2, \phi}(u))\}^2 du \quad (4.2)$$

e

$$D_K(\sigma^2, \phi) = \int_0^{1500} \{(\hat{K}(u))^{0,25} - (K_{\sigma^2, \phi}(u))^{0,25}\}^2 du. \quad (4.3)$$

Para o caso temporal, utilizando a medida de autocovariância no tempo, o critério de contraste utilizado foi

$$D_C(\theta) = \sum_{v=1}^{30} \sum_{t=v+1}^n \{\hat{C}(t, v) - C(t, v; \theta)\}^2, \quad (4.4)$$

em que  $C(t, v; \theta)$  e  $\hat{C}(t, v)$  são definidos pelas equações (2.71) e (2.72), respectivamente.

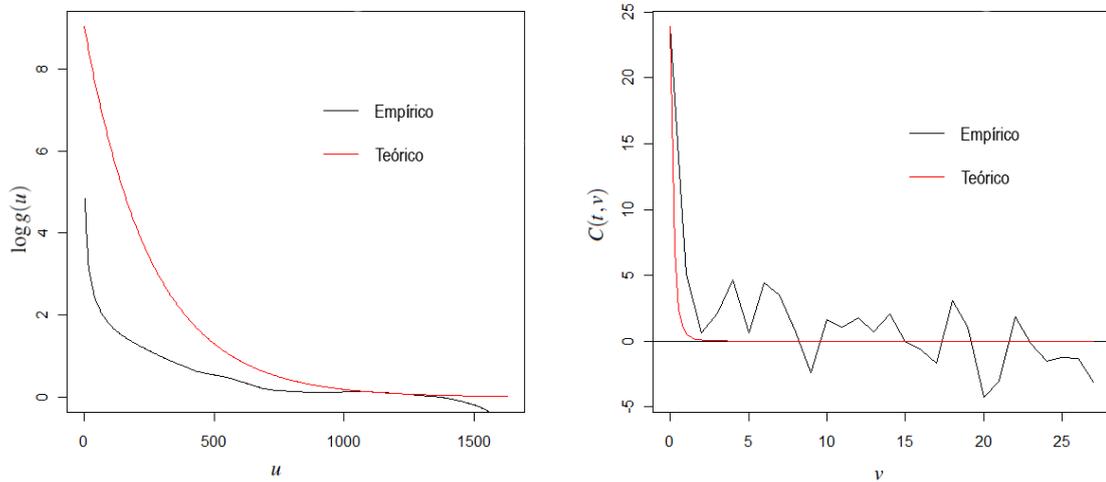
O resultado obtido, após a estimação, é apresentado na Tabela ??, utilizando as Funções  $g$  e  $K$ , respectivamente.

Tabela 4.2 – Valores estimados para  $\sigma$ ,  $\phi$  e  $\theta$  utilizando a Função  $g$  e  $K$ .

Funções	Parâmetros estimados			$D_g^2(\sigma^2, \phi)$
	$\phi$	$\sigma_Z$	$\theta$	
<b>Função <math>g</math></b>	258,03	3,007	0,553	58,9
<b>Função <math>K</math></b>	270,448	3,210	0,549	$5,18561 \times 10^{13}$

Para decidir sobre qual função utilizar, foi verificado o valor da discrepância quadrática das estimativas,  $D_g^2(\sigma^2, \phi)$ . Logo, a função escolhida foi a Função  $g$ , apresentando uma discrepância quadrática de 58,9. As Figuras 4.5(a) e 4.5(b) apresentam os gráficos da Função  $g$  e da função de autocovariância no tempo, respectivamente, resultantes do método de mínimos contrastes.

Figura 4.5 – Gráfico da função de correlação entre pares (Função  $g$ ) e da função de autocovariância temporal ajustados pelo método de mínimos contrastes.



(a) Logaritmo da função de correlação entre pares de eventos espacial.

(b) Função de autocovariância temporal.

Dessa forma, a função de correlação espacial do campo aleatório Gaussiano ajustada por mínimos contrastes é dada por

$$c_1(u) = \exp\{-258,03 u\}. \quad (4.5)$$

O valor estimado para o parâmetro  $\phi$  está relacionado com o alcance da dependência espacial do processo, isto é, representa a distância em que a dependência espacial entre eventos deixa de existir. No contexto da dengue, este valor pode estar relacionado com a capacidade de dispersão que o mosquito apresenta a partir de um ponto de origem. Estudos apontam que as fêmeas da espécie do *Aedes aegypti* podem voar entre 100 e 500 metros (PIROLA, 2016). Especificamente em Três Corações, estimou-se que o mosquito *Aedes aegypti* pode voar, em média, cerca de 258,03 metros.

Já a função de correlação temporal correspondente a  $Z$  é dada por

$$c_2(v) = \exp\{-0,553 v\}. \quad (4.6)$$

No contexto temporal, acredita-se que o parâmetro  $\theta$  possa estar relacionado com o tempo máximo que o mosquito *Aedes aegypti* leva para se alimentar de sangue humano nova-

mente. Neste contexto, a fêmea necessita se alimentar de sangue para a reprodução de ovos, conhecido como o período de maturação. De acordo com Instituto Oswaldo Cruz (2017), um mosquito comum se alimenta apenas uma única vez durante o período produtivo. No entanto, o mosquito da dengue tem uma certa particularidade, podendo se alimentar mais de uma vez durante esse período. Relatos revelam que um mesmo mosquito já se alimentou de cinco pessoas em um mesmo dia (INSTITUTO OSWALDO CRUZ, 2017). Para a cidade de Três Corações, foi estimado um alcance temporal de 0,553 dias, mostrando, assim, que um mosquito leva, em média, aproximadamente 13 horas para se alimentar novamente de sangue humano.

### 4.3 Predição espaço-temporal

Um dos resultados que se espera com este trabalho é identificar áreas de grande risco dos casos de dengue na região urbana do município de Três Corações e como esse risco se modifica ao longo do tempo. Assim, uma vez estimadas as componentes fixas e os parâmetros do modelo de correlação do processo de Cox log-Gaussiano, o próximo passo foi detectar essas áreas de risco por meio da predição espaço-temporal.

Para a realização da predição de  $R(\mathbf{s}, t)$ , utilizou-se somente o ano de 2015. Este ano apresentou um maior número de casos de dengue em relação aos anos anteriores (2010 à 2014). Assim, por meio dessas ocorrências, é possível observar como o padrão de risco se comporta na região espacial, ao longo do tempo. No entanto, foi considerado somente o período de grande surto epidêmico da doença aconteceu, aproximadamente, entre os dias 1850 e 1993, que corresponde ao período de 24 de janeiro à 16 de junho de 2015.

Dessa forma, foi realizada uma simulação utilizando o algoritmo de Langevin Hastings. Para tanto, foi considerada uma cadeia com 15000 iterações, das quais 5000 primeiras foram eliminadas. Além disso, foi adotado um salto de tamanho 100 a fim de se obter uma amostra aproximadamente independente da distribuição preditiva. Foi utilizado, também, uma variância adaptativa na distribuição do passeio aleatório a fim de se obter uma taxa de aceitação da cadeia em torno de 0,57, o que é recomendado por Roberts e Rosenthal (1998).

Uma forma de identificar estas áreas de grande risco é por meio das probabilidades de excedência, como descrito na subseção 2.7.2. Para a obtenção de um limiar,  $\varepsilon$ , que identificasse, de forma satisfatória, estas áreas com altas probabilidades, optou-se por um valor de  $\varepsilon$ , tal que  $Pr(R(\mathbf{s}, t) > \varepsilon \mid \{X_{W,1} \cup \dots \cup X_{W,t}\}) = 0,9$ . Desse modo, o valor obtido para  $\varepsilon$ , que satisfaça esta igualdade, foi de 0,51.

Os mapas apresentados nas Figuras 4.6 à 4.14 mostram os resultados das probabilidades de excedência para cada dia no período de grande surto epidêmico no ano de 2015. Assim, cada pixel contido no mapa está associado a uma probabilidade, de tal forma que áreas do mapa com alta probabilidade de ocorrências de casos de dengue apresentarão cores mais intensas (vermelho).

Figura 4.6 – Mapa de probabilidade de excedência da predição para os dias 24 de Janeiro à 08 de Fevereiro de 2015.



Figura 4.7 – Mapa de probabilidade de excedência da previsão para os dias 09 de Fevereiro à 24 de Fevereiro de 2015.



Figura 4.8 – Mapa de probabilidade de excedência da previsão para os dias 25 de Fevereiro à 12 de Março de 2015.

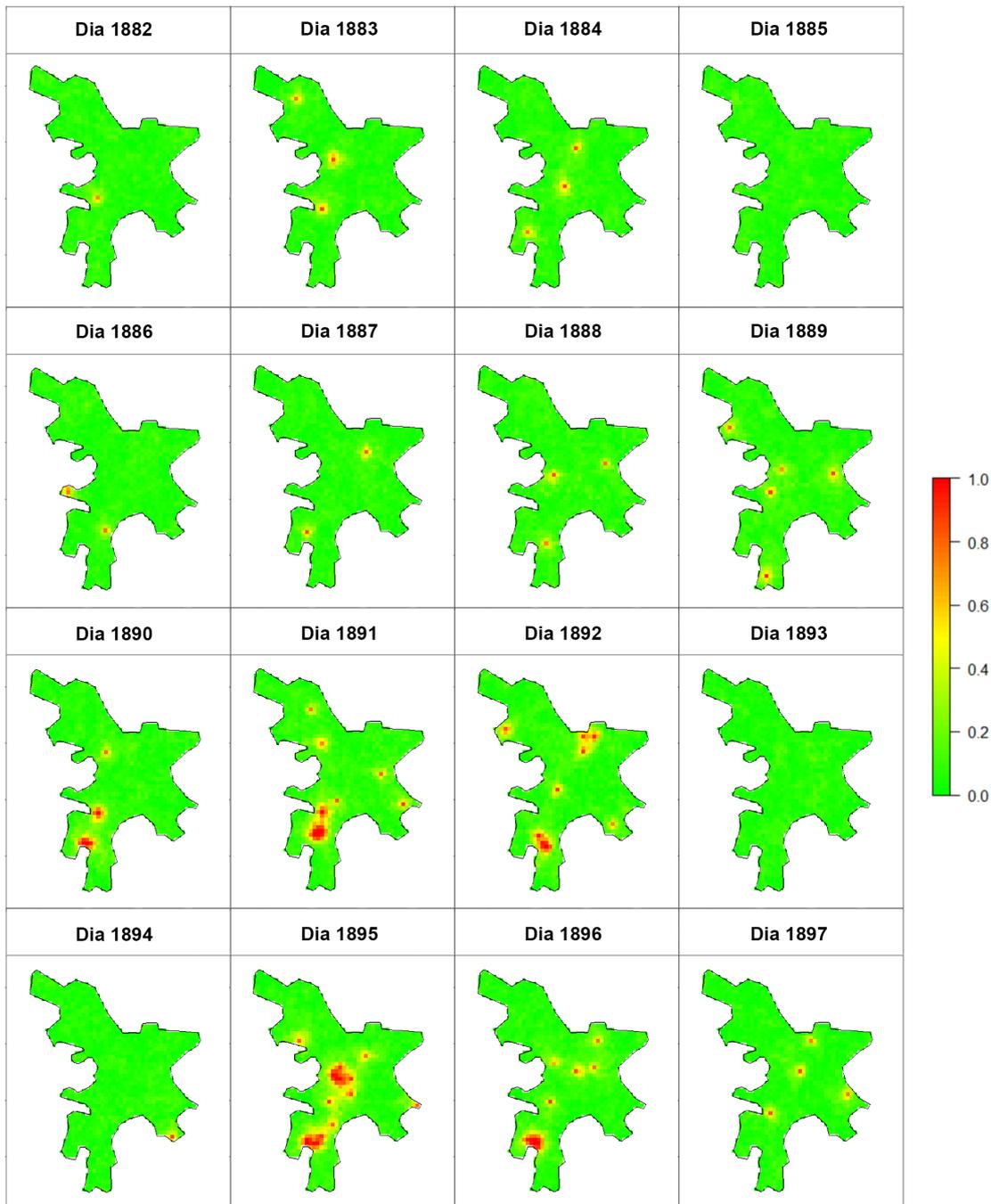


Figura 4.9 – Mapa de probabilidade de excedência da predição para os dias 13 de Março à 28 de Março de 2015.



Figura 4.10 – Mapa de probabilidade de excedência da predição para os dias 29 de Março à 13 de Abril de 2015.

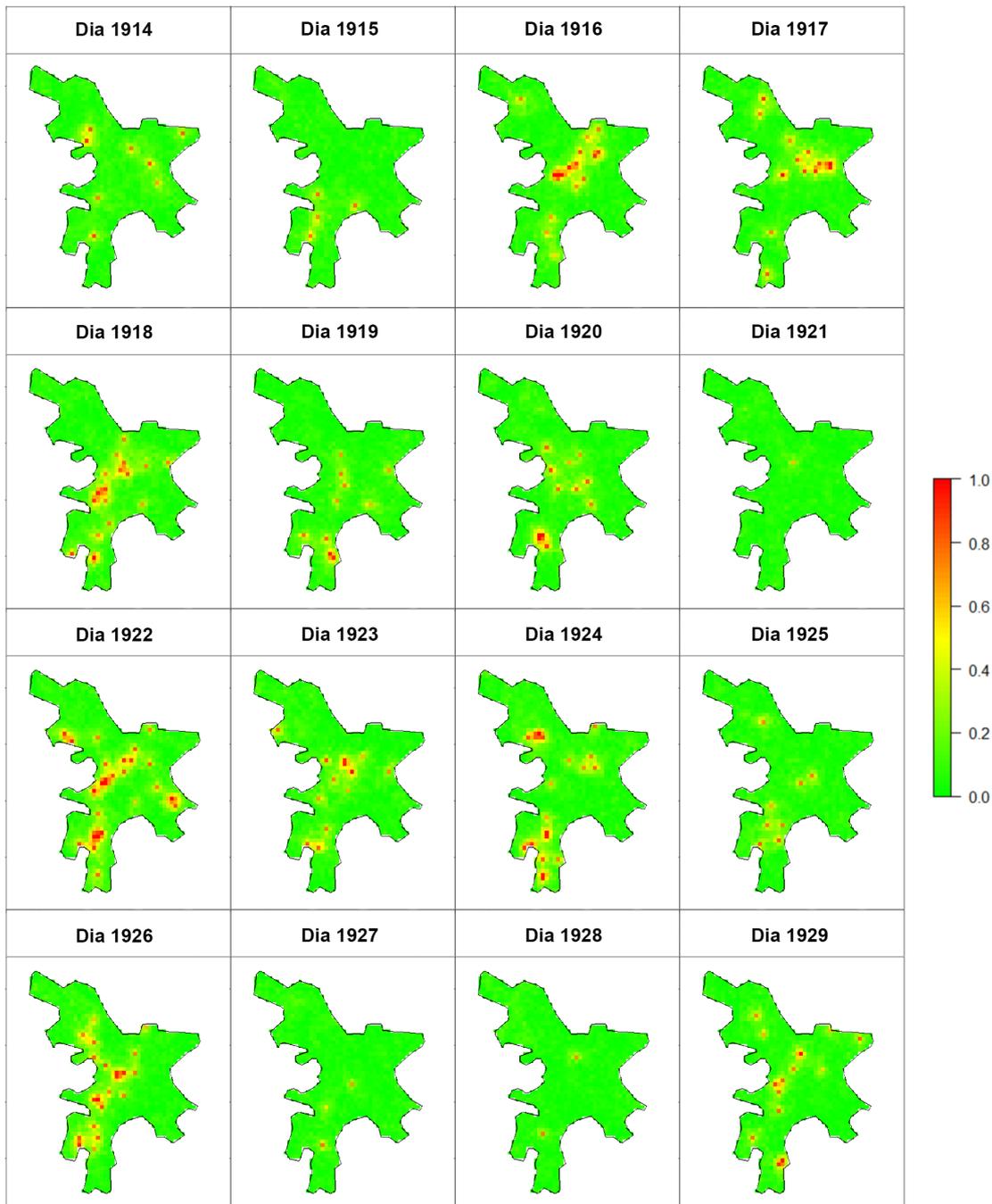


Figura 4.11 – Mapa de probabilidade de excedência da previsão para os dias 14 de Abril à 29 de Abril de 2015.

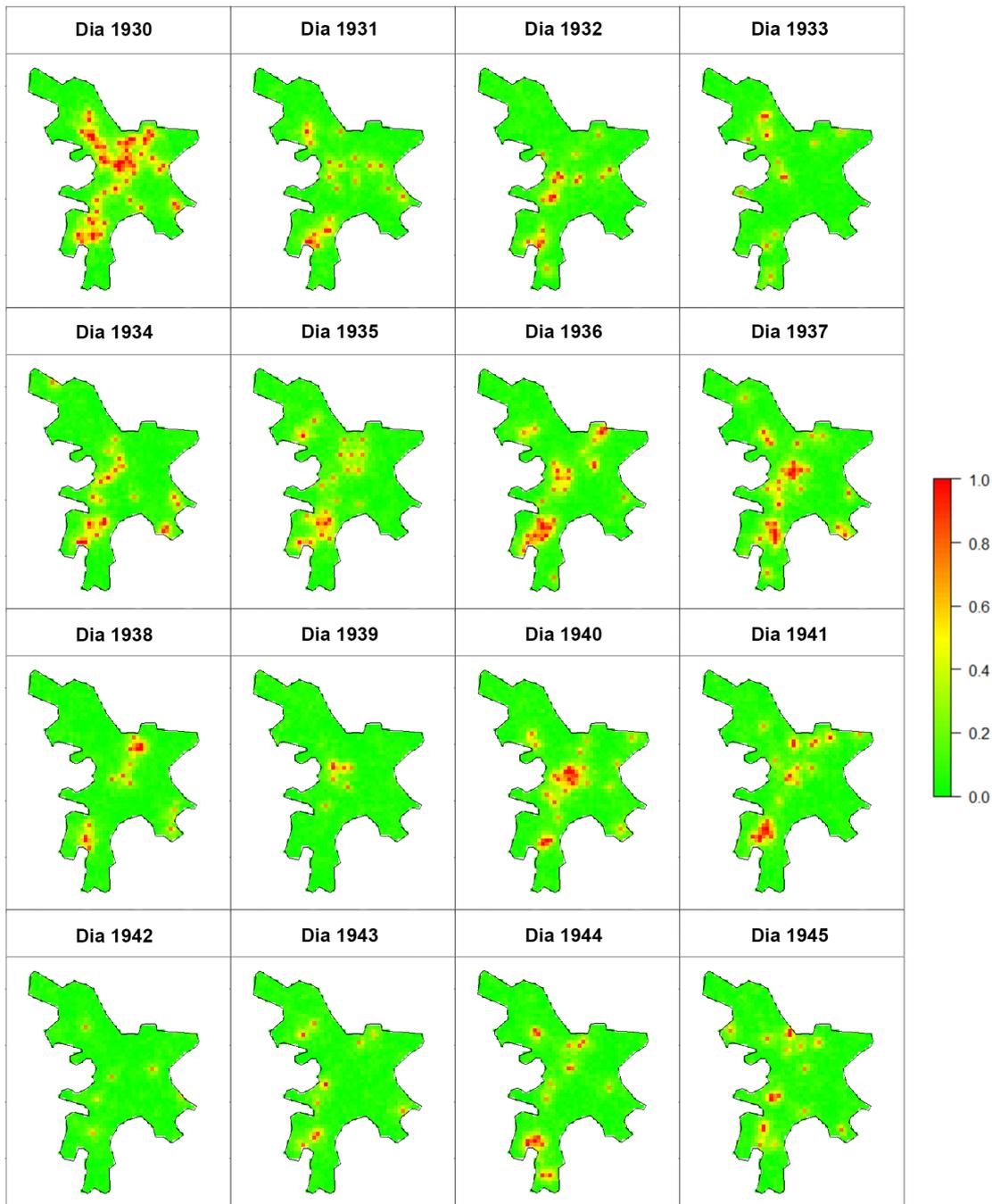


Figura 4.12 – Mapa de probabilidade de excedência da previsão para os dias 30 de Abril à 15 de Maio de 2015.

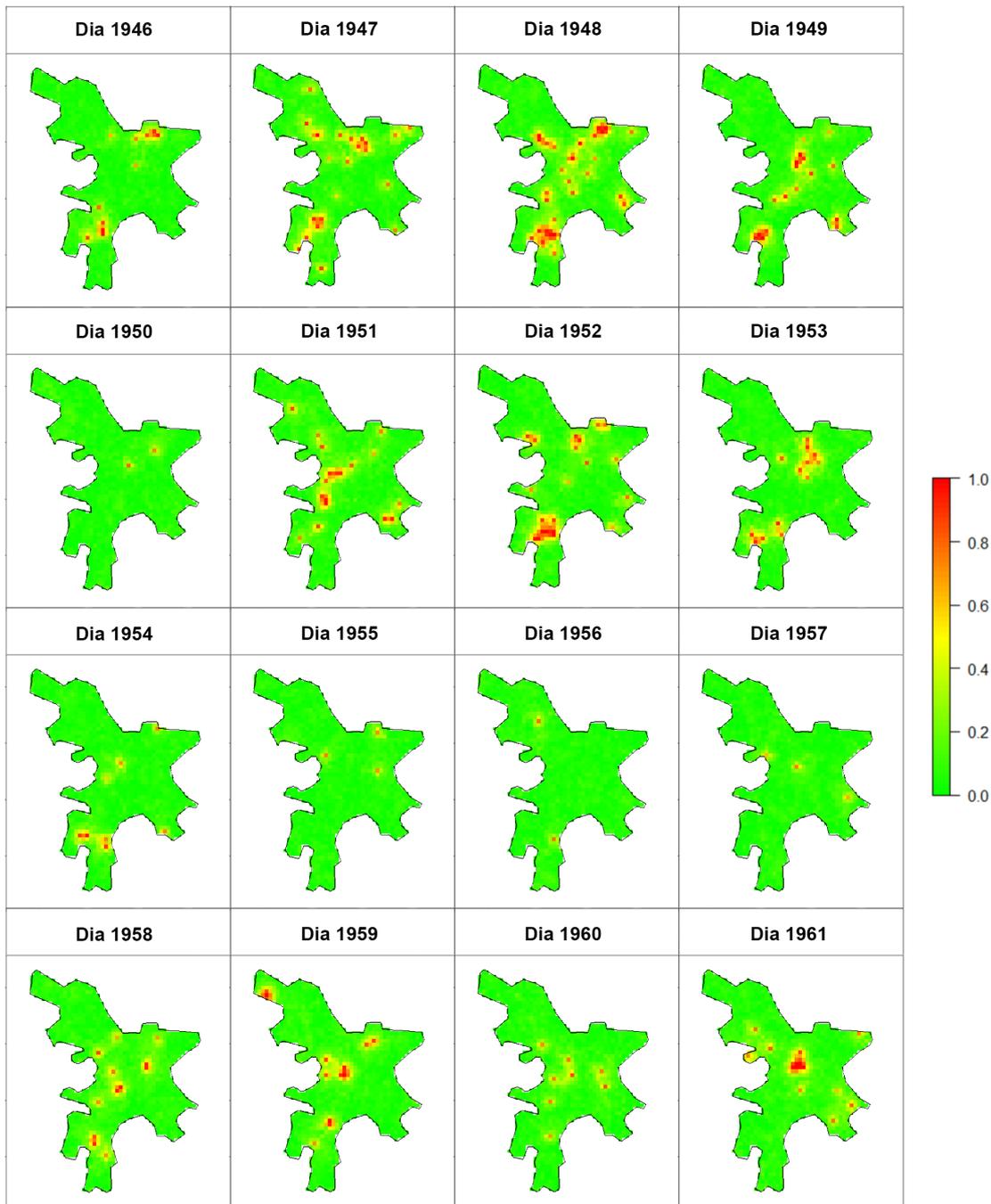


Figura 4.13 – Mapa de probabilidade de excedência da previsão para os dias 16 de Maio à 31 de Maio de 2015.

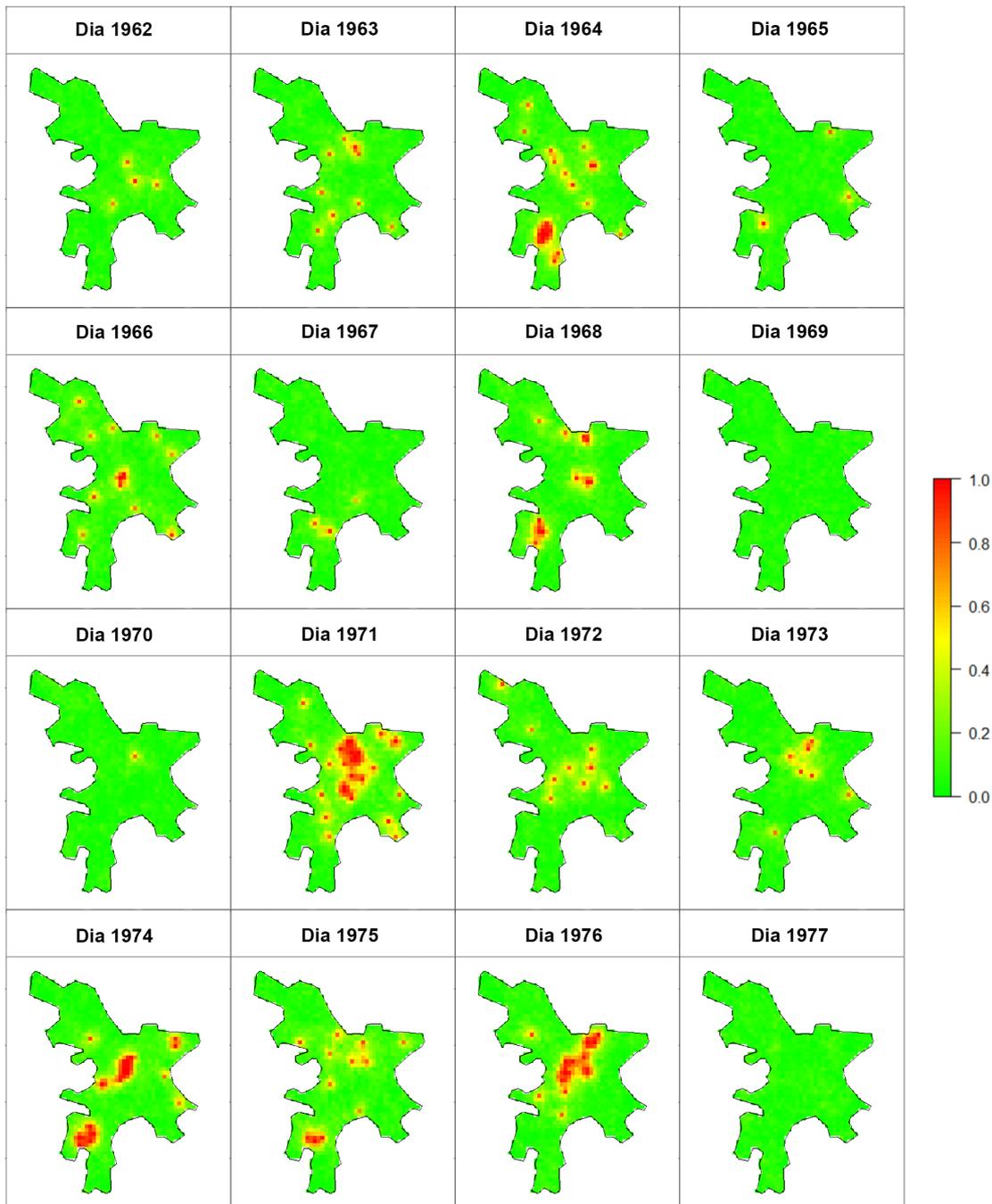
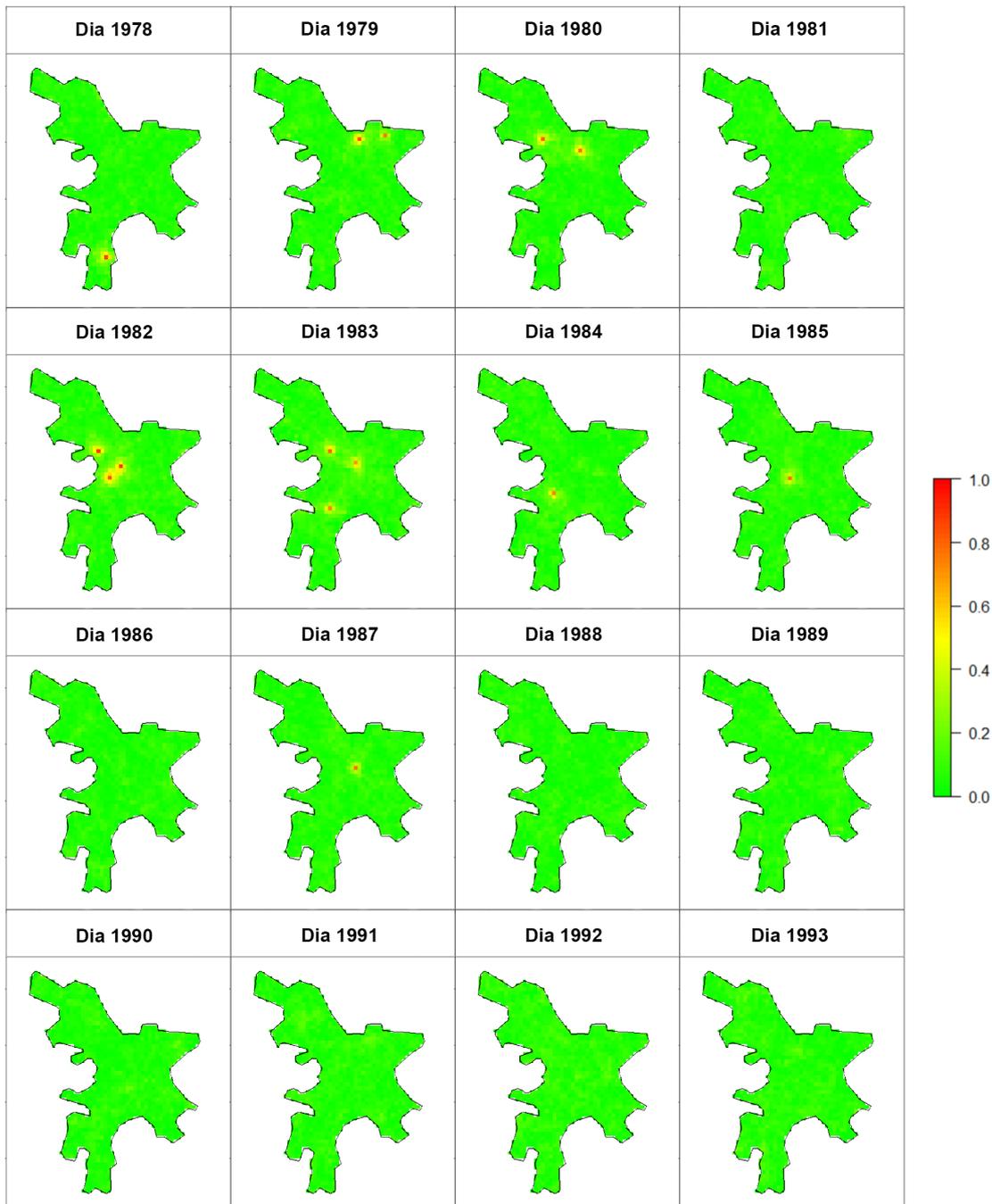


Figura 4.14 – Mapa de probabilidade de excedência da predição para os dias 01 de Junho à 16 de Junho de 2015.



As figuras apresentadas anteriormente apresentam mapas advindos da probabilidade de  $R(s, t)$  exceder o valor 0,51 condicionado aos eventos observados. Estas figuras apresentaram mapas de risco que descrevem possíveis anomalias dos casos de dengue na região espacial ao longo do tempo.

Os mapas apresentados entre os dias 1850 e 1889, correspondente ao período de 24 de janeiro à 04 de março, apresentaram algumas áreas na região espacial com alta probabilidade. No entanto, durante este período, essas áreas de alta incidência não apresentaram um padrão espaço-temporal, ocorrendo aleatoriamente em alguns locais do município de Três Corações, porém sem evolução no tempo.

A partir do dia 1890, que corresponde a 05 de março de 2015, passou-se a observar um aumento de áreas com probabilidades de risco consideradas altas, que se perpetuou até o dia 1976, que corresponde a data de 30 de maio do mesmo ano. Durante este período de tempo, pode-se observar que algumas dessas áreas se mantiveram com altas probabilidades por sucessivos dias. Este comportamento reflete uma possível anomalia da doença na região de estudo. Estas anomalias, no contexto da dengue, sugere possíveis surtos epidêmicos de tal forma que, a chance de casos de dengue acontecerem nessas áreas é muito maior em relação a outras localizações.

Com relação às regiões que apresentaram uma alta chance de ocorrências da doença, destacam-se os bairros Jardim Paraíso, Jardim Califórnia, Vila Santo Afonso e o bairro Vila Fernão Dias, localizados na região sul do município. Em relação à região central do mapa, pode-se observar um risco maior que cobre os bairros Cotia, Jardim Santa Tereza, Centro, Vila Gesse e Monte Alegre. Destacam-se, também, os bairros São Jerônimo e Cinturão Verde, situados mais ao sudeste do município. Mais ao norte da cidade, os bairros Però Um, Però Dois, Santana, Parque São José, Vila Lima, Loteamento Bela Vista e o Odilon Rezende também apresentaram riscos elevados. Vale destacar a proximidade existente entre os bairros citados, ressaltando a presença de aglomerados de casos, característica típica da doença.

Após este período da possibilidade de grandes anomalias na região de estudo, as probabilidades de excedência em  $R(s,t)$  passaram, novamente, a apresentar baixas evidências de grande risco em localizações, ocorrendo aleatoriamente até o momento em que estas probabilidades passam a ser praticamente iguais a zero.

A Figura 4.15 apresenta, de forma resumida, todas as áreas o qual foram consideradas com risco elevado da doença durante o período de 05 de março à 30 de maio de 2015, por meio dos mapas apresentados entre as Figuras 4.8 e 4.13.



tificação da doença. Para alguns casos ocorridos, estas informações estavam incompletas e, portanto, tiveram de ser retiradas da base de dados. Deste modo, deve-se ressaltar a importância de que tais informações sejam inseridas, durante o registro da ocorrência, de forma correta e completa por parte dos postos de saúde do município. Garantindo, assim, a qualidade destas informações, estudos estatísticos voltados para análise espacial e/ou temporal poderão ser realizados, garantindo, assim, bons resultados.

Outro fator limitador, o qual foi imposto, é que a localização espacial assumida para a caracterização do evento foi a residência da pessoa infectada. No entanto, esse ponto geográfico não reflete a real localização em que, de fato, o evento ocorreu, o que impacta, também, nos resultados encontrados.

Outra possível justificativa para obtenção de erros tão elevados na predição pode estar relacionado com a escolha do método de estimação adotado neste trabalho, isto é, o método de mínimos contrastes. A ausência de covariáveis explicando o comportamento da intensidade espaço-temporal pode, também, ter corroborado para tais resultados.

Em relação à estrutura de correlação espaço-temporal para o campo aleatório Gaussiano,  $Z(\mathbf{s}, t)$ , foi assumido uma estrutura de separabilidade. Esta suposição é imposta devido ao ganho computacional e a fim de que os métodos de estimação, por meio de mínimos contrastes, possam ser utilizados de forma satisfatória. No entanto, considerar uma estrutura de separabilidade espaço-temporal pode ocasionar em perdas de informações caso haja algum tipo de interação entre tempo e espaço.

Com relação aos trabalhos futuros, pretende-se utilizar outros métodos inferenciais como, por exemplo, os métodos bayesianos para a estimação de parâmetros do modelo em estudo a fim de verificar a melhoria de tais resultados. Outra abordagem a ser considerada é a inclusão de marcas relacionadas aos eventos, a fim de que se possa extrair mais informações relacionadas ao fenômeno estudado.

## 5 CONCLUSÕES

Este trabalho teve por objetivo modelar o comportamento espaço-temporal dos casos de dengue na cidade de Três Corações, Minas Gerais. O modelo proposto foi o modelo de Cox log-Gaussiano, baseado na teoria de processos pontuais.

Assim, pode-se concluir que o modelo de Cox log-Gaussiano espaço-temporal apresentou um ajuste relativamente bom, contribuindo para a detecção de áreas e períodos de grande incidência de casos de dengue na cidade de Três Corações - MG.

Com estes resultados, espera-se que os órgãos públicos do município de Três Corações voltados para a saúde pública possam utilizar destas informações, a fim de que haja uma atenção maior para estas regiões mais críticas, onde o risco da doença é elevado. Assim, com as devidas políticas de prevenção e conscientização da população, é possível reduzir estes números para os anos subsequentes.

## REFERÊNCIAS

- BADDELEY, A.; RUBAK, E.; TURNER, R. **Spatial point patterns: Methodology and applications with R**. Florida: Chapman and Hall/CRC, 2015. 810p.
- 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.
- BICALHO, C. C.; SAFADI, T.; CHARRET, I. The influence of climatic factors on dengue epidemics in the cities Cuiaba (Mato Grosso State) and Lavras (Minas Gerais State), Brazil, using statistical methods. **Revista Brasileira de Biometria**, v. 32, n. 2, p. 308–322, 2014.
- BIVAND, R. S.; PEBESMA, E. J.; GÓMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**. New York: Springer, 2008.
- BRAGA, I. A.; VALLE, D. *Aedes aegypti*: histórico do controle no Brasil. **Epidemiologia e Serviços de Saúde**, v. 16, n. 2, p. 113–118, abr/jun 2007.
- BRILLINGER, D. R.; GUTTORP, P. M.; SCHOENBERG, F. P. Point processes, temporal. **Wiley StatsRef: Statistics Reference Online**, Wiley Online Library, v. 03, p. 1577–1581, 2002.
- BRIX, A.; DIGGLE, P. J. Spatiotemporal prediction for log-Gaussian Cox processes. **Journal of the Royal Statistical Society. Series B**, v. 63, n. 4, p. 823–841, 2001.
- CASTELLANOS, J.; BELLO, J.; VELANDIA-ROMERO, M. Manifestaciones neurológicas durante la infección por el virus del dengue. **Infectio**, v. 18, n. 4, p. 167–176, sep/dec 2014. Disponível em: <[http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-93922014000400008&nrm=iso](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-93922014000400008&nrm=iso)>.
- CATÃO, R. C. **Dengue no Brasil: abordagem geográfica na escala nacional**. Dissertação (Dissertação de Mestrado Acadêmico) — Universidade Estadual Paulista (UNESP), Presidente Prudente, fev. 2011.
- CHILÈS, J.-P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2. ed. Hoboken: John Wiley & Sons, 2009.
- COX, D.; ISHAM, V. **Point processes**. 1. ed. Florida: Chapman and Hall/CRC, 1980. (Monographs on applied probability and statistics).
- CRESSIE, N. A. **Statistics for spatial data**. Hoboken: John Wiley & Sons, 1993.
- DALBEM, A. G. et al. Dengue clássica e febre hemorrágica da dengue: etiologia, fisiologia, epidemiologia e fatores de risco. **Revista Ciência e Estudos Acadêmicos de Medicina**, v. 01, n. 01, p. 18–36, jan. - jul. 2014.
- DAVIES, T. M.; HAZELTON, M. L. Assessing minimum contrast parameter estimation for spatial and spatiotemporal log-Gaussian Cox processes. **Statistica Neerlandica**, v. 67, n. 4, p. 355–389, 2013.
- DIAS, L. B. et al. Dengue: transmissão, aspectos clínicos, diagnóstico e tratamento. **Medicina (Ribeirao Preto. Online)**, v. 43, n. 2, p. 143–152, 2010.

- DICK, O. B. et al. Review: The History of Dengue Outbreaks in the Americas. **The American journal of tropical medicine and hygiene**, v. 87, n. 4, p. 584– 593, 2012. Disponível em: <<http://www.ajtmh.org/content/87/4/584.short>>.
- DIGGLE, P. J. A Kernel Method for Smoothing Point Process Data. **Applied statistics**, v. 34, n. 2, p. 138–147, 1985.
- DIGGLE, P. J. **Statistical analysis of spatial and spatio-temporal point patterns**. 3. ed. Florida: Chapman and Hall/CRC, 2013.
- DIGGLE, P. J. et al. Second-order analysis of space-time clustering. **Statistical Methods in Medical Research**, Sage Publications, v. 4, n. 2, p. 124–136, jun. 1995.
- DIGGLE, P. J. et al. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. **Statistical Science**, v. 28, n. 4, p. 542–563, 2013.
- DIGGLE, P. J.; ROWLINGSON, B.; SU, T. li. Point process methodology for on-line spatio-temporal disease surveillance. **Environmetrics**, v. 16, n. 5, p. 423–434, 2005.
- FINKENSTÄDT, B.; HELD, L.; ISHAM, V. **Statistical methods for spatio-temporal systems**. Florida: Chapman and Hall/CRC, 2006.
- GABRIEL, E.; DIGGLE, P. J. Second-order analysis of inhomogeneous spatio-temporal point process data. **Statistica Neerlandica**, Wiley Online Library, v. 63, n. 1, p. 43–51, 2009.
- GABRIEL, E.; ROWLINGSON, B.; DIGGLE, P. J. stpp: an R package for plotting, simulating and analyzing Spatio-Temporal Point Patterns. **Journal of Statistical Software**, v. 53, n. 2, p. 1–29, abr. 2013.
- GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. 2. ed. Florida: Chapman and Hall/CRC, 2006.
- GELFAND, A. E. et al. **Handbook of spatial statistics**. Florida: Chapman and Hall/CRC, 2010.
- GOOGLE. **Google Maps**. 2016. Site. Disponível em: <<https://www.google.com.br/maps/>>.
- GUBLER, D. J. Dengue and Dengue Hemorrhagic Fever. **Clinical Microbiology Reviews**, v. 11, n. 03, p. 480 – 496, jul 1998.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, Biometrika Trust, v. 57, n. 1, p. 97–109, 1970.
- IBGE. - **Instituto Brasileiro de Geografia e Estatística**. 2016. Acesso em: 14 mai. 2016. Disponível em: <<http://ibge.gov.br/home/>>.
- INSTITUTO OSWALDO CRUZ. **Curiosidades sobre o A. Aegypti**. 2017. Site. Acesso em: 08 mar. 2017. Disponível em: <<http://www.ioc.fiocruz.br/dengue/textos/curiosidades.html>>.
- LOECHER, M.; ROPKINS, K. RgoogleMaps and loa: Unleashing R Graphics Power on Map Tiles. **Journal of Statistical Software**, v. 63, n. 4, p. 1–18, 2015. Disponível em: <<http://www.jstatsoft.org/v63/i04/>>.

METROPOLIS, N. et al. Simulated annealing. **Journal of Chemical Physics**, v. 21, p. 1087–1092, 1953.

MINISTÉRIO DA SAÚDE. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 52, 2015. **Boletim Epidemiológico**, v. 47, n. 3, 2016.

MØLLER, J.; SYVERSVEEN, A. R.; WAAGEPETERSEN, R. P. Log-Gaussian Cox processes. **Scandinavian journal of statistics**, Wiley Online Library, v. 25, n. 3, p. 451–482, 1998.

PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. **R news**, v. 05, n. 02, p. 9–13, 2005.

PINHEIRO, F.; NELSON, M. Re-Emergence of Dengue and Emergence of Dengue Haemorrhagic Fever in the Americas. **Dengue Bulletin**, v. 21, p. 16 – 24, dec 1997.

PIROLA, C. L. **Análise espacial e temporal da distribuição de Aedes aegypti e Aedes albopictus na Associação dos Municípios da Região Carbonífera (AMREC) com uma proposta metodológica de monitoramento populacional dos vetores**. Trabalho de conclusão de curso (bacharelado em ciências biológicas). Criciúma, 2016.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.

REZENDE, J. M. Notas históricas e filológicas sobre a palavra dengue. **Revista de Patologia Tropical**, v. 26, n. 2, 1997.

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.

ROBERT, C.; CASELLA, G. **Introducing Monte Carlo Methods with R**. New York: Springer, 2009.

ROBERTS, G. O.; ROSENTHAL, J. S. Optimal scaling of discrete approximations to Langevin diffusions. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 60, n. 1, p. 255–268, 1998.

ROSS, S. M. **Stochastic Processes**. Second edition. New York: John Wiley & Sons, 1996.

SCANDAR, S. A. S. **Análise espacial da distribuição dos casos de dengue e a relação com fatores entomológicos, ambientais e socioeconômicos no município de São José do Rio Preto-SP-Brasil**. Tese (Tese de doutorado) — Universidade de São Paulo, 2007.

SCHABENBERGER, O.; GOTWAY, C. A. **Statistical methods for spatial data analysis**. Florida: Chapman and Hall/CRC, 2004.

SHERMAN, M. **Spatial statistics and spatio-temporal data: covariance functions and directional properties**. Chichester: John Wiley & Sons, 2011.

SILVA, A. S. da; RIBEIRO JR, P. J.; ELMATZOGLOU, I. Modelagem geoestatística utilizando a família de Gneiting de funções de covariâncias espaço-temporais. **Revista de Matemática e Estatística**, v. 25, n. 1, p. 65–83, jan.-mar. 2007.

STOYAN, D.; STOYAN, H. **Fractals, random shapes and point fields**. Chichester: John Wiley & Sons, 1994.

TAPIA-CONYER, R.; MENDEZ-GALVAN, J. F.; GALLARDO-RINCON, H. The growing burden of dengue in Latin America. **Journal of Clinical Virology**, v. 46, p. S3–S6, 2009. Disponível em: <[http://www.journalofclinicalvirology.com/article/S1386-6532\(09\)70286-0/abstract](http://www.journalofclinicalvirology.com/article/S1386-6532(09)70286-0/abstract)>.

TAUIL, P. L. Urbanização e ecologia do dengue. **Cadernos de Saúde Pública**, scielosp, v. 17, p. 99 – 102, jan 2001. Disponível em: <[http://www.scielosp.org/scielo.php?script=sci\\_arttext&pid=S0102-311X2001000700018&nrm=iso](http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0102-311X2001000700018&nrm=iso)>.

TAYLOR, B. M. et al. Igcpx - An R Package for inference with Spatio-Temporal Log-Gaussian Cox Processes. **Journal of Statistical Software**, v. 52, n. 4, jan 2013.

TAYLOR, B. M. et al. Bayesian inference and data augmentation schemes for spatial, spatio-temporal and multivariate Log-Gaussian Cox processes in R. **Journal of Statistical Software**, v. 63, p. 1–48, 2015.

TAYLOR, B. M.; DIGGLE, P. J. INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 84, n. 10, p. 2266–2284, 2014.

TEIXEIRA, M. G. Few characteristics of dengue's fever epidemiology in Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, scielo, v. 54, p. 1 – 4, 10 2012. ISSN 0036-4665. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0036-46652012000700002&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0036-46652012000700002&nrm=iso)>.

VASCONCELOS, P. F. C. et al. Aedes aegypti, Dengue and Re-urbanization of Yellow Fever in Brazil and other South American Countries - Past and Present Situation and Future Perspectives. **Dengue Bulletin**, v. 23, p. 55–66, 1999.

WALLER, L. A.; GOTWAY, C. A. **Applied Spatial Statistics for Public Health Data**. Hoboken: John Wiley & Sons, 2004. v. 368.

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

**APÊNDICE A – Mapas para o erro de predição**

Figura 1 – Erro padrão de predição para os dias 24 de Janeiro à 08 de Fevereiro.

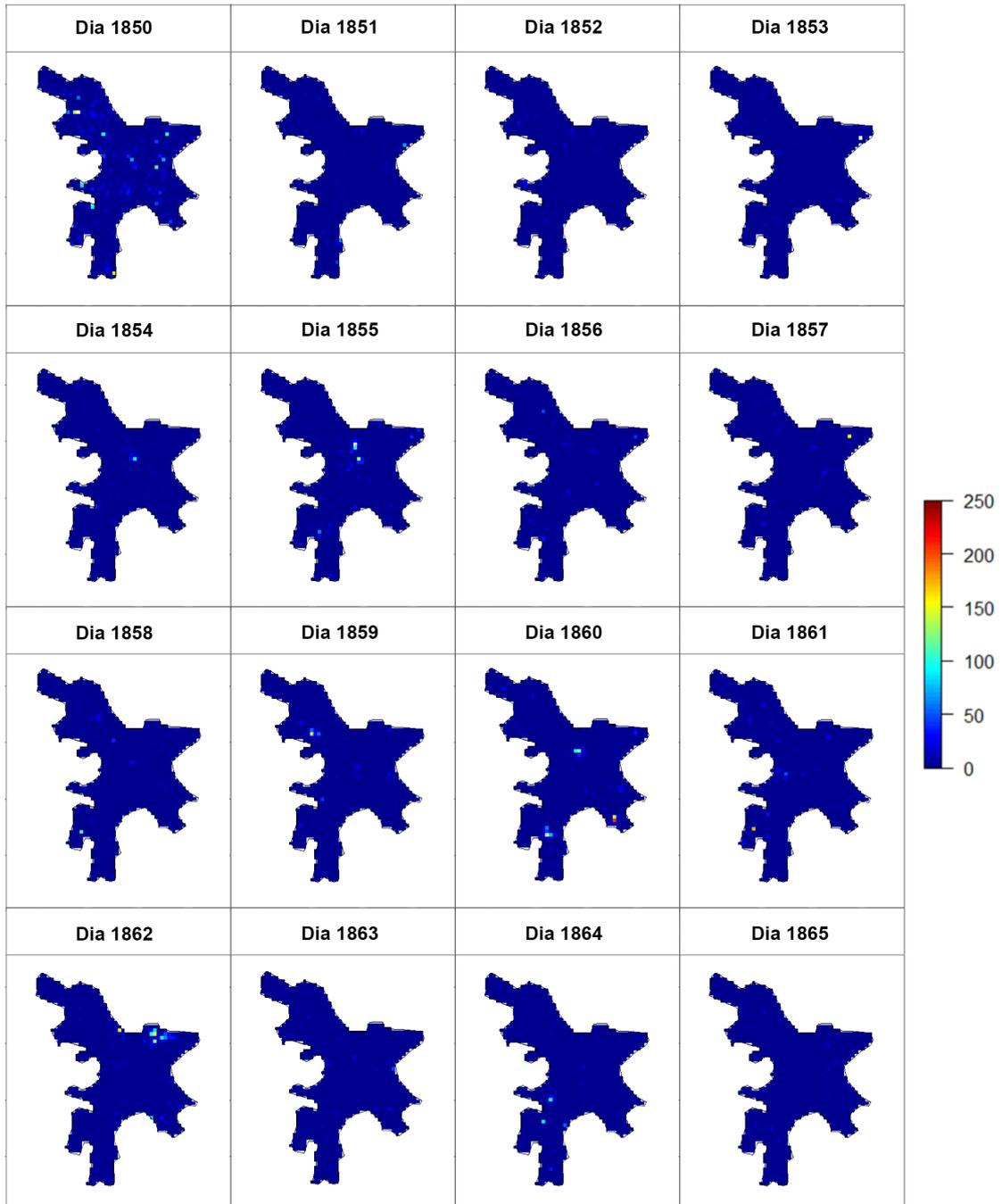


Figura 2 – Erro padrão de predição para os dias 09 de Fevereiro à 24 de Fevereiro.

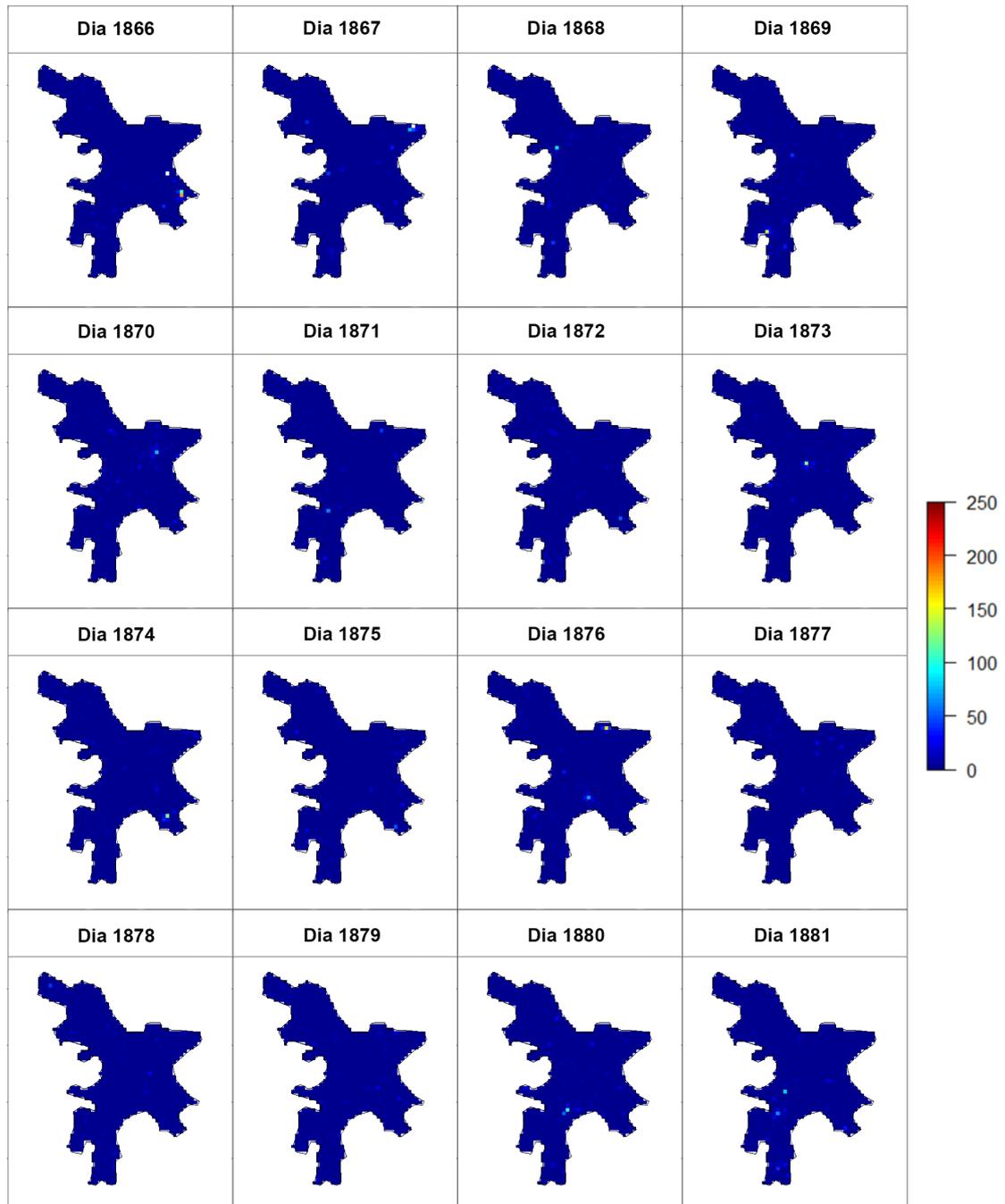


Figura 3 – Erro padrão de predição para os dias 25 de Fevereiro à 12 de Março.

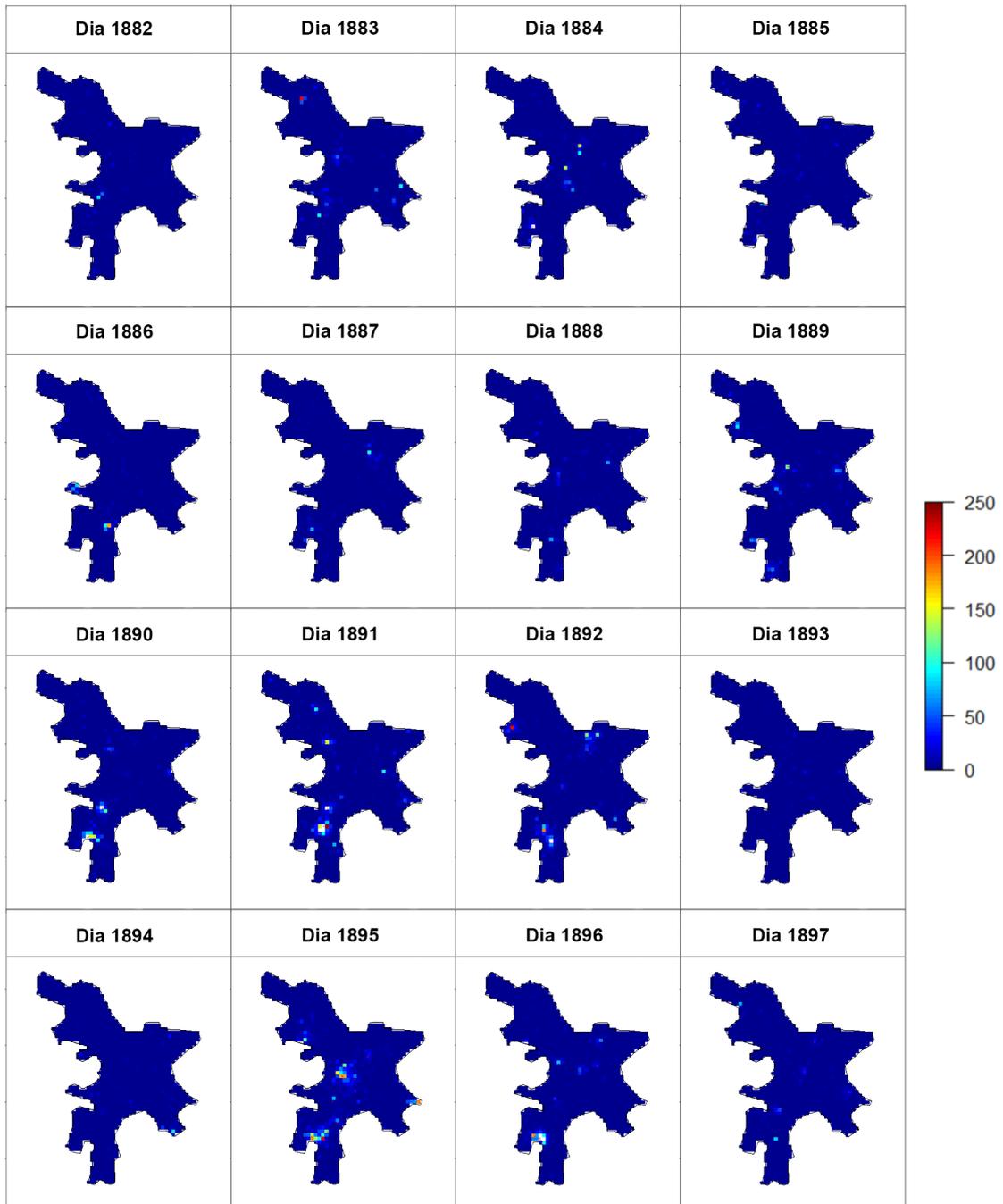


Figura 4 – Erro padrão de predição para os dias 13 de Março à 28 de Março.

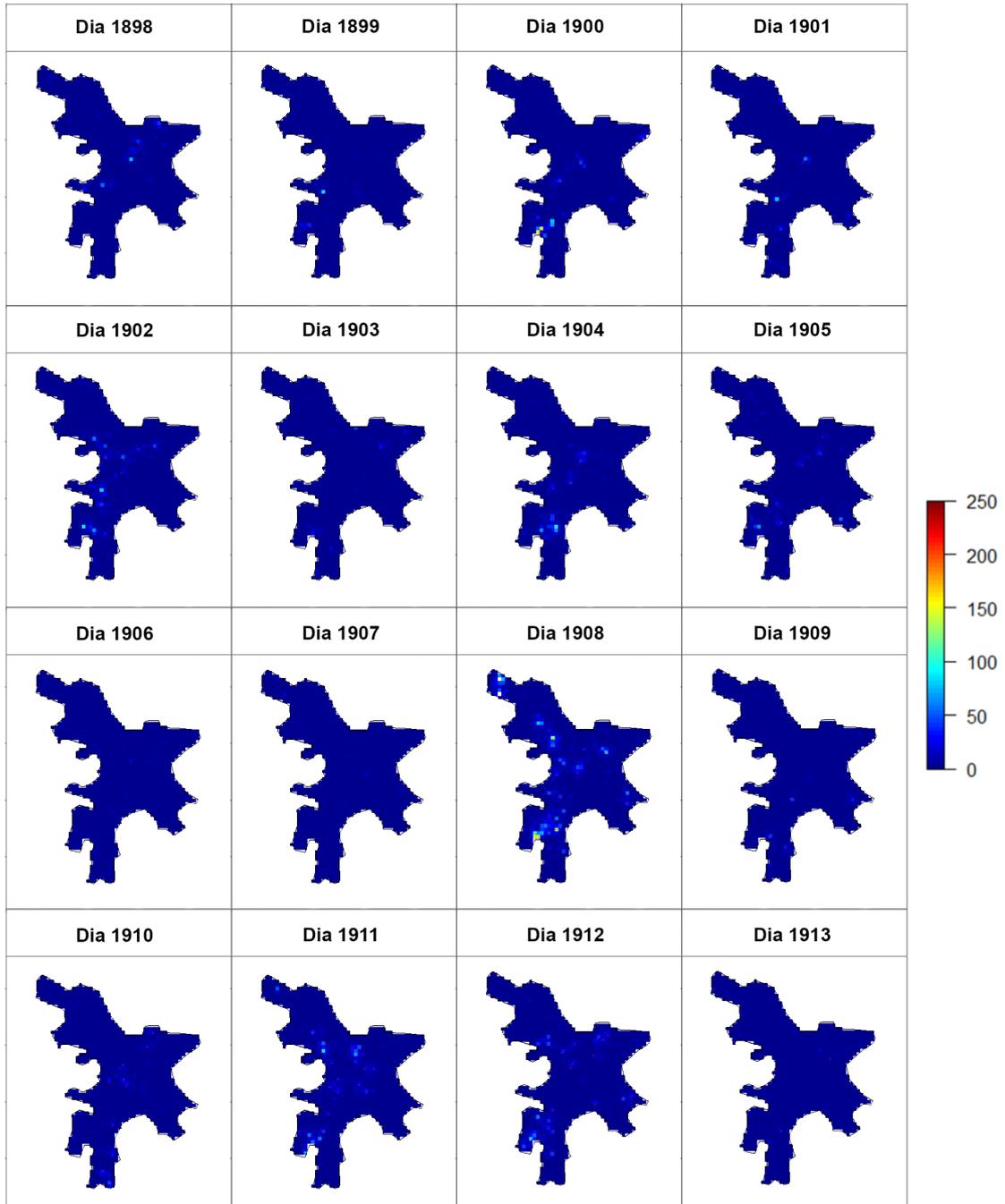


Figura 5 – Erro padrão de predição para os dias 29 de Março à 13 de Abril.

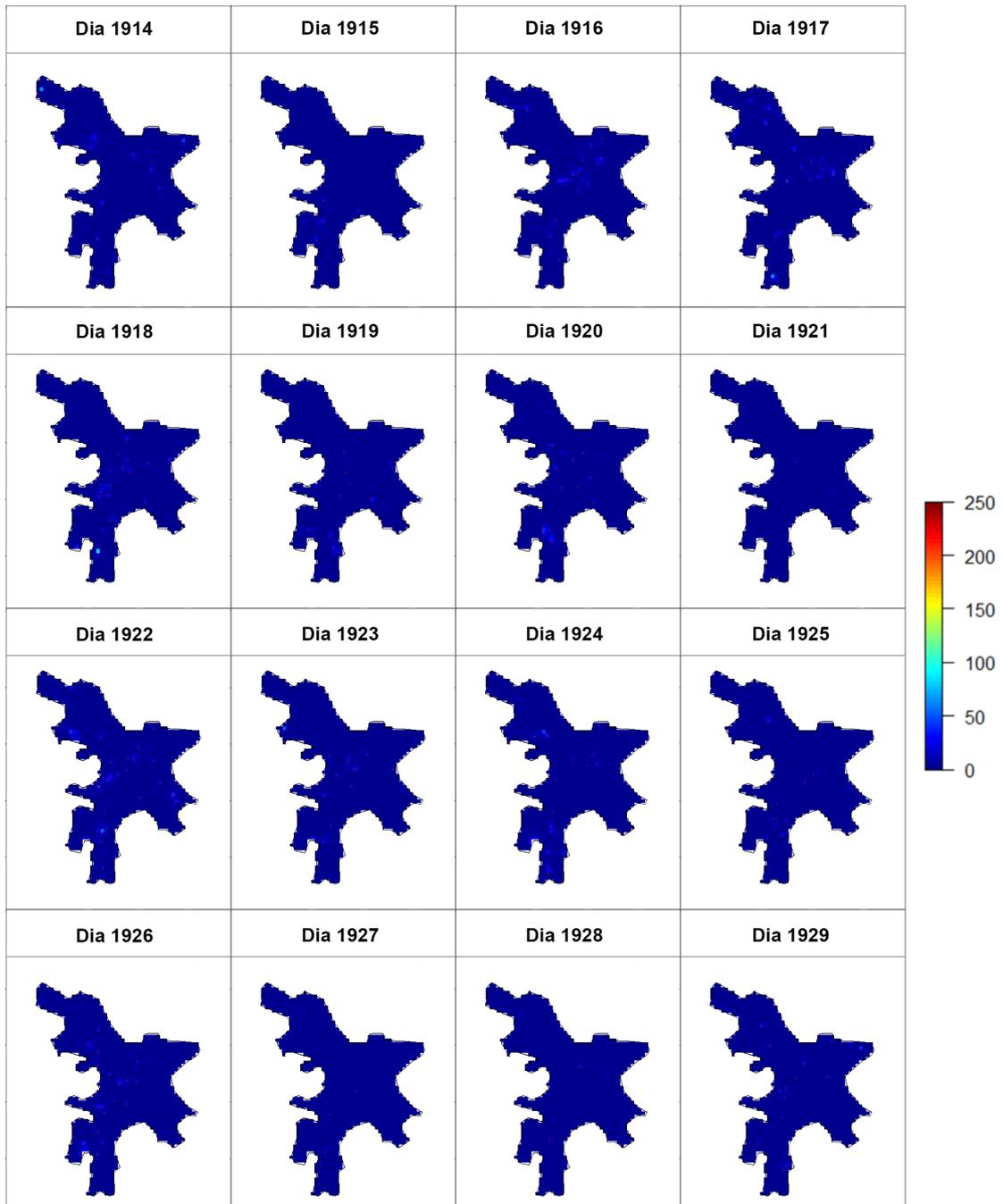


Figura 6 – Erro padrão de predição para os dias 14 de Abril à 29 de Abril.

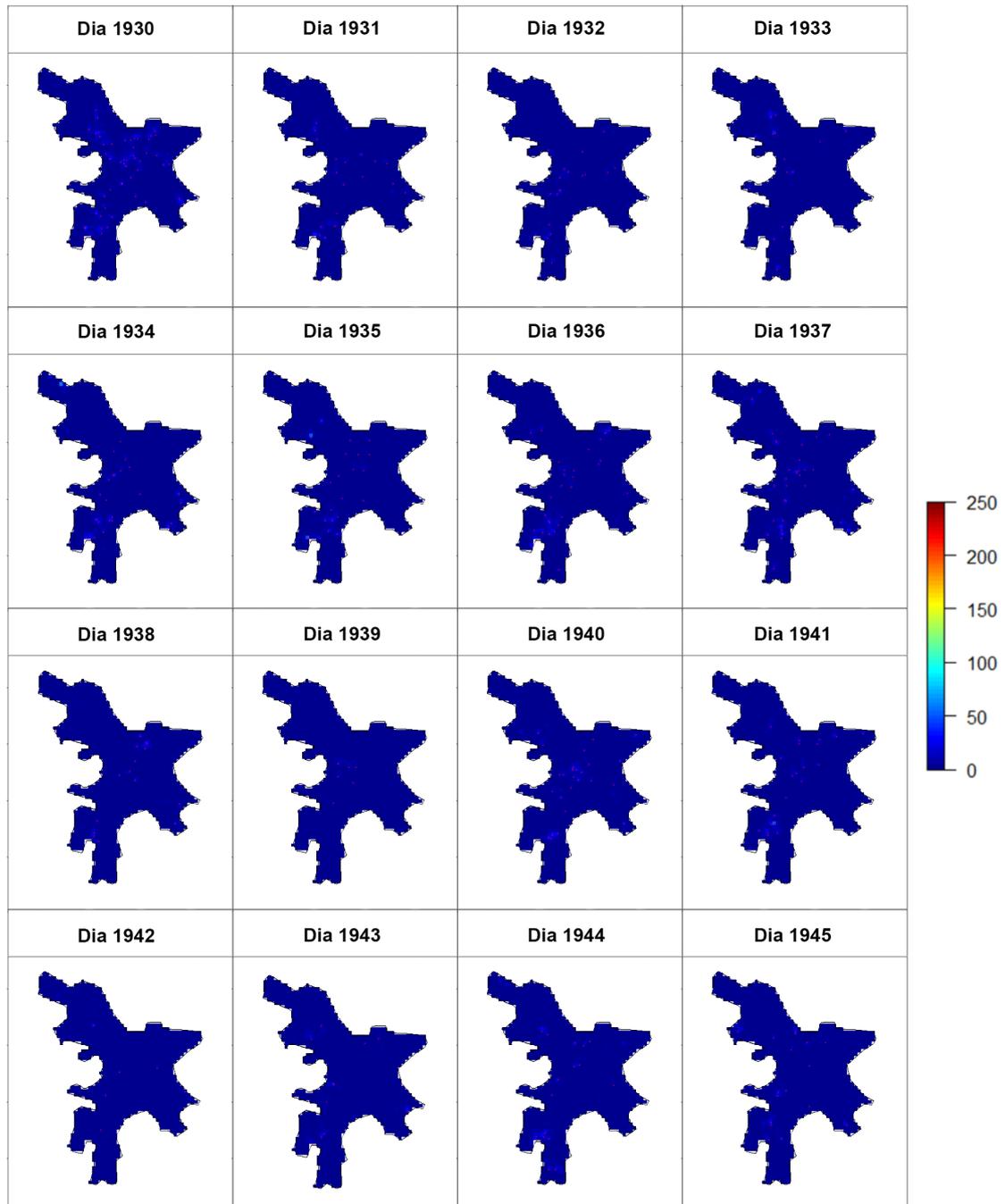


Figura 7 – Erro padrão de predição para os dias 30 de Abril à 15 de Maio.

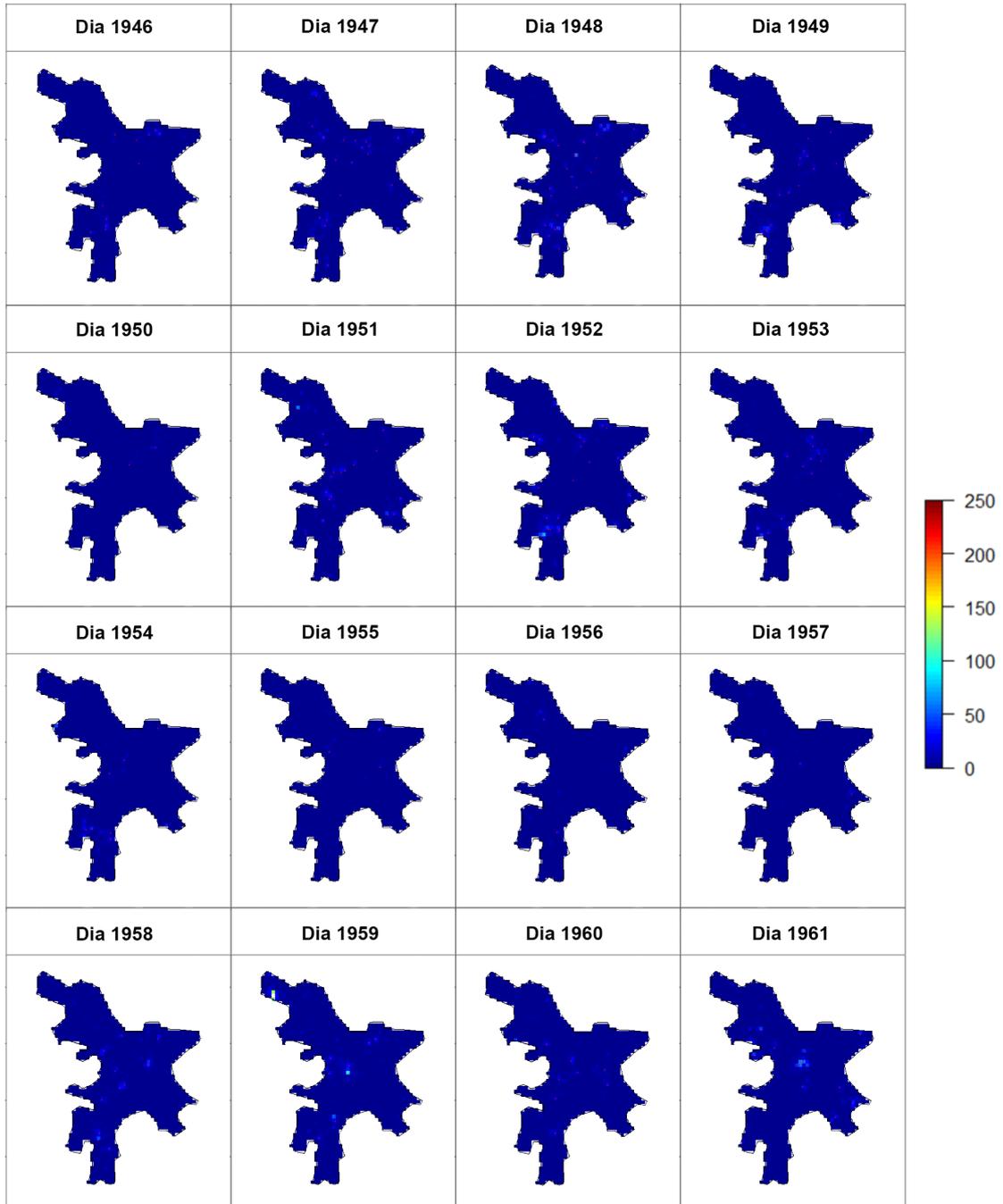


Figura 8 – Erro padrão de predição para os dias 16 de Maio à 31 de Maio.

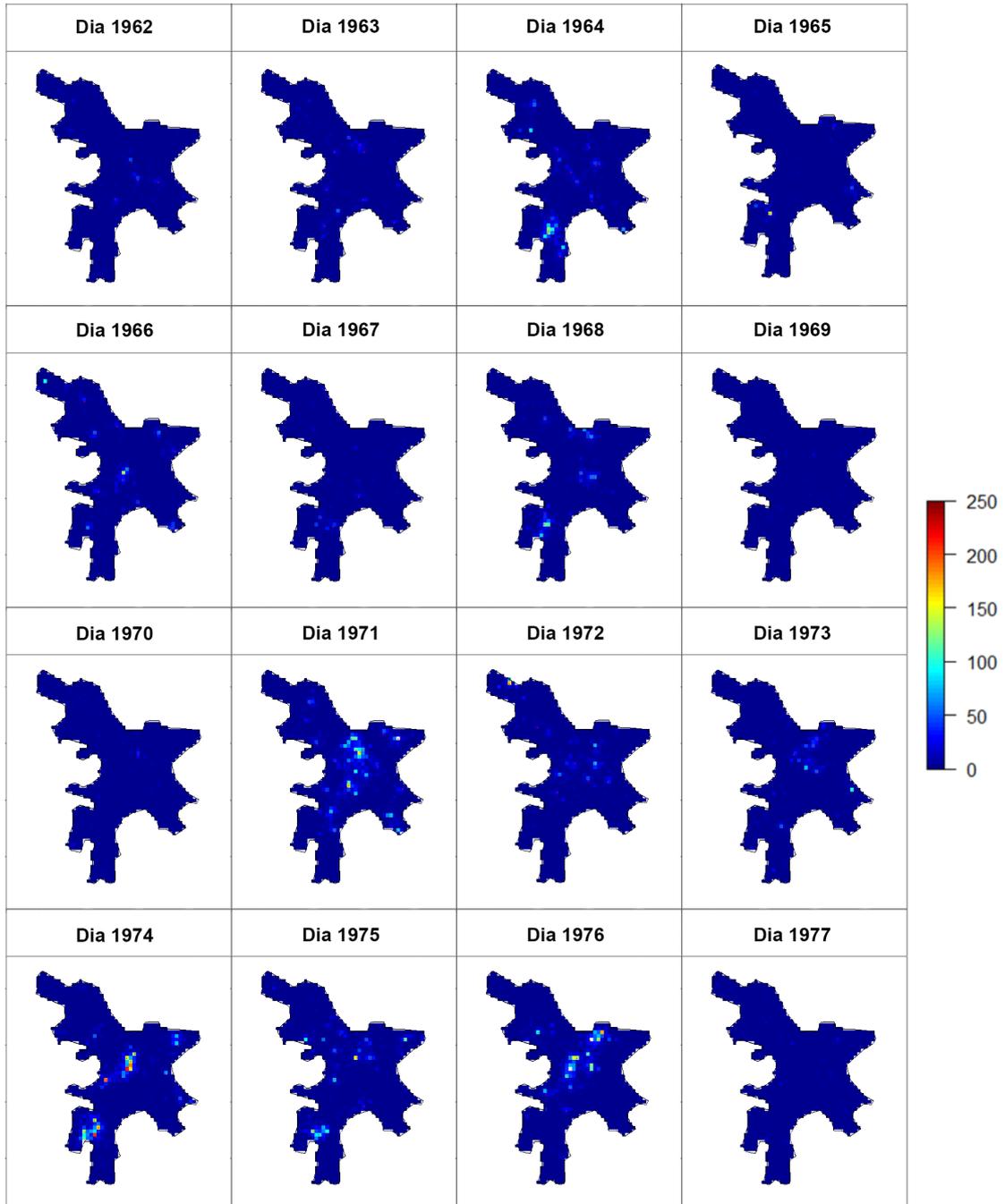


Figura 9 – Erro padrão de predição para os dias 01 de Junho à 16 de Junho.

