



**RODRIGO FERREIRA DE ABREU**

**VARIOGRAMA MARCADO NA ANÁLISE DE PROCESSOS  
PONTUAIS ESPAÇO-TEMPORAIS COM MARCAS**

**LAVRAS – MG**

**2020**

**RODRIGO FERREIRA DE ABREU**

**VARIOGRAMA MARCADO NA ANÁLISE DE PROCESSOS PONTUAIS  
ESPAÇO-TEMPORAIS COM MARCAS**

Tese 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 Doutor.

Prof. DSc. João Domingos Scalon  
Orientador

**LAVRAS – MG  
2020**

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

Abreu, Rodrigo Ferreira de.

Variograma marcado na análise de processos pontuais espaço-  
temporais com marcas / Rodrigo Ferreira de Abreu. - 2020.

93 p.

Orientador(a): João Domingos Scalon.

Tese (doutorado) - Universidade Federal de Lavras, 2020.

Bibliografia.

1. Estatística espacial. 2. Variograma. 3. Focos de calor.
  4. Poteência Radiativa do Fogo. I. Scalon, João Domingos.
- II. Título.

**RODRIGO FERREIRA DE ABREU**

**VARIOGRAMA MARCADO NA ANÁLISE DE PROCESSOS PONTUAIS  
ESPAÇO-TEMPORAIS COM MARCAS  
MARKED VARIOGRAM IN THE ANALYSIS OF SPATIO-TEMPORAL  
PROCESSES WITH MARKS**

Tese 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 obtenção do título de Doutor.

APROVADA em 24 de Novembro de 2020

Prof. DSc. Marcelo Ângelo Cirilo	DES-UFLA
Prof. DSc. Renato Ribeiro de Lima	DES-UFLA
Prof. DSc. Peter de Matos Campos	DEMAT-UFSJ
Prof. DSc. Fernando Luiz Pereira de Oliveira	DEEST-UFOP



Prof. Dr. João Domingos Scalon  
Orientador

**LAVRAS –MG  
2020**

*À todas as pessoas que mesmo tendo todos os motivos para desistir, mesmo o mundo o tempo  
todo querendo subjugar-las, preservaram em meio à adversidade e tentam fazer desse mundo  
um lugar melhor.  
Dedico.*

## **AGRADECIMENTOS**

À Deus Pai, Filho e Espírito Santo sempre em primeiro lugar, por meio de quem todas as coisas são possíveis.

À Maria mãe de Cristo, pela sua proteção.

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

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

À minha esposa Fiorita, pelo amor e apoio incondicional.

Ao professor Dr. João Domingos Scalon, pela orientação e confiança no meu trabalho.

À banca examinadora, pelas contribuições.

Aos colegas do doutorado Rafael, Vânia e Patrícia, companheiros de empreitada.

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

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

À Universidade Federal de Lavras e ao Departamento de Estatística, pela oportunidade de alcançar voos mais altos do que um dia imaginei poder.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

*“No futuro, o pensamento estatístico será tão necessário  
para a cidadania eficiente como saber ler e escrever.”  
Herbert George Wells*

## RESUMO

A estatística espacial vem sendo amplamente estudada e aplicada a diversos fenômenos, como eventos sísmicos, incêndios florestais, infecção por uma determinada doença, localização de plantas de determinada espécie em uma floresta, crimes em uma cidade, entre outros. Esses tipos de fenômenos podem ser caracterizados como sendo realizações de processos pontuais, em que cada ocorrência observada pode ser representada como um ponto no espaço  $d$ -dimensional. Se  $d = 1$  tem-se um processo pontual unidimensional ou um processo pontual temporal (PPT), se  $d = 2$  tem-se um processo pontual espacial (PPE), e se há tanto a informação do espaço como o tempo, o que se tem é um processo pontual espaço-temporal (PPET). Em muitos casos, além da localização e tempo dos eventos, há ainda uma informação adicional denominada *marca*, que é um atributo específico do evento observado, como por exemplo, a altura de cada árvore. Nesse contexto, o processo pontual é dito *marcado*, e uma ferramenta que fornece valiosa informação sobre a estrutura de correlação das marcas é o variograma marcado, que mede o grau de correlação entre elas. Além disso, o variograma marcado também pode ser utilizado como ferramenta para análise de um PPET não marcado, transformando-o em dois processos pontuais marcados, ora os tempos como marcas das localizações, ora as localizações como marcas dos tempos. O variograma marcado, no entanto, não foi explorado ainda como uma ferramenta para a análise de dados provenientes de um processo pontual espaço-tempo marcado (PPETM). Diante desse contexto, o objetivo principal desse trabalho foi apresentar uma versão espaço-temporal do variograma marcado, para análise conjunta das componentes espacial e temporal de dados provenientes de um PPETM, para verificar se essa abordagem apresenta ganhos em relação à uma análise considerando essas componentes separadamente. O método proposto foi utilizado em dados simulados sob o modelo nulo de rotulagem aleatória (marcas independentes e identicamente distribuídas (iid) e independentes dos pontos), para verificar a compatibilidade entre o resultado teórico e o calculado, e em dados reais de focos de calor obtidos por sensoriamento remoto por satélites. A marca dos focos de calor é Potência Radiativa do Fogo, que é uma medida da energia radiante do fogo durante o processo de queima, e está associada a quantidade de biomassa queimada em um incêndio, por exemplo. Os dados foram obtidos junto ao Instituto Nacional de Pesquisas Espaciais - INPE e foram referentes aos anos de 2018 e 2019, para o município de Itapuã do Oeste - RO, na região amazônica. Verificou-se que a estrutura gráfica do variograma marcado proposto é condizente com o resultado teórico para o modelo nulo de rotulagem aleatória, e que na análise dos dados de focos de calor, o resultado apresentou ganhos em relação à análise marginal das componentes espacial e temporal, uma vez que foi possível identificar uma estrutura de correlação para a variabilidade das marcas, em várias escalas espaço-temporais que não foram observadas ao considerar os processos separadamente.

**Palavras-chave:** Estatística espacial, Variograma, Focos de calor, Potência Radiativa do Fogo

## ABSTRACT

Spatial statistics has been widely studied and applied to several phenomena, such as seismic events, forest fires, infection by a specific disease, location of plants of a specific species in a forest, crimes in a city, among others. These types of phenomena can be characterized as realizations of point processes, in which each observed occurrence can be represented as a point in two-dimensional space. If  $d = 1$  there is a one-dimensional point process or a temporal point process (TPP), if  $d = 2$  there is a spatial point process (SPP), and if there is both information about space and time, which has is a space-time point process (STPP). In many cases, in addition to the location and timing of the events, there is also additional information called a mark, which is a specific attribute of the observed event, such as the height of each tree. In this context, the point process is said to be marked, and a tool that provides valuable information about the correlation structure of the marks is the marked variogram, which measures the degree of correlation between them. In addition, the marked variogram can also be used as a tool for the analysis of an unmarked STPP, transforming it into two marked point processes, treating the times as marks of the locations and after, the locations as marks of the times. The marked variogram, however, has not yet been explored as a tool for the analysis of data from a space-time marked point process (STMPP). Given this context, the main objective of this work was to present a spatio-temporal version of the marked variogram, for joint analysis of the spatial and temporal components of data from a STMPP, to verify whether this approach presents gains in relation to an analysis considering these components separately. The proposed method was used in simulated data under the null model of random labeling (independent and identically distributed marks (iid) and independent of points), to verify the compatibility between the theoretical and the calculated result, and in real data of heat points obtained by remote sensing by satellites. The heat point mark is Radiative Fire Power, which is a measure of the radiant energy of the fire during the burning process, and is associated with the amount of biomass burned in a fire, for example. The data were obtained from the National Institute for Space Research - INPE and were for the years 2018 and 2019, for the municipality of Itapuã do Oeste - RO, in the Amazon region. It was found that the graphic structure of the proposed marked variogram is consistent with the theoretical result for the null model of random labeling, and that in the analysis of the data of heat points, the result showed gains in relation to the marginal analysis of the spatial and temporal components, since it was possible to identify a correlation structure for the variability of the marks, in several space-time scales that were not observed when considering the processes separately.

**Keywords:** Spatial statistics, Variogram, Heat points, Fire Radiative Power

## LISTA DE FIGURAS

Figura 2.1 – Representação gráfica de padrões pontuais com diferentes intensidades . . .	16
Figura 2.2 – Localizações de terremotos e as bordas das placas tectônicas ao redor do mundo . . . . .	17
Figura 2.3 – Representação gráfica da estimação da intensidade usando função kernel . . .	20
Figura 2.4 – Diferentes realizações de um processo pontual . . . . .	22
Figura 2.5 – Função $K$ para diferentes realizações de PPE (linha contínua) e o esperado teórico para um PPH (linha tracejada). . . . .	25
Figura 2.6 – Função $K$ para diferentes realizações de PPE com os respectivos envelopes construídos a partir de 199 simulações de um PPH. . . . .	26
Figura 2.7 – Pontos do PPET que contribuem para o calculo da kernel espaço-temporal .	30
Figura 2.8 – Aplicação do suavizador de Nadaraya-Watson aos dados <i>longleaf</i> . . . . .	39
Figura 2.9 – Variograma marcado para o conjunto de dados <i>longleaf</i> obtido com o pacote <i>spatstat</i> . . . . .	41
Figura 2.10 – Pinheiros em um estande de floresta na região da Saxônia, Alemanha, marcados com o diâmetro à altura do peito. . . . .	42
Figura 2.11 – Variograma marcado para o conjunto de dados <i>spruces</i> obtido com o pacote <i>spatstat</i> . . . . .	43
Figura 2.12 – Gráfico do modelo gaussiano teórico . . . . .	52
Figura 2.13 – Gráfico do modelo gaussiano teórico com efeito pepita maior que zero . . .	52
Figura 3.1 – Localização do município de Itapuã do Oeste - RO e a Flona Jamari . . . . .	57
Figura 3.2 – Gráfico da precipitação média e da temperatura média mensais para Itapuã do Oeste - RO. . . . .	58
Figura 4.1 – Distribuição mensal dos focos de calor em Itapuã do Oeste - RO entre 2018 e 2019 . . . . .	61
Figura 4.2 – Distribuição espacial dos focos de calor em Itapuã do Oeste - RO entre 2018 e 2019. As unidades dos mapas são dadas em quilômetros. . . . .	62
Figura 4.3 – Histograma dos valores de FRP associados a cada foco de calor . . . . .	63
Figura 4.4 – Distribuição dos valores de FRP associados a cada foco de calor em Itapuã do Oeste - RO entre 2018 e 2019 . . . . .	63
Figura 4.5 – Distribuição dos espacial do valores de FRP associados a cada foco de calor em Itapuã do Oeste - RO entre 2018 e 2019 . . . . .	64

Figura 4.6 – Valores de FRP espacialmente suavizados pelo estimador de Nadaraya-Watson para os focos de calor em Itapuã do Oeste - RO entre 2018 e 2019 .....	65
Figura 4.7 – Resultados da análise espaço-temporal para o período 2018-2019 utilizando o modelo gaussiano e os envelopes de simulação. ....	67
Figura 4.8 – Variogramas marginais $\hat{\gamma}_S(r)$ e $\hat{\gamma}_T(t)$ estimados, e os envelopes obtidos a partir de 199 permutações sob o modelo nulo de rotulagem aleatória, para os anos de 2018 a 2019. ....	72
Figura 4.9 – Variogramas médios obtidos para o modelo nulo de rotulagem aleatória para 100 permutações das marcas em diferentes estruturas espaço-temporais. ....	74
Figura 4.10 – Variogramas médios obtidos para o modelo nulo de rotulagem aleatória para 100 permutações das marcas em diferentes estruturas espaço-temporais. ....	75
Figura 4.11 – Variogramas médios obtidos para o modelo nulo de rotulagem aleatória para 100 permutações das marcas em um PPNH espaço-temporal. ....	76
Figura 4.12 – Resultado de $\hat{\gamma}_{ST}(r, t)$ aplicado aos dados de focos de calor . . . . .	77

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>14</b>
<b>2.1</b>	<b>Processos Pontuais</b>	<b>14</b>
<b>2.2</b>	<b>Processos Pontuais Espaciais</b>	<b>15</b>
<b>2.2.1</b>	<b>Efeitos de primeira ordem</b>	<b>16</b>
<b>2.2.1.1</b>	<b>Estimação da instensidade usando suavização de kernel</b>	<b>18</b>
<b>2.2.2</b>	<b>Efeitos de segunda ordem</b>	<b>21</b>
<b>2.3</b>	<b>Processos Pontuais Temporais</b>	<b>27</b>
<b>2.4</b>	<b>Processos Pontuais Espaço-Temporais</b>	<b>29</b>
<b>2.4.1</b>	<b>Propriedades de primeira ordem</b>	<b>29</b>
<b>2.4.2</b>	<b>Propriedades de segunda ordem</b>	<b>31</b>
<b>2.5</b>	<b>Processos Pontuais Marcados</b>	<b>36</b>
<b>2.5.1</b>	<b>Propriedades de primeira ordem</b>	<b>38</b>
<b>2.5.2</b>	<b>Propriedades de segunda ordem</b>	<b>40</b>
<b>2.5.2.1</b>	<b>O variograma marcado</b>	<b>41</b>
<b>2.6</b>	<b>Processos Pontuais Espaço-Temporais Marcados</b>	<b>44</b>
<b>2.6.1</b>	<b>Propriedades de primeira e segunda ordem</b>	<b>45</b>
<b>2.6.2</b>	<b>Estacionariedade de intensidade de segunda ordem reponderada</b>	<b>47</b>
<b>2.6.3</b>	<b>Função <math>K</math> não homogênea marcada</b>	<b>47</b>
<b>2.7</b>	<b>Análise de PPET usando o variograma marcado</b>	<b>48</b>
<b>2.7.1</b>	<b>Tempo e espaço como marcas</b>	<b>48</b>
<b>2.7.2</b>	<b>O variograma marcado espacial</b>	<b>49</b>
<b>2.7.3</b>	<b>O variograma marcado temporal</b>	<b>50</b>
<b>2.7.4</b>	<b>Estimação dos variogramas marcados <math>\gamma_{sp}(r)</math> e <math>\gamma_e(t)</math></b>	<b>51</b>
<b>2.7.5</b>	<b>O modelo gaussiano</b>	<b>51</b>
<b>2.7.6</b>	<b>Considerações sobre os variogramas <math>\gamma_{sp}(r)</math> e <math>\gamma_e(t)</math></b>	<b>53</b>
<b>2.8</b>	<b>Focos de calor e a potência radiativa do fogo (FRP)</b>	<b>54</b>
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>57</b>
<b>3.1</b>	<b>O conjunto de dados</b>	<b>57</b>
<b>3.1.1</b>	<b>Dados reais</b>	<b>57</b>
<b>3.1.2</b>	<b>Dados simulados</b>	<b>58</b>

3.2	<b>Análise do PPET usando o variograma marcado . . . . .</b>	58
3.3	<b>Software . . . . .</b>	59
4	<b>Resultados e discussão . . . . .</b>	61
4.1	<b>Análises relativas às componentes espacial e temporal dos dados de focos de calor . . . . .</b>	61
4.1.1	<b>Análise descritiva . . . . .</b>	61
4.1.2	<b>Análise do PPET usando o variograma marcado . . . . .</b>	66
4.2	<b>Proposta de análise de um PPETM utilizando o variograma marcado . . .</b>	68
4.2.1	<b>Análise marginal do PPETM . . . . .</b>	68
4.2.2	<b>Análise conjunta do PPETM . . . . .</b>	70
4.2.3	<b>Resultado da análise marginal . . . . .</b>	72
4.2.4	<b>Resultado da análise conjunta . . . . .</b>	73
5	<b>Conclusões . . . . .</b>	79
	<b>REFERÊNCIAS . . . . .</b>	80
	<b>APENDICE A – Implementação em R do variograma marcado espaço-temporal <math>\hat{\gamma}_{ST}(r,t)</math> . . . . .</b>	87
	<b>APENDICE B – Construção dos envelopes para <math>\hat{\gamma}_{ST}(r,t)</math> a partir de 199 simulações do processo pontual sob a hipótese nula de rotulagem aleatória e taxa de erro <math>\alpha = 0,05</math> . . . . .</b>	91

## 1 INTRODUÇÃO

A teoria de processos pontuais é uma área da Estatística Espacial cujo objeto de análise são dados que contém informações de onde determinados eventos ocorrem em uma região contínua de algum espaço matemático. Essa região pode ser tanto o o espaço unidimensional (tempo ou linha reta), quanto o bidimensional, ou ainda o produto cartesiano entre conjuntos, como o espaço-tempo, por exemplo.

Muitos tipos de fenômenos podem ser estudados considerando a ótica de processos pontuais, dos quais pode-se destacar: eventos sísmicos (OGATA, 1998; SCHOENBERG, 2003; VERE-JONES; DENG, 1988), incêndios florestais (ARAGÓ et al., 2016; MØLLER; DÍAZ-AVALOS, 2010; PODUR; MARTELL; CSILLAG, 2003; TURNER, 2009), infecção por uma determinada doença (DIGGLE, 2013; LAWSON; ZHOU, 2005; PEREZ; WARD; CARPENTER, 2004), localização de plantas de determinada espécie em uma floresta (COMAS; MATEU, 2007; PENTTINEN; STOYAN; HENTTONEN, 1992; STOYAN; PENTTINEN, 2000), aglomerados de galáxias (BEISBART; KERSCHER, 2000), crimes em uma cidade (GERVINI, 2016; MOHLER et al., 2011), entre outros.

O tipo de processo a ser considerado depende da estrutura dos dados disponíveis. Por exemplo, se as informações são das localizações de acidentes ao longo de uma rodovia, o processo pontual pode ser tratado como unidimensional, mas, se as informações são dos tempos em que os acidentes ocorreram, tem-se um processo pontual temporal (PPT). Se as informações são de localizações das ocorrências de crimes em uma cidade, tratamos o processo como sendo espacial (PPE). Mas, se além das localizações o momento do crime também foi registrado, pode-se considerar então, um processo pontual espaço-temporal (PPET).

Pode-se considerar ainda outras estruturas possíveis para um processo pontual, como o caso do processo pontual marcado (PPM), em que, além da localização, há uma informação adicional que é própria do evento, como por exemplo, a altura de uma árvore, o tipo de um crime ou a área de um incêndio. Essa informação adicional (marca) pode estar anexada a um ponto que possui apenas a sua localização espacial ou temporal, ou ainda espaço-temporal, de modo que, nesse último caso tem-se um processo pontual espaço-temporal marcado (PPETM). Para cada um dos tipos de processos pontuais há varias ferramentas de análise (vide por exemplo, (BADDELEY; RUBAK; TURNER, 2015; DIGGLE, 2013; ILLIAN et al., 2008)). Para um PPETM, no entanto, a literatura disponível não é tão extensa como nos outros casos. Isso faz com que a quantidade de ferramentas de análise de um PPETM disponível seja menor.

Os principais trabalhos sobre PPETM são dedicados a modelar os processos a partir funções de intensidade como em Assunção e Maia (2007), Cronie e Särkkä (2011), Holden, Sannan e Bungum (2003), Ogata (1998), Särkkä e Renshaw (2006), Schoenberg (2004) e, portanto, fazem uso na análise das propriedades de primeira ordem. O trabalho de Iftimi, Cronie e Montes (2016), por outro lado, apresenta uma análise para PPETM cujo principal objetivo é desenvolver estimadores de estatísticas sumárias para descrição das propriedades de segunda ordem.

Uma ferramenta que fornece valiosa informação sobre a estrutura de correlação das marcas é o variograma marcado, que mede o grau de correlação entre as marcas de um processo marcado espacial (ou temporal), à medida que os pontos se tornam mais distantes (CRESSIE, 1993; ILLIAN et al., 2008; STOYAN; WÄLDER, 2000). O variograma marcado fornece informações importantes, que incluem a rapidez com que a correlação entre marcas diminui com o aumento da distância, o alcance dessa correlação, se pontos próximos tendem a ter marcas parecidas ou não, ou mesmo se não há correlação alguma. Além disso, mesmo o variograma marcado sendo uma ferramenta para análise de um PPM, ele também pode ser utilizado como ferramenta para análise de um PPET não marcado, como propôs Stoyan et al. (2017), transformando o PPET em dois processos pontuais marcados, ora os tempos como marcas das localizações, ora as localizações como marcas dos tempos. O variograma marcado, no entanto, não foi explorado ainda como uma ferramenta para a análise das componentes espacial e temporal conjuntamente, de dados provenientes de um PPETM.

Diante desse contexto, o presente estudo tem o objetivo desenvolver novos métodos para análise para um PPETM tendo como base o variograma marcado, estendendo a abordagem do caso puramente espacial (ou temporal), para o espaço-tempo.

Como objetivos específicos tem-se:

- Propor uma versão espaço-temporal do variograma marcado, para a análise conjunta das componentes espacial e temporal de um PPETM;
- Aplicar os métodos propostos em dados de focos de calor obtidos por sensoriamento remoto por satélites no município de Itapuã do Oeste - RO, na região amazônica entre os anos de 2018 e 2019.

Com esse estudo espera-se identificar tendências espaço-temporais nos dados de focos de calor, bem como no padrão de severidade da queima.

Como hipótese, acredita-se que a análise conjunta de um PPETM utilizando o variograma marcado possibilita ganhos em relação a uma análise que leve em conta apenas as componentes puramente espacial ou temporal, de modo que seja possível identificar uma estrutura de correlação para a variabilidade das marcas em várias escalas espaço-temporais, o que pode não ocorrer ao considerar os processos separadamente.

## 2 REFERENCIAL TEÓRICO

### 2.1 Processos Pontuais

Há muitos tipos de análises estatísticas em que o interesse reside no comportamento de determinado fenômeno em relação à sua distribuição espacial ou temporal, ou nos dois aspectos simultaneamente (espaço-tempo), como por exemplo, em dados de ciências florestais, epidemiologia, sismologia, entre outros. Para que se faça o uso de ferramentas de análise adequadas é preciso que se tenha uma compreensão do tipo de dados a ser analisado. Considerando os dados que apresentam uma estrutura espacial, Cressie (1993) aponta que podem ser analisados a partir das seguintes perspectivas:

- **Geoestatística:** São dados de superfície contínua em que o fenômeno está distribuído continuamente sobre uma superfície, como, por exemplo, teor de um determinado nutriente no solo, temperatura, estimação da extensão de uma jazida mineral, etc.
- **Dados de áreas (lattice):** São dados indexados a sub-regiões (polígonos) que constituem uma partição de um domínio contínuo. São exemplos os dados agregados por município, setor censitário, bairros, etc, em que não se sabe exatamente onde os eventos ocorrem mas dispõe-se de um valor agregado a uma área.
- **Dados de processos pontuais:** São dados em que a informação é a própria localização do evento. Ocorrência de crimes, epicentros de terremotos, epidemias e localização de plantas de uma determinada espécie são alguns exemplos desse tipo de dados.

Dados que apresentam essas estruturas são objeto de estudo da Estatística Espacial, que pode ser caracterizada como o conjunto de conceitos e métodos de análise de dados, em que a localização geográfica é utilizada explicitamente, podendo ser modelada como um processo estocástico (CRESSIE, 1993). Um processo estocástico é, conforme Snyder e Miller (1991), uma família ou uma coleção de variáveis aleatórias que podem ser identificadas (indexadas), por um determinado parâmetro  $\theta$ , que toma valores em um conjunto de parâmetros  $\Theta$  chamado de conjunto de índices do processo. Se  $\Theta$  é um conjunto contável  $\{\theta_1, \theta_2, \dots, \theta_n\}$ , tem-se então, segundo os autores, um processo de parâmetro discreto. Se  $\Theta \subset \mathbb{R}^d$  o processo é dito de parâmetro contínuo. Esse parâmetro pode ser o tempo, o espaço ou ainda uma combinação desses dois (espaço-tempo). A partir dessa definição, um processo estocástico pode ser representado no espaço, no tempo e no espaço-tempo, respectivamente por:

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

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

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

Nas expressões acima tem-se que  $S$  é uma sub-região do espaço  $d$ -dimensional, em geral, considerada como sendo finita. Para fins práticos, tem-se usualmente  $d = 2$  ou  $d = 3$ , e ao longo deste trabalho, será considerado o caso bidimensional, de modo que  $\mathbf{s} = (x, y)'$ .  $T$  é um subconjunto de  $\mathbb{R}$  e corresponde a um intervalo de tempo, podendo ser considerado como o tempo  $t$  decorrido após algum instante inicial  $t_0$ ,  $T = \{t : t \geq t_0\}$ .

Para o desenvolvimento deste trabalho, os dados são considerados a partir da perspectiva de um processo pontual, que pode ser entendido, conforme Snyder e Miller (1991), como um modelo matemático para um fenômeno físico, caracterizado por pontos cujas localizações se distribuem aleatoriamente em uma região contínua, como por exemplo, um intervalo de tempo ou o espaço uni ou multidimensional. Se  $X$  denota essa região contínua, uma realização do processo pontual em  $X$ , nada mais é do que um conjunto de pontos em  $X$ , e pode ser chamada de *configuração pontual*.

## 2.2 Processos Pontuais Espaciais

Um Processo Pontual Espacial (PPE) é um mecanismo estocástico que gera um conjunto contável finito ou infinito de pontos  $\{\mathbf{s}_i : i = 1, 2, \dots\}$ , tais que  $\mathbf{s}_i = (x, y) \in S \subset \mathbb{R}^2$ , ou seja, cada variável aleatória denota a localização de um ponto específico no espaço bidimensional. Um ponto observado na realização do processo estocástico, também é comumente chamados de *evento*, para distingui-lo de um ponto aleatório qualquer na região de estudo (DIGGLE, 2013).

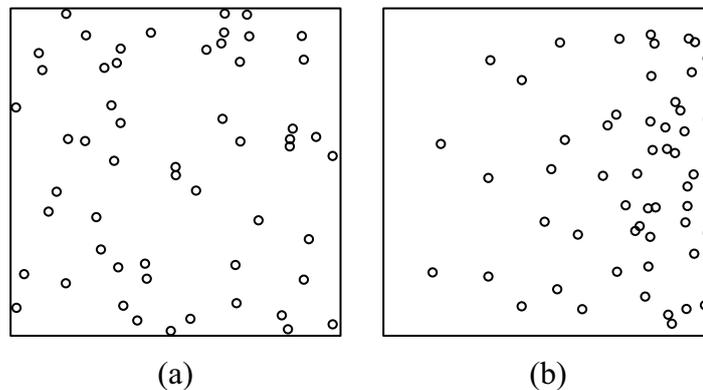
De acordo com Baddeley, Bárány e Schneider (2007), pode-se descrever os processos pontuais em termos de momentos, como o valor esperado (efeitos de primeira ordem) e variância ou covariância (efeitos de segunda ordem) de modo análogo ao que é feito para uma variável aleatória. Essas quantidades, segundo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

### 2.2.1 Efeitos de primeira ordem

De acordo com Câmara e Carvalho (2003), os efeitos de primeira ordem são aqueles ditos globais ou de larga escala, e correspondem a variações no valor médio do processo no espaço. Na análise desses efeitos, o interesse está na intensidade  $\lambda$  do processo, ou seja, no número de eventos por unidade de área. Dentro da região de estudo, a intensidade pode ser homogênea ou não homogênea. No primeiro caso, intensidade homogênea quer dizer constante em toda a parte, ou seja, o número esperado de eventos em qualquer sub-região da área de estudo é o mesmo. Por outro lado, não homogênea significa que o número esperado de eventos varia dentro da área de estudo.

Na Figura 2.1, temos em (a) uma configuração pontual com intensidade constante em toda a região, ou seja, intensidade homogênea. Em (b) temos uma configuração pontual com intensidade não homogênea, ou seja, existe uma tendência nas localizações dos eventos. Um exemplo desse caso poderia ser relacionado a uma região com declividade, ou então a um rio que passa por uma das laterais da região, propiciando que haja uma maior ocorrência de determinada espécie arbórea nas proximidades do rio.

Figura 2.1 – Representação gráfica de padrões pontuais com diferentes intensidades



Legenda: (a) padrão pontual com intensidade homogênea. (b) padrão pontual com intensidade não homogênea.

Fonte: Do autor (2020)

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

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

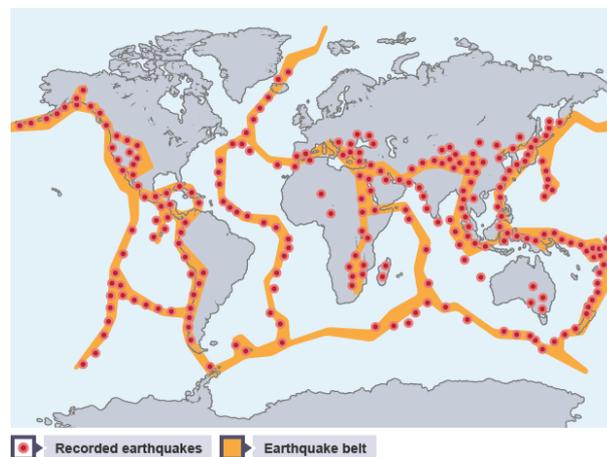
em que  $ds$  define uma pequena região em torno da localização  $s$ ,  $|ds|$  é a sua área e  $N(ds)$  é número de eventos localizados dentro de  $ds$ .

A partir da função intensidade pode-se pensar no número esperado de eventos, dentro de uma região  $B$  onde o processo pontual  $X$  ocorre. Conforme Baddeley, Rubak e Turner (2015), consideremos uma pequena vizinhança em torno de uma localização  $s$ , com uma pequena área  $a$ . Temos que o número esperado de pontos nessa vizinhança será  $\lambda(s)a$ . A partir disso, imaginemos que a região  $B$  possa ser dividida em pixels e se calcula em cada pixel, o número esperado de pontos. O número de pontos esperado em  $B$  será então a soma de todos os valores encontrados para cada pixel. Essa soma, de modo efetivo é a integral da função intensidade dada por

$$\mathbb{E}[N(X \cap B)] = \mu(B) = \int_B \lambda(s) ds. \quad (2.5)$$

De acordo com Baddeley, Rubak e Turner (2015),  $\lambda(s)$  pode ser pensada como uma superfície ondulada, cuja altura em cada ponto do domínio representa a intensidade e o número esperado de pontos dentro de alguma região de interesse é igual ao volume abaixo da superfície nessa região. Os autores salientam que há processos pontuais em que não há uma função intensidade. Um exemplo seria o caso de um processo em que os eventos são terremotos, que ocorrem exatamente nas bordas de placas tectônicas (Figura 2.2). Devido a isso tem-se a necessidade de utilizar uma linguagem mais geral.

Figura 2.2 – Localizações de terremotos e as bordas das placas tectônicas ao redor do mundo



Fonte: <<https://app.emaze.com/@AFZFQRLO#5>>

Consideremos então, um processo pontual  $X$  qualquer. As características de intensidade são, de acordo com os autores, completamente descritas se soubermos o número esperado de

pontos

$$\Lambda(B) = \mathbb{E}[N(X \cap B)] \quad (2.6)$$

dentro de uma região  $B$ .

Se a intensidade do processo pontual é homogênea, isto é,  $\lambda(\mathbf{s}) = \lambda \geq 0$  para todo  $\mathbf{s} \in B$ , então  $\Lambda(B) = \lambda|B|$ , em que  $|B|$  é a área de  $B$ , ou seja, a intensidade  $\lambda$  é dada como  $\lambda = \frac{\mathbb{E}[n(X \cap B)]}{|B|}$  e pode ser estimada por

$$\hat{\lambda} = \frac{n}{|B|}. \quad (2.7)$$

em que  $n = N(X \cap B)$ .

Baddeley, Rubak e Turner (2015) chamam a atenção para o fato de  $\Lambda$  ser uma “função conjunto”, ou seja, uma função cujo argumento é um conjunto ou região  $B$  e que retorna um valor numérico  $\Lambda(B)$ . Os autores denominam  $\Lambda$  como a *medida média* do processo pontual  $X$ , e afirmam que isso nos permite falar sobre a intensidade de qualquer processo pontual. Outro fato, é que em alguns textos  $\Lambda$  é chamada de *medida de intensidade*, o que pode trazer alguma confusão, uma vez que os valores  $\Lambda(B)$  não são intensidades, ou seja, número de pontos por unidade de área, mas sim números médios de pontos. A medida média  $\Lambda$  é chamada *difusa* se tiver uma função de intensidade  $\lambda(\mathbf{s})$  no sentido da equação (2.5) e é dita *singular*, se os pontos estão confinados a um conjunto com área zero, como por exemplo o caso dos terremotos mencionado anteriormente.

### 2.2.1.1 Estimação da intensidade usando suavização de kernel

Para o caso da intensidade  $\lambda(\mathbf{s})$  de um processo pontual não ser constante, uma forma de se fazer a sua estimação é utilizando a suavização/alisamento de kernel. Esse é um método não paramétrico, que estima a intensidade do PPE em uma determinada região de estudo, de modo que o resultado é uma superfície contínua, em que a altura representa o valor da intensidade em cada ponto. A função kernel calcula a densidade de uma variável aleatória, e a sua relação com a intensidade de um PPE está descrita a seguir.

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

$$\widehat{f}(y_0) = \frac{1}{nh} \sum_{i=1}^n \kappa\left(\frac{y_i - y_0}{h}\right) \quad (2.8)$$

em que  $\kappa(u)$  é a densidade uniforme em  $-1 \leq u \leq 1$ ,

$$\kappa(u) = \begin{cases} 0, & \text{se } |y_i - y_0| > h \\ 1, & \text{caso contrário.} \end{cases} \quad (2.9)$$

e  $h$  é um parâmetro chamado largura de banda, que controla o alisamento da kernel na superfície espacial. De acordo com Schabenberger e Gotway (2005), se a vizinhança  $h$  é pequena,  $\widehat{f}(y_0)$  é um estimador quase não viesado de  $f(y_0)$ , mas sofre de grande variabilidade. Com o aumento da largura de banda  $h$ , a estimativa torna-se mais suave, menos variável e mais viesada. Assim, um procedimento baseado em erro quadrático médio, como validação cruzada, pode ser usado para determinar um valor apropriado para o parâmetro  $h$ .

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

$$\kappa(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \quad (2.10)$$

ou a kernel Epanechnikov (quadrática)

$$\kappa(u) = \begin{cases} \frac{3}{4}(1-u)^2, & \text{se } |u| \leq 1 \\ 0, & \text{caso contrário.} \end{cases} \quad (2.11)$$

Além dessas há várias outras funções kernel. A exigência para que uma função  $\kappa(u)$  possa servir como um kernel é que  $\int \kappa(u)du = 1$  e  $\int u\kappa(u)du = 0$ .

Para o caso de um processo pontual espacial, a obtenção da densidade produz uma estimativa da probabilidade de se observar um evento na localização  $\mathbf{s}$ . A relação entre a densidade  $f_B(\mathbf{s})$  em  $B$  e a intensidade  $\lambda(\mathbf{s})$ , de acordo com Schabenberger e Gotway (2005), é dada por:

$$\lambda(\mathbf{s}) = f_B(\mathbf{s}) \int_B \lambda(\mathbf{u})d\mathbf{u} = f_B(\mathbf{s})\mu(B). \quad (2.12)$$

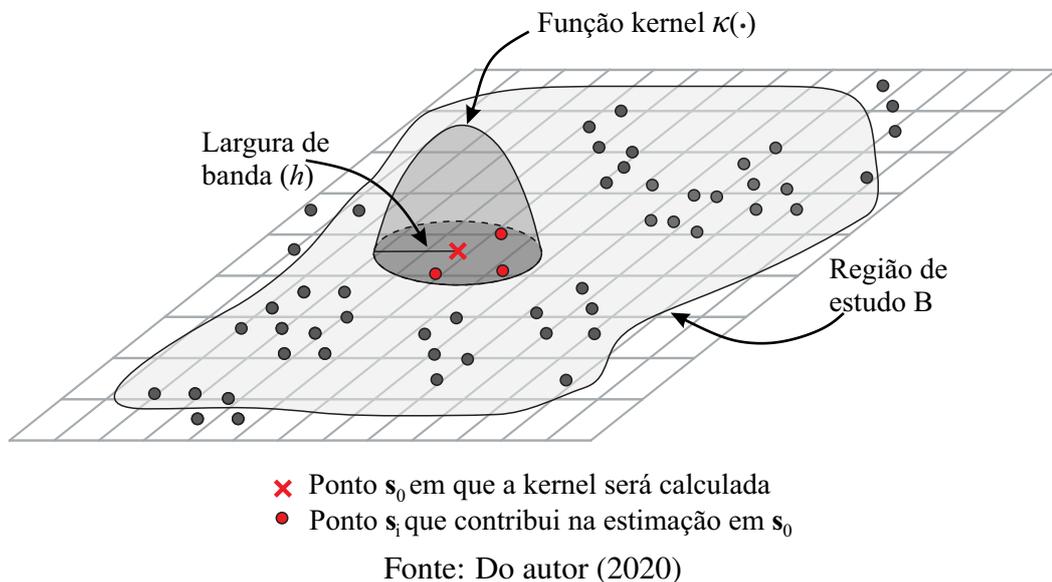
Para a construção do estimador kernel da intensidade de primeira ordem, considera-se  $\nu_d(D)$  a medida de Lebesgue de um conjunto arbitrário  $D \in \mathbb{R}^d$ . Se  $d = 1$  então  $\nu_d(D)$  é a

medida do comprimento de um intervalo, no  $\mathbb{R}^2$  é a medida da área de  $D$  e no  $\mathbb{R}^3$  é a medida do volume de  $D$ . A partir disso, para um processo em  $\mathbb{R}$ , Schabenberger e Gotway (2005) sugerem uma modificação do estimador da densidade na equação (2.8) como

$$\hat{\lambda}(s_0) = \frac{1}{v(B)h} \sum_{i=1}^n \kappa\left(\frac{s_i - s_0}{h}\right) \quad (2.13)$$

para obter o estimador kernel da intensidade de primeira ordem (Figura 2.3).

Figura 2.3 – Representação gráfica da estimação da intensidade usando função kernel



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

$$\hat{\lambda}(s_0) = \frac{1}{v(B)h_x h_y} \sum_{i=1}^n \kappa\left(\frac{x_i - x_0}{h_x}\right) \kappa\left(\frac{y_i - y_0}{h_y}\right) \quad (2.14)$$

em que  $h_x$  e  $h_y$  são larguras de banda nas direções do sistema de coordenadas.

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

As expressões acima no entanto, não consideram os efeitos de borda. Esses efeitos surgem na análise de padrões pontuais quando a região  $B$  na qual se observa o padrão é parte de uma região maior na qual o processo subjacente opera (DIGGLE, 2013). O que pode acontecer é que, eventos não observados fora de  $B$  podem interagir com eventos observados dentro de  $B$ , mas pelo fato de que os eventos não são observados, é difícil levar isso em consideração. Um estimador kernel para a intensidade corrigido para o efeito de bordas, proposto por Diggle (1985) é dado por

$$\hat{\lambda}(\mathbf{s}) = \frac{1}{p_h(\mathbf{s})} \sum_{i=1}^n \frac{1}{h^2} \kappa\left(\frac{\mathbf{s} - \mathbf{s}_i}{h}\right) \quad (2.15)$$

em que  $p_h(\mathbf{s}) = \int_B \frac{1}{h^2} \kappa\left(\frac{\mathbf{s} - \mathbf{u}}{h}\right) d\mathbf{u}$  serve como fator de correção do efeito de bordas.

### 2.2.2 Efeitos de segunda ordem

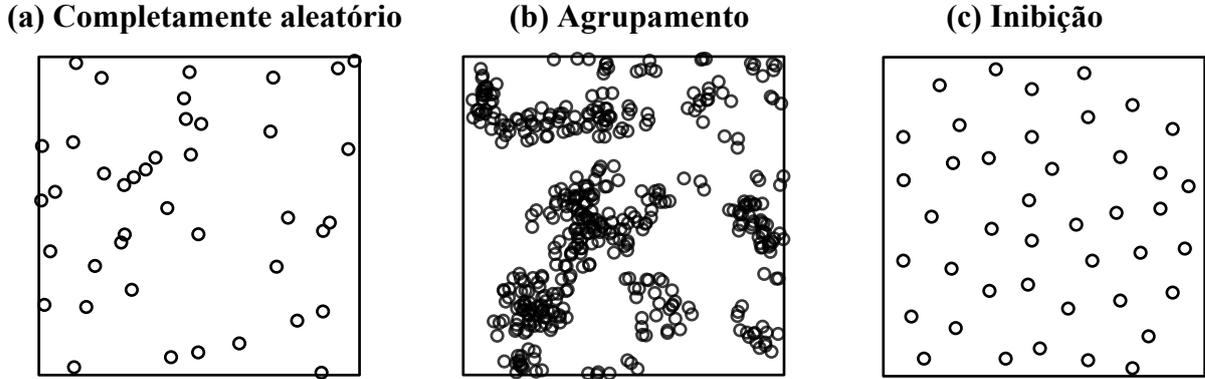
Os efeitos de segunda ordem, Câmara e Carvalho (2003) abordam que são aqueles considerados locais ou de pequena escala e representam a dependência espacial no processo que é proveniente de uma estrutura de correlação espacial. Na análise dos efeitos de segunda ordem procura-se responder se os eventos estão mais próximos ou mais distantes entre si, exibindo algum grau de relação diferente do que seria esperado no caso de completa aleatoriedade espacial (CAE). A CAE, segundo Schabenberger e Gotway (2005), ocorre se o padrão pontual atende aos seguintes critérios:

- O número médio de eventos por unidade de área, ou seja, a intensidade  $\lambda(\mathbf{s})$  é homogênea em  $B$ ,
- o número de eventos em duas sub-regiões sem sobreposição  $B_1$  e  $B_2$  são independentes,
- o número de eventos em qualquer sub-região segue distribuição Poisson.

Deste modo, os eventos distribuem-se uniformemente e independentemente ao longo do domínio e a manifestação matemática da CAE é o chamada de processo de Poisson homogêneo (PPH). Esse é um processo que não exhibe qualquer estrutura espacial e serve como hipótese nula para muitas investigações estatísticas em padrões pontuais. Os padrões de pontos observados são testados inicialmente contra a hipótese de CAE e se essa hipótese for rejeitada, então o pesquisador geralmente segue com análises mais específicas para investigar a natureza do padrão pontual espacial (SCHABENBERGER; GOTWAY, 2005).

Na Figura 2.4 estão representadas as três configurações típicas de padrões pontuais: independência (CAE), agrupamentos e inibição/regularidade.

Figura 2.4 – Diferentes realizações de um processo pontual



Fonte: Do autor (2020)

No caso de agrupamento, a ocorrência de um evento aumenta a probabilidade de ocorrer outro evento nas suas proximidades. Já para o caso de regularidade, o comportamento é oposto, a ocorrência de um evento diminui a probabilidade de que outro evento ocorra nas suas proximidades.

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

$$\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lim_{|ds_i|, |ds_j| \rightarrow 0} \frac{\mathbb{E}[N(ds_i)N(ds_j)]}{|ds_i||ds_j|}, \quad (2.16)$$

em que  $N(ds_i)$  e  $N(ds_j)$  representam o número de eventos no interior de pequenos círculos centrados nas localizações  $\mathbf{s}_i$  e  $\mathbf{s}_j$  com áreas iguais a  $|ds_i|$  e  $|ds_j|$ , respectivamente. De acordo com Bivand et al. (2008), a intensidade de segunda ordem de dois pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$  representa a probabilidade de qualquer par de eventos ocorrer nas vizinhanças de  $\mathbf{s}_i$  e  $\mathbf{s}_j$ , respectivamente.

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

$$g(\mathbf{s}_i, \mathbf{s}_j) = \frac{\lambda_2(\mathbf{s}_i, \mathbf{s}_j)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \quad (2.17)$$

e a densidade de covariância dada por

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\mathbf{s}_i, \mathbf{s}_j) - \lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j). \quad (2.18)$$

No estudo das propriedades de segunda ordem, as suposições de estacionariedade e isotropia sobre um processo pontual são feitas a fim de facilitar as análises. De acordo com Lloyd (2010), o termo estacionariedade é muitas vezes tomado para se referir ao resultado de algum processo que tem propriedades semelhantes em todos os locais na região de interesse, mais especificamente, o processo é dito estacionário se as suas propriedades estatísticas são invariantes sobre translação, ou seja, a intensidade de primeira ordem é constante em toda parte ( $\lambda(\mathbf{s}) = \lambda$ ) e a relação entre dois eventos, medida pela intensidade de segunda ordem, depende somente da sua distância relativa ( $\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\mathbf{s}_i - \mathbf{s}_j)$ ). Schabenberger e Gotway (2005) apontam que, se além de estacionário, o processo for isotrópico, a função intensidade de segunda ordem depende então apenas das distâncias entre os eventos e não da direção entre eles, ou seja,  $\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lambda_2(\|\mathbf{s}_i - \mathbf{s}_j\|) = \lambda_2(r)$ .

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

$$g(r) = \frac{\lambda_2(\mathbf{s}_i, \mathbf{s}_j)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \quad (2.19)$$

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

Esses conceitos possibilitam a estruturação da função  $K$  espacial proposta por Ripley (1976) que é uma das medidas propostas para caracterizar as propriedades de segunda ordem de um PPE. Ela pode ser definida por

$$K(r) = \lambda^{-1}\mathbb{E}[N(r)], \quad (2.20)$$

em que  $N(r)$  é o número de eventos encontrados dentro de uma distância  $r$  e  $\lambda$  é a intensidade de primeira ordem do processo pontual.

Pode-se definir a função  $K$  também a partir da função de correlação entre pares de eventos. Ela é dada conforme Diggle (2013) por

$$K(r) = 2\pi \int_0^r g(\mathbf{s}_i, \mathbf{s}_j) x dx. \quad (2.21)$$

Considerando as suposições de estacionariedade e isotropia, pode-se mostrar que, para o caso de um PPE de Poisson, ou seja, no caso de CAE, tem-se  $K(r) = 2\pi r^2$ .

Pode-se estimar a função  $K$  espacial, conforme Baddeley, Rubak e Turner (2015), por

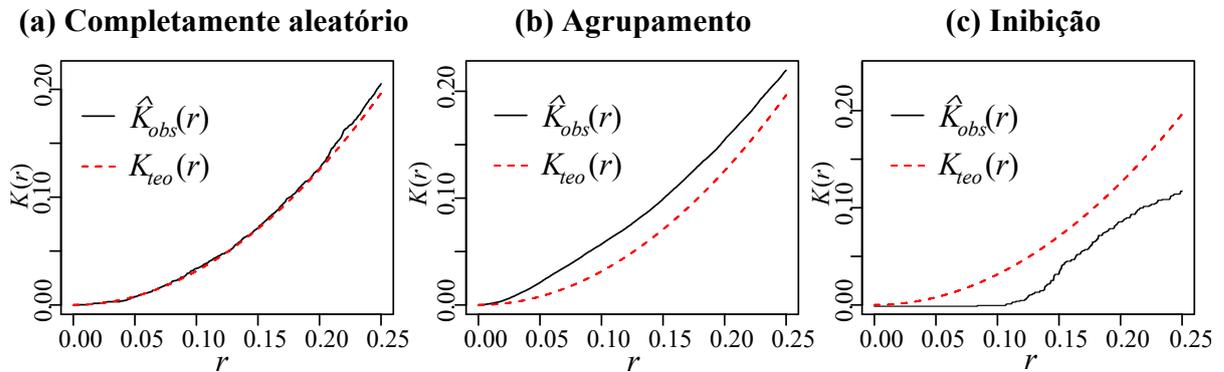
$$\hat{K}(r) = \frac{|B|}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{I}_{(\|\mathbf{s}_i - \mathbf{s}_j\| \leq r)} w_{ij}^{(r)} \quad (2.22)$$

em que  $\mathbb{I}_{(\bullet)}$  é a função indicadora que retorna 1, se o argumento “ $\bullet$ ” é satisfeito e 0, caso contrário. Uma correção do efeito de bordas  $w_{ij}^{(r)}$  que pode ser considerada, é a chamada *isotrópica*, que consiste em considerar a proporção de uma circunferência de raio  $r$ , centrada em  $\mathbf{s}_i$ , que está dentro da região de estudo  $B$ . Para outros tipos de correção do efeito de bordas, ver Baddeley, Rubak e Turner (2015), capítulo 7.

Considerando uma realização de um PPE, se o gráfico da função  $K$  estimada ficar acima da função  $K$  teórica (Figura 2.5 (a)), o que se tem é um padrão que exhibe agrupamentos, pois o número de eventos observado para várias escalas de distância é maior do que o esperado para o processo de Poisson homogêneo. Já para o caso de regularidade, o número de eventos observado é menor do que o esperado pelo processo de Poisson homogêneo, para várias escalas de distância, de modo que o gráfico da função  $K$  estimada fica abaixo da função  $K$  teórica (Figura 2.5 (c)).

Assim como para uma amostra aleatória é necessário realizar algum teste de hipótese, para verificar se os dados seguem uma determinada distribuição ou não, para o caso de um PPE recomenda-se utilizar uma abordagem equivalente. De acordo com Wiegand e Moloney (2013), abordagens utilizando estatísticas sumárias como a função  $K$  requerem a comparação de padrões observados e padrões criados por um modelo nulo para avaliar sua correspondência. Os autores apontam que a alternativa mais prática é usar simulações de Monte Carlo, para produzir realizações do processo pontual estocástico subjacente ao modelo nulo específico (que no caso da função  $K$  pode ser o PPH), com o objetivo de indicar a faixa de variação estatística sob o modelo nulo.

Figura 2.5 – Função  $K$  para diferentes realizações de PPE (linha contínua) e o esperado teórico para um PPH (linha tracejada).



Fonte: Do autor (2020)

Nesse procedimento, cada simulação do modelo nulo gera um padrão de pontos e a estatística de teste é calculada para cada padrão simulado. Wiegand e Moloney (2013) apontam que, como o modelo nulo é estocástico, os resultados das estatísticas de teste em cada um dos diferentes padrões simulados, não apresentam uma correspondência exata entre si, mas, em vez disso, exibem flutuações em torno do valor teoricamente esperado. Assim, o objetivo da abordagem utilizando o envelope de simulação é determinar se a estatística de teste dos dados observados está fora da faixa de flutuações estocásticas produzidas pelo modelo nulo.

Para esse procedimento, consideremos conforme Baddeley, Rubak e Turner (2015) uma função sumária  $S$  qualquer para processos pontuais. Para realizar o teste, primeiro calcula-se a função  $S$  para os dados observados,  $S_{obs}(r)$ . Em seguida, procede-se com a geração de  $m$  processos pontuais sintéticos, simulados a partir da hipótese nula e se obtém as suas respectivas funções  $S$  estimadas para cada simulação,  $S^{(1)}(r), S^{(2)}(r), \dots, S^{(m)}(r)$ .

Para aplicar o princípio dos testes Monte Carlo, Baddeley, Rubak e Turner (2015) apontam que é necessário reduzir a função  $S(r)$  a um simples número  $T$ , que servirá como estatística de teste. Uma forma apresentada pelos autores consiste em especificar uma distância  $r = r_0$  e definir  $T = S(r_0)$ , que é o valor da função sumária nessa distância específica. De acordo com os autores, sob a hipótese nula, os valores de  $\hat{S}(r_0)$  para os dados e para os padrões simulados são variáveis aleatórias estatisticamente equivalentes, e o princípio do teste Monte Carlo se aplica. Assim, o teste de Monte Carlo bilateral, com nível de significância  $\alpha = 2/(m + 1)$  rejeita a hipótese nula se o valor  $t_{obs} = S_{obs}(r_0)$  fica fora do alcance de todas os valores obtidos por simulação  $t_1 = S^{(1)}(r_0), t_2 = S^{(2)}(r_0), \dots, t_m = S^{(m)}(r_0)$ , isto é, se  $t_{obs} > \max_j t_j$  ou  $t_{obs} < \min_j t_j$ . Esta regra é, de acordo com Baddeley, Rubak e Turner (2015), equivalente a construir o gráfico

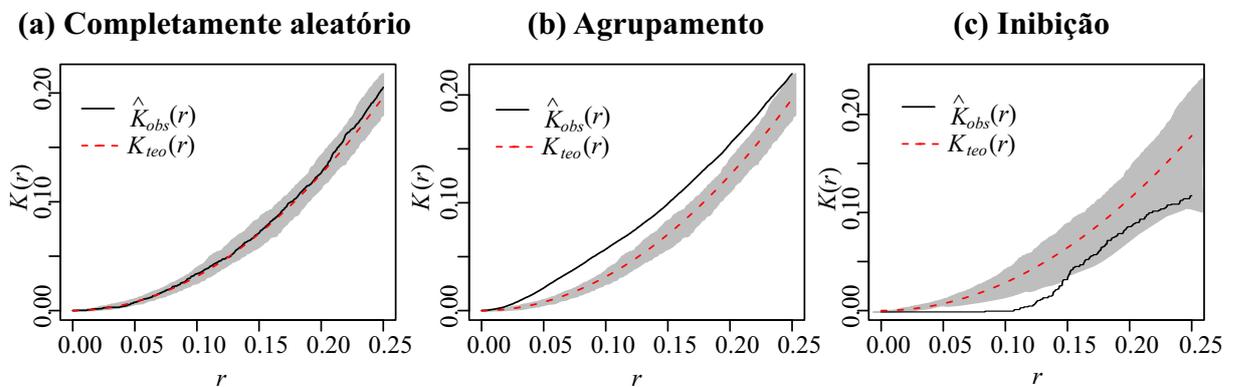
dos envelopes superior e inferior

$$\begin{aligned} U(r) &= \max_j S^{(j)}(r) \left\{ S^{(1)}(r), S^{(2)}(r), \dots, S^{(m)}(r) \right\} \\ L(r) &= \min_j S^{(j)}(r) \left\{ S^{(1)}(r), S^{(2)}(r), \dots, S^{(m)}(r) \right\}, \end{aligned} \quad (2.23)$$

cortando-o uma linha vertical que intercepta o eixo horizontal em  $r = r_0$ , e a hipótese nula é rejeitada se o valor da função observada  $S_{obs}(r)$  estiver fora do intervalo limitado por esses envelopes nessa distância pré-especificada  $r = r_0$ .

Na Figura 2.6 estão apresentados os mesmos gráficos da Figura 2.5, mas agora com os envelopes de simulação construídos a partir de 199 simulações do modelo nulo de completa aleatoriedade e considerando um nível de significância  $\alpha = 0,05$ , para cada um dos casos. Note que para o caso de inibição (2.6 (c)), a função  $K$  estimada fica fora do envelope até uma distância  $r = 0,15$  e depois permanece no seu interior. Isso indica que a configuração pontual exibe um padrão de inibição até  $r = 0,15$  e passa a ser aleatório depois disso.

Figura 2.6 – Função  $K$  para diferentes realizações de PPE com os respectivos envelopes construídos a partir de 199 simulações de um PPH.



Fonte: Do autor (2020)

De acordo com Wiegand e Moloney (2013), o uso de 199 simulações e não de 200, deve-se ao fato de que o cálculo da taxa de erro é baseado nas 199 simulações do modelo nulo mais o padrão observado. Considere por exemplo, que um padrão observado siga de fato um PPH. Teremos nesse caso, que o risco de  $\hat{K}(r)$  estimada para esse padrão observado esteja fora dos limites do envelope é  $5/200$  para estar abaixo e  $5/200$  para estar acima. De modo mais geral, pode-se definir os envelopes de simulação como os  $s$ -ésimos valores mais altos e mais baixos da estatística de teste, tomados a partir de  $m$  simulações do modelo nulo com nível de significância

$\alpha = 2 \frac{s}{m+1}$  para o teste bicaudal. No caso da figura 2.6, os envelopes foram construídos considerando os quintos maiores e menores valores observados para os dados simulados.

Outras ferramentas de análise para um PPE podem ser consultadas em Baddeley, Rubak e Turner (2015), Cressie (1993), Diggle (2013), Schabenberger e Gotway (2005), entre outros.

### 2.3 Processos Pontuais Temporais

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

De acordo com Cox e Isham (2018), um processo pontual temporal pode ser estudado, assim como no espacial, pelas propriedades de primeira e segunda ordem. Os autores mostram que a função intensidade que descreve o número médio de eventos por unidade de tempo pode ser expressa por

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{E}[N(dt)]}{|dt|}, \quad (2.24)$$

em que  $N(dt)$  é o número de eventos dentro de um pequeno intervalo em torno do tempo  $t$ , com comprimento igual a  $|dt|$ . Se essa intensidade for constante em todo o intervalo, ou seja,  $\lambda(t) = \lambda$ , o processo pontual temporal é dito homogêneo.

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

$$\lambda(t_i, t_j) = \lim_{|t_i|, |t_j| \rightarrow 0} \frac{\mathbb{E}[N(t_i)N(t_j)]}{\lambda(dt_i)\lambda(dt_j)} \quad (2.25)$$

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

$$K(v) = \lambda^{-1} E[N(v)], \quad (2.26)$$

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

O valor teórico da função sob a hipótese de estacionariedade e independência é

$$K(v) = 2v, \quad (2.27)$$

e pode ser estimada, conforme Diggle et al. (1995), por

$$\hat{K}(v) = \frac{T}{n} \sum_{i=1}^n \sum_{i \neq j} \frac{1}{w_{ij}^{(t)}} \frac{\mathbb{I}_{(|t_i - t_j| \leq v)}}{n-1}, \quad (2.28)$$

em que  $n$  é o número de eventos observados no intervalo de tempo  $T$ ,  $\mathbb{I}_{(\cdot)}$  é a função indicadora e  $w_{ij}^{(t)}$  é um fator de correção do efeito bordas, tal que  $w_{ij}^{(t)} = 1$  se ambas as extremidades do intervalo, de comprimento  $2v$ , com centro em  $t_i$  estiver dentro do intervalo  $T$  e  $w_{ij}^{(t)} = \frac{1}{2}$ , caso contrário. A interpretação do gráfico da função  $K$  temporal é da mesma forma que o caso espacial, com a diferença que enquanto no caso temporal, a função  $K$  teórica é uma reta, no caso espacial, o gráfico é de uma parábola.

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

Baddeley, Bárány e Schneider (2007) mostram que, alternativamente, pode-se estudar os tempos de inter-chegada  $S_i = T_{i+1} - T_i$ . Estes têm a vantagem de que, para alguns modelos especiais como o Poisson, as variáveis aleatórias  $S_1, S_2, \dots$  são independentes. Outro modo apresentado pelo autor consiste em formular um processo pontual em termos de um processo de contagem cumulativa

$$N_t = \sum_{i=1}^{\infty} \mathbb{I}_{(T_i \leq t)}, \quad (2.29)$$

ou seja,  $N_t$  é o número de ocorrências observadas até o tempo  $t$ . Esta abordagem tem a vantagem de converter o processo numa função aleatória de tempo contínuo  $t$ , mas tem a desvantagem de que os valores  $N_t$  para diferentes  $t$  são altamente dependentes (BADDELEY; BÁRÁNY; SCHNEIDER, 2007).

Há ainda outras formas de se analisar um processo pontual no tempo e um aprofundamento nessa teoria pode ser feito em Cox e Isham (2018), Baddeley, Bárány e Schneider (2007) e Daley e Vere-Jones (2003).

## 2.4 Processos Pontuais Espaço-Temporais

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

Formalmente, um processo pontual espaço-temporal (PPET) é um processo estocástico que gera um conjunto contável de pontos  $\{(s_i, t_i) : i = 1, \dots, n\}$ , em que  $s_i \in S \subset \mathbb{R}^2$  e  $t_i \in T \subset \mathbb{R}^+$ . Deste modo, o estudo dos processos pontuais espaço-temporais é considerado no produto cartesiano  $\mathbb{R}^2 \times \mathbb{R}^+$  (DIGGLE, 2013).

### 2.4.1 Propriedades de primeira ordem

Um PPET, assim como no PPE e PPT, também pode ser analisado em função das propriedades de primeira e segunda ordem e, conforme Gabriel et al. (2013), as propriedades de primeira ordem de um processo pontual espaço-temporal podem ser descritas pela sua função de intensidade espaço-temporal, dada por

$$\lambda(\mathbf{s}, t) = \lim_{|ds|, |dt| \rightarrow 0} \left\{ \frac{E[N(ds, dt)]}{|ds||dt|} \right\}, \quad (2.30)$$

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

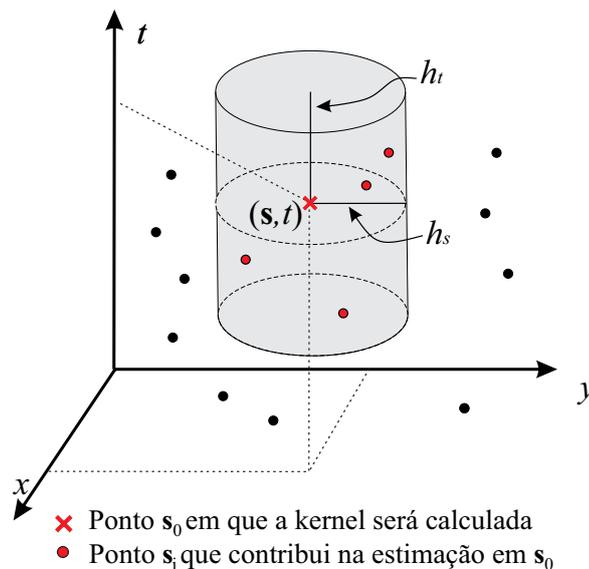
Assim como no caso espacial, é possível utilizar a função kernel para verificar a intensidade dos pontos no domínio espaço-temporal. Uma extensão para a função kernel espaço-temporal, apresentada por Brunson, Corcoran e Higgs (2007), é dada por

$$\hat{f}(x, y, t) = \frac{1}{nh_s^2 h_t} \sum_{i=1}^n \kappa_s \left( \frac{x-x_i}{h_s}, \frac{y-y_i}{h_s} \right) \kappa_t \left( \frac{t-t_j}{h_t} \right), \quad (2.31)$$

em que  $x$  e  $y$  são as coordenadas do ponto  $\mathbf{s}$ ,  $\kappa_s$  é a função kernel no espaço com largura de banda  $h_s$  e  $\kappa_t$  é a função kernel no tempo com largura de banda  $h_t$ .

No caso espacial, a largura de banda forma um círculo (ou elipsóide, se forem consideradas larguras de banda diferentes nas direções dos eixos  $x$  e  $y$ ) em torno do ponto arbitrário  $\mathbf{s}_0$  avaliado (ver Figura 2.3) e apenas os pontos observados dentro desse círculo contribuem na estimação da intensidade. Para o caso espaço-tempo, deve-se levar em consideração que tempo e espaço não possuem as mesmas métricas, e a análise é dada em  $\mathbb{R}^2 \times \mathbb{R}^+$  e não no  $\mathbb{R}^3$ , assim, as larguras de banda  $h_s$  e  $h_t$  formam um cilindro em torno do ponto arbitrário  $(\mathbf{s}_0, t_0)$  e apenas os pontos dentro desse cilindro contribuem no cálculo da kernel, como mostrado na Figura 2.7. Se estivéssemos considerando apenas o caso tridimensional no espaço ( $\mathbb{R}^3$ ), teríamos uma esfera (ou de modo mais geral, um elipsóide) em torno do ponto em que a kernel é calculada.

Figura 2.7 – Pontos do PPET que contribuem para o cálculo da kernel espaço-temporal



Fonte: Do autor (2020)

### 2.4.2 Propriedades de segunda ordem

No que se refere às propriedades de segunda ordem de um processo pontual espaço-temporal, Gabriel et al. (2013) apontam que elas descrevem a relação entre números de eventos em pares de sub-regiões dentro de  $S \times T$  e são dadas em termos da função de intensidade

$$\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lim_{|ds_i \times dt_i|, |ds_j \times dt_j| \rightarrow 0} \left\{ \frac{E[N(ds_i \times dt_i)N(ds_j \times dt_j)]}{|ds_i \times dt_i||ds_j \times dt_j|} \right\}, \quad (2.32)$$

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

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

$$\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) - \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) \quad (2.33)$$

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

$$g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \frac{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}. \quad (2.34)$$

Ainda de acordo com Gabriel et al. (2013), a densidade de covariância pode ser interpretada de modo análogo a covariância de variáveis aleatórias e a função de correlação entre pares pode ser interpretada, informalmente, como a densidade de probabilidade padronizada de ocorrer um evento em cada um de dois pequenos volumes centrados nos pontos  $(\mathbf{s}_i, t_i)$  e  $(\mathbf{s}_j, t_j)$ .

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

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

De acordo com esses autores, para o caso do processo de Poisson homogêneo, as intensidades de primeira e de segunda ordem reduzem-se a constantes  $\lambda(\mathbf{s}, t) = \lambda$  e  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda^2$ . Deste modo, a partir de 2.33 e 2.34 temos, respectivamente, que densidade de covariância é dada por

$$\begin{aligned} \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \underbrace{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}_{=\lambda^2} - \underbrace{\lambda(\mathbf{s}_i, t_i)}_{=\lambda} \underbrace{\lambda(\mathbf{s}_j, t_j)}_{=\lambda} \\ &= \lambda^2 - \lambda^2 \\ &= 0 \end{aligned} \quad (2.35)$$

é identicamente zero e a função de correlação entre pares

$$\begin{aligned} g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \frac{\overbrace{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}^{=\lambda^2}}{\underbrace{\lambda(\mathbf{s}_i, t_i)}_{=\lambda} \underbrace{\lambda(\mathbf{s}_j, t_j)}_{=\lambda}} \\ &= \frac{\lambda^2}{\lambda^2} \\ &= 1 \end{aligned} \quad (2.36)$$

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

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

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

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

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

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

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

Gabriel et al. (2013) afirmam que para um processo de Poisson com intensidade  $\lambda(\mathbf{s}, t)$ , a intensidade de segunda ordem é  $\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)$ . Assim, de modo análogo às equações (2.35) e (2.36), tem-se que a densidade de covariância

$$\begin{aligned} \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) - \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) \\ &= \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) - \lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j) \\ &= 0 \end{aligned} \quad (2.37)$$

é identicamente zero e a função de correlação entre pares

$$\begin{aligned} g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) &= \frac{\lambda_2((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \\ &= \frac{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \\ &= 1 \end{aligned} \quad (2.38)$$

é identicamente 1.

Assim como no caso espacial, para o caso em que a intensidade de primeira ordem não é constante, a tendência espaço-temporal nas localizações dos eventos deve ser levada em consideração. O conceito de estacionariedade de segunda ordem reponderada, tratado para o caso espacial em Baddeley, Møller e Waagepetersen (2000), foi estendido diretamente para o caso

espaço-temporal em Gabriel e Diggle (2009). Esta suposição mais fraca de estacionariedade permite uma intensidade não constante que é obrigada a ser limitada a partir de zero e assume que a função de correlação entre pares, dada na equação 2.34, depende apenas do vetor de diferenças espaço-temporais  $(r, t)$ , ou seja,

$$g(r, t) = \frac{\lambda_2(r, t)}{\lambda(\mathbf{s}_i, t_i)\lambda(\mathbf{s}_j, t_j)} \quad (2.39)$$

Essas suposições permitem a estruturação da função  $K$  para investigar a dependência espaço-temporal do processo. Diggle et al. (1995) estenderam a ideia da função  $K$ , em (2.20) e (2.26) para o caso espaço temporal, definindo-a como

$$K(r, t) = \lambda^{-1} \mathbb{E}[N(r, t)], \quad (2.40)$$

em que  $N(r, t)$  é o número de eventos encontrados dentro de uma distância  $r$  e tempo  $t$  e  $\lambda$  é a intensidade homogênea do processo pontual. A função  $K$  pode ser definida diretamente da esperança, como em (2.40), ou de acordo com Diggle e Gabriel (2010), a partir da função de correlação entre pares de eventos, por

$$K(r, t) = 2\pi \int_{-t}^t \int_0^r g((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) ds dt. \quad (2.41)$$

Sob a hipótese da completa aleatoriedade espaço-temporal, ou seja, quando se tem um processo de Poisson homogêneo, pode-se mostrar que o valor da função  $K$  teórica é  $K(r, t) = 2\pi r^2 t$ .

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

Considerando agora o caso em que o processo não é estacionário, ou seja, a sua intensidade de primeira ordem não é constante, Gabriel e Diggle (2009) se utilizaram da suposição de estacionariedade de segunda ordem reponderada para definir a função  $K$  não homogênea como

$$K_I(r, t) = 2\pi \int_{-t}^t \int_0^r g(r, t) s ds dt, \quad (2.42)$$

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

Sob a hipótese de o processo ser de Poisson não homogêneo, pode-se mostrar que a função de correlação entre pares é igual a 1, e assim a função  $K_I(r, t)$  na equação (2.42) se reduz a

$$\begin{aligned} K_I(r, t) &= 2\pi \int_{-t}^t \int_0^r 1 s ds dt \\ &= \pi r^2 \int_{-t}^t dt \\ &= 2\pi r^2 t. \end{aligned} \quad (2.43)$$

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

$$\hat{K}_I(r, t) = \frac{1}{|S \times T|} \frac{n}{n_t} \sum_{i=1}^{n_t} \sum_{j>i}^{n_t} \frac{1}{w_{ij}} \frac{1}{\lambda(s_i, t_i) \lambda(s_j, t_j)} I_{(\|\mathbf{s}_i - \mathbf{s}_j\| \leq r; t_j - t_i \leq t)}. \quad (2.44)$$

em que  $n_t$  é o número de eventos para os quais  $t_i \leq T_1 - t$ ,  $w_{ij} = w_{ij}^{(s)} w_{ij}^{(t)}$  é um fator de correção do efeito de bordas, sendo  $w_{ij}^{(s)}$  e  $w_{ij}^{(t)}$  as correções de borda espacial e temporal, respectivamente, e  $I_{(\cdot)}$  é função indicadora.

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

$$\lambda(\mathbf{s}, t) = m(s) \mu(t) \quad \forall (\mathbf{s}, t) \in S \times T, \quad (2.45)$$

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

Sob essa suposição de separabilidade, Gabriel e Diggle (2009) afirmam que quaisquer efeitos não separáveis são interpretados como efeitos de segunda ordem, ao invés de primeira. Ainda de acordo com os autores, estimativas adequadas de  $m(\mathbf{s})$  e  $\mu(t)$  em (2.45) dependerão das características de cada aplicação. De um modo geral,  $\hat{m}(\mathbf{s})$  e  $\hat{\mu}(t)$  podem ser obtidas a partir de funções kernel. O procedimento de análise a partir da função  $K$  não homogênea é análogo à função  $K$  homogênea e a interpretação dos resultados é a mesma.

## 2.5 Processos Pontuais Marcados

Um processo pontual é dito marcado quando, além da localização espacial dos eventos, há uma informação adicional (marca) anexada a cada evento. De acordo com Chiu et al. (2013), um processo pontual marcado em  $\mathbb{R}^d$  é uma sequência aleatória  $X_m = \{(\mathbf{s}_i, m_i)\}$ , em que os pontos  $\mathbf{s}_i$  juntos constituem o processo pontual não marcado em  $\mathbb{R}^d$ , e os valores  $m_i$  são as marcas correspondentes a cada ponto  $\mathbf{s}_i$ , isto é,  $m_i = m(\mathbf{s}_i)$ . Considerando aqui  $d = 2$  e que as marcas pertencem a espaço das marcas  $\mathcal{M}$ , o estudo de processos pontuais marcados é dado em  $\mathbb{R}^2 \times \mathcal{M}$ . Daley e Vere-Jones (2002) referem ao processo pontual não marcado  $X$  como *processo de base*, o que também será considerado aqui.

A respeito da natureza das marcas, estas podem ser variáveis contínuas ou discretas/categóricas. Um exemplo de variável contínua é o diâmetro à altura do peito de árvores em um talhão de floresta e, no caso discreto, a espécie das árvores.

É importante destacar que dados de padrão de pontos marcados tem semelhanças com dados geoestatísticos, que também consistem em informações sobre localizações  $\mathbf{s}$  e valores  $Z(\mathbf{s})$  associados. No entanto, conforme nos apontam Illian et al. (2008), o objetivo da geoestatística é estimar fenômenos espacialmente contínuos com base em medições feitas em pontos escolhidos para esse fim. No caso de padrões pontuais marcados, os pontos não representam os locais que foram escolhidos para fins de medição, mas as localizações exatas em que os objetos analisados ocorreram. Assim, de acordo com os autores, a análise de dados geoestatísticos e dados de processos pontuais marcados possuem objetivos completamente diferentes e aplicam métodos diferentes, de modo que o mau uso de métodos geoestatísticos em processos pontuais marcados pode produzir resultados incorretos. Como só existem marcas nas localizações dos eventos e não em toda a região de estudo, uma variável que ocorra em toda a região, que se pretende incorporar à análise de padrões de pontos deve ser tratada então, como uma covariável e não como uma marca.

Considerando que as marcas podem ser de duas naturezas, há métodos específicos para se analisar dados com marcas contínuas ou categóricas. No entanto, pode-se também tomar os dados de marcas contínuas e estabelecer limiares de modo a criar categorias, de tal modo que dados com marcas contínuas podem ser analisados como se fossem discretos.

Neste trabalho não será dado enfoque para métodos de análise de pontos com marcas discretas, uma vez que o objetivo é trabalhar com marcas contínuas (temos nesse caso, que  $\mathcal{M} = \mathbb{R}$  e o processo  $X_m$  é considerado em  $\mathbb{R}^2 \times \mathbb{R}$ ). Métodos de análise para marcas categóricas podem ser encontrados em Baddeley, Rubak e Turner (2015) e Illian et al. (2008).

Assim como nos processos mostrados anteriormente, a suposição de estacionariedade também é utilizada no caso marcado. A definição é similar ao caso puramente espacial, como mostram Illian et al. (2008):

Seja o processo pontual marcado  $X_m = \{(\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2), \dots\}$  e  $X_m^s = \{(\mathbf{s}_1 + \mathbf{s}, m_1), (\mathbf{s}_2 + \mathbf{s}, m_2), \dots\}$  é o processo transladado por  $\mathbf{s}$ . Note que em  $X_m^s$  as marcas continuam as mesmas, mas os pontos são transladados. A partir disso temos que um processo pontual marcado  $X_m$  é estacionário se e somente se

$$X_m \stackrel{d}{=} X_m^s,$$

ou seja, o processo pontual marcado  $X_m$  e o processo pontual marcado transladado por  $\mathbf{s}$  tem a mesma distribuição. Se o processo além de possuir essa invariância quanto à translação, também for invariante quanto à rotação, ele será dito isotrópico.

Como nos casos apresentados anteriormente, um processo pontual marcado pode ser estudado pelas propriedades de primeira e segunda ordem, dadas a partir de funções de intensidade. De um modo geral, Zhang e Zhuang (2014), expressam a função de intensidade de  $k$ -ésima ordem por

$$\lambda_k \left( (\mathbf{s}_1, m_1), \dots, (\mathbf{s}_k, m_k) \right) = \lim_{|d\mathbf{s}_i \times dm_i| \rightarrow 0, i=1, \dots, k} \left\{ \frac{E[N(d\mathbf{s}_1 \times dm_1) \dots N(d\mathbf{s}_k \times dm_k)]}{|d\mathbf{s}_1 \times dm_1| \dots |d\mathbf{s}_k \times dm_k|} \right\} \quad (2.46)$$

em que  $(\mathbf{s}_i, m_i)$  são distintos pares de pontos e marcas em  $\mathbb{R}^2 \times \mathbb{R}$ ,  $d\mathbf{s}_i \times dm_i$  é uma região infinitesimal contendo  $(\mathbf{s}_i, m_i)$  e  $|d\mathbf{s}_i \times dm_i|$  é a medida de  $d\mathbf{s}_i \times dm_i$  (a intensidade de primeira ordem  $\lambda_1$  será denotada simplesmente por  $\lambda$ ).

### 2.5.1 Propriedades de primeira ordem

Consideremos que o processo pontual marcado  $X_m$  ocorre no produto cartesiano  $A \times B$ , em que  $A \subset \mathbb{R}^2$  e  $B \subset \mathcal{M}$ . A medida média de  $X_m$ , ou seja, o número esperado de eventos na região  $A \times B$  é denotada, conforme Zhang e Zhuang (2014), por

$$\mathbb{E}[N(A \times B)] = \mu(A \times B) = \int_A \int_B \lambda(\mathbf{s}, m) dm d\mathbf{s}. \quad (2.47)$$

A intensidade  $\lambda(\mathbf{s}, m)$  é, de acordo com Baddeley, Rubak e Turner (2015), para o processo pontual em duas dimensões com marcas de valor real, uma função de três coordenadas (duas coordenadas espaciais e a variável marca). Integrando a função de intensidade em relação à variável marca e leva à intensidade do processo pontual das localizações:

$$\lambda_{\bullet}(\mathbf{s}) = \int_{-\infty}^{\infty} \lambda(\mathbf{s}, m) dm. \quad (2.48)$$

Dado um ponto  $\mathbf{s}$ , a razão

$$p(m|\mathbf{s}) = \frac{\lambda(\mathbf{s}, m)}{\lambda_{\bullet}(\mathbf{s})} \quad (2.49)$$

expressa, conforme Baddeley, Rubak e Turner (2015), a densidade de probabilidade condicional da marca  $m$  sobre  $\mathbf{s}$ , e o valor esperado da marca sobre essa localização é

$$e(\mathbf{s}) = \int_{-\infty}^{\infty} m p(m|\mathbf{s}) dm = \frac{\int_{-\infty}^{\infty} m \lambda(\mathbf{s}, m) dm}{\int_{-\infty}^{\infty} \lambda(\mathbf{s}, m) dm}. \quad (2.50)$$

Para a estimação dessas funções são utilizados estimadores baseados em suavizadores de kernel e, no que se refere à função  $\lambda(\mathbf{s}, m)$ , de acordo com Baddeley, Rubak e Turner (2015), utiliza-se um kernel em três dimensões:

$$\hat{\lambda}(\mathbf{s}, m) = \kappa_3((\mathbf{s}_i, m_i) - (\mathbf{s}, m)) \quad (2.51)$$

em que  $\kappa_3$  é uma densidade de probabilidade em  $\mathbb{R}^3$ .

A intensidade das localizações  $\lambda_{\bullet}(\mathbf{s})$  pode ser estimada, como apontam os autores, por um suavizador de kernel das localizações que corresponde ao kernel marginal bidimensional

$$\kappa((x, y)) = \int_{-\infty}^{\infty} \kappa_3((x, y, z)) dz. \quad (2.52)$$

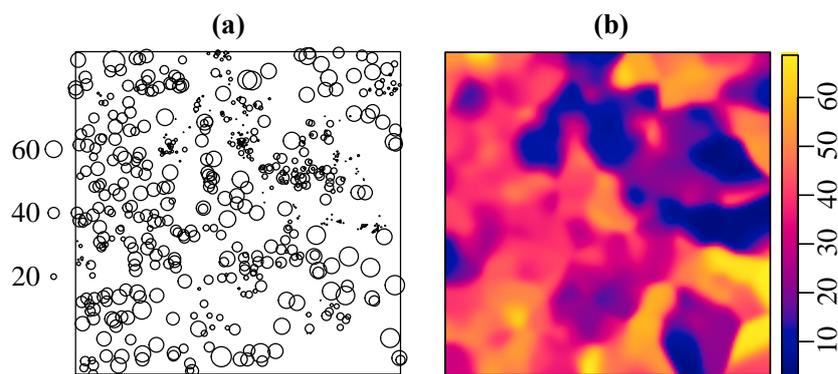
Isso assegura, de acordo com Baddeley, Rubak e Turner (2015), que  $\hat{\lambda}_\bullet$  é a integral marginal de  $\hat{\lambda}$  e a razão  $\hat{p}(m|\mathbf{s}) = \frac{\hat{\lambda}(\mathbf{s}, m)}{\hat{\lambda}_\bullet(\mathbf{s})}$  é uma densidade de probabilidade. A respeito de  $e(\mathbf{s})$ , os autores afirmam que a estimação pode ser feita utilizando o suavizador de Nadaraya-Watson (NADARAYA, 1964; WATSON, 1964; NADARAYA, 1989), dado por

$$\tilde{m}(\mathbf{s}) = \frac{\sum_i m_i \kappa(\mathbf{s} - \mathbf{s}_i) b(\mathbf{s}_i)}{\sum_i \kappa(\mathbf{s} - \mathbf{s}_i) b(\mathbf{s}_i)}, \quad (2.53)$$

em que  $\kappa$  é um suavizador de kernel em duas dimensões, e  $b(\mathbf{s}_i)$  é um fator de correção do efeito de bordas. Vejamos a aplicação do suavizador de Nadaraya-Watson sobre o conjunto de dados *longleaf*, disponível no pacote *spatstat* (BADDELEY; TURNER, 2005) do R (R Core Team, 2020).

Na Figura 2.8 tem-se em **(a)** a representação das 584 localizações e medidas do diâmetro à atura do peito da espécie *Pinus palustris*, nativa do sudeste dos Estados Unidos em uma janela quadrada de 200 metros de cada lado. Em **(b)** tem-se o resultado da suavização das marcas pelo suavizador de Nadaraya-Watson, em que, através do mapa pode-se notar que há uma faixa de árvores mais jovens no quarto superior direito do talhão. Esta é uma ferramenta interessante que permite observar, em média, o valor esperado das marcas para toda a região, condicionadas ao processo pontual observado.

Figura 2.8 – Aplicação do suavizador de Nadaraya-Watson aos dados *longleaf*



Legenda: **(a)** 584 localizações e diâmetros à atura do peito de *Pinus palustris*. Os círculos são proporcionais aos diâmetros das árvores. **(b)** Variação espacial do diâmetro médio das árvores de *Pinus palustris* em centímetros, calculada pelo suavizador de Nadaraya-Watson utilizando o pacote *spatstat*

Fonte: Do autor (2020)

Note que, neste exemplo, não foi abordado a forma como os eventos estão distribuídos, se a intensidade é constante ou não, ou se existe alguma estrutura de correlação espacial en-

tre eles. No caso de processos pontuais marcados com marcas quantitativas, o interesse está, conforme apontam Wiegand e Moloney (2013), na caracterização do processo que distribuiu as marcas sobre os eventos, enquanto as características do próprio padrão não marcado geralmente não são de interesse.

### 2.5.2 Propriedades de segunda ordem

Considerando um processo pontual puramente espacial, as estatísticas de segunda ordem são úteis para que se possa avaliar a variabilidade na distribuição dos pontos, ou seja, a forma como os pontos interagem entre si. No caso de um processo pontual marcado, além do mencionado acima, busca-se descrever a variabilidade na distribuição das marcas, como também as correlações existentes entre as marcas e os pontos. Uma suposição que é feita em muitos casos é assumir que marcas são independentes entre si e identicamente distribuídas (i.i.d.). Entretanto, a marca  $m_i$  pode depender da sua localização  $s_i$ .

Em processos pontuais marcados, uma forma alternativa de aleatoriedade é conhecida como *rotulagem aleatória*. Enquanto a CAE descreve eventos que ocorrem aleatoriamente em uma área de estudo, a rotulagem aleatória descreve simplesmente os pontos de um processo sendo marcados aleatoriamente, independente da distribuição do processo de base, seja agrupamento, regularidade ou completamente aleatório.

No caso da rotulagem aleatória, as marcas do processo são, conforme Pawlas (2009), independentes e identicamente distribuídas (i.i.d.) e além disso, são independentes dos pontos.

Para testar um conjunto de dados sob a hipótese de rotulagem aleatória, o procedimento é análogo ao executado para a função  $K$  utilizando os envelopes de simulação. No entanto para a rotulagem aleatória, as  $m$  “simulações” são obtidas mantendo-se fixas as localizações do processo de base e permutando  $m$  vezes as marcas, de modo que as marcas são atribuídas de modo aleatório para cada localização. Para cada uma das permutações computa-se a estatística de teste de interesse, constrói-se os envelopes e compara-se com o resultado dos dados originais, do mesmo modo como mostrado anteriormente. Rejeitar a hipótese de rotulagem aleatória significa então, que as marcas não são i.i.d e não foram atribuídas de modo aleatório as suas localizações, indicando que existe uma estrutura de correlação entre elas e com os pontos.

Nesta tese será abordado aqui apenas o variograma marcado para as análises de segunda ordem do processo pontual marcado. Outros métodos propostos para esse tipo de análise podem

ser encontradas em Baddeley, Rubak e Turner (2015), Illian et al. (2008), Chiu et al. (2013), Cronie e Lieshout (2016), entre outros.

### 2.5.2.1 O variograma marcado

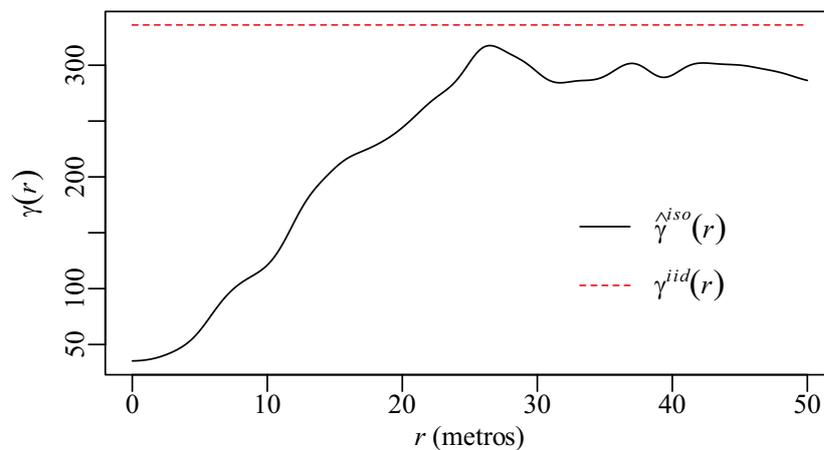
O variograma marcado (CRESSIE, 1993; WÄLDER; STOYAN, 1996; STOYAN; WÄLDER, 2000; ILLIAN et al., 2008; BADDELEY; RUBAK; TURNER, 2015) para um processo pontual marcado estacionário é definido por

$$\gamma(r) = \frac{1}{2} \mathbb{E} \left[ (m_i - m_j)^2 \mid \|\mathbf{s}_i - \mathbf{s}_j\| = r \right]. \quad (2.54)$$

Temos então que  $2\gamma(r)$  é a diferença quadrática esperada entre os valores de duas marcas cujos pontos correspondentes estão separados por uma distância  $r$ . De acordo com Stoyan et al. (2017), o comportamento de  $\gamma(r)$  como uma função de  $r$  fornece uma valiosa informação sobre as correlações espaciais entre as marcas do processo. Três aspectos dessa informação que são destacados pelos autores estão descritos a seguir.

1. Se as marcas de pontos próximos forem semelhantes para pequenos valores de  $r$ ,  $\gamma(r)$  tem valores pequenos. Este é um comportamento natural em muitos casos. Vejamos o exemplo dos dados *longleaf*, cujo variograma marcado está apresentado na Figura 2.9.

Figura 2.9 – Variograma marcado para o conjunto de dados *longleaf* obtido com o pacote *spatstat*.

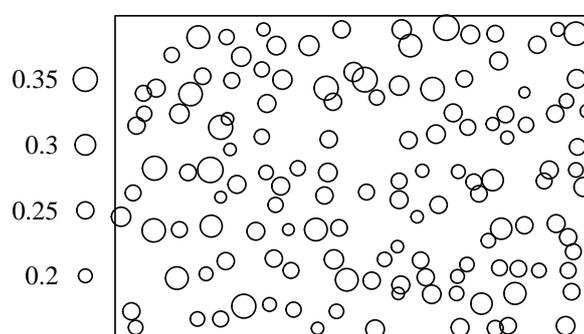


Legenda: Linha contínua: variograma estimado considerando uma correção de bordas isotrópica. Linha tracejada: variograma esperado para o caso de marcas i.i.d.

Fonte: Do autor (2020)

Stoyan et al. (2017) destacam, no entanto, que esta não é a única possibilidade e relatam o exemplo apresentado em Baddeley, Rubak e Turner (2015) a respeito do conjunto de dados *spruces* (pinheiro típico de regiões temperadas) disponível no pacote *spatstat*. Na Figura 2.10 está a representação desse conjunto de dados, com a localização de 134 pinheiros em um estande de floresta natural na Saxônia, Alemanha, com cada marca sendo o diâmetro à altura do peito.

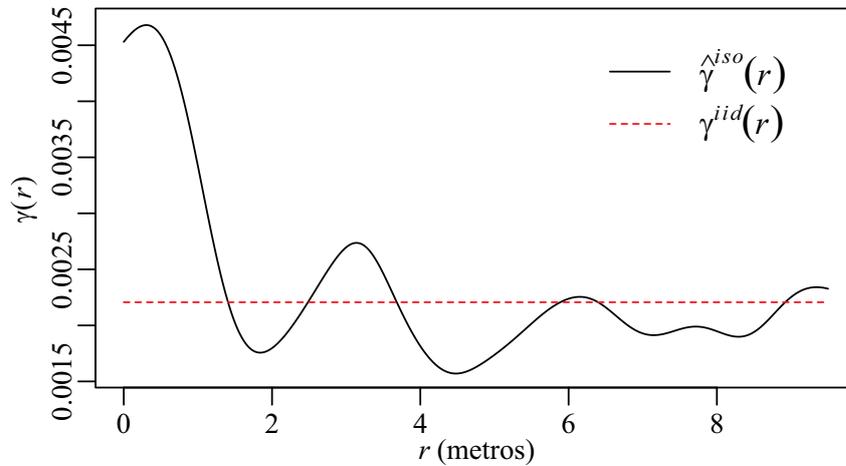
Figura 2.10 – Pinheiros em um estande de floresta na região da Saxônia, Alemanha, marcados com o diâmetro à altura do peito.



Fonte: Do autor (2020)

O variograma marcado para esse conjunto de dados está apresentado na Figura 2.11. Nota-se, conforme apontam Baddeley, Rubak e Turner (2015), que há um aumento da variação à distâncias muito curtas. Uma interpretação possível, de acordo com os autores é que, para árvores bastante grandes encontradas próximas umas das outras, deve haver pelo menos uma que é pequena.

Figura 2.11 – Variograma marcado para o conjunto de dados *spruces* obtido com o pacote *spatstat*.



Legenda: Linha contínua: variograma estimado considerando uma correção de bordas isotrópica. Linha tracejada: variograma esperado para o caso de marcas i.i.d.

Fonte: Do autor (2020)

2. Com o aumento de  $r$ ,  $\gamma(r)$  frequentemente aumenta e mostra uma convergência assintótica em direção a um valor constante, ou seja (ILLIAN et al., 2008):

$$\lim_{r \rightarrow \infty} \gamma(r) = \sigma_{\mu}^2. \quad (2.55)$$

Esse valor constante é chamado *patamar* ou *variância das marcas*. Pode acontecer, entretanto, que  $\gamma(r)$  esteja diminuindo em alguns intervalos de  $r$ , como por exemplo no variograma da Figura 2.11. De acordo com Stoyan et al. (2017), a velocidade da convergência está relacionada à velocidade de declínio da correlação espacial das marcas com o aumento das distâncias inter-pontos. E o valor de  $r$  em que o patamar é atingido é chamado de *alcance*. Se existente, o alcance é um valioso parâmetro de correlação.

3. O comportamento  $\gamma(r)$  próximo à origem, denotado por  $\gamma(0+)$ , também é frequentemente de interesse.  $\gamma(0+)$  é chamado de *efeito pepita* e pode acontecer, de acordo com os autores, que seja positivo e não desapareça, como talvez seja esperado.  $\gamma(0+)$  reflete a agregação espacial de pontos próximos com marcas diferentes. Ou seja, um valor positivo de  $\gamma(0+)$  indica que os pontos próximos tendem a ter marcas significativamente diferentes. E no caso de  $\gamma(0+)$  ser positivo e constante para  $r > 0$ , se refere então a um *efeito pepita puro*.

A definição apresentada na equação (2.54) para o variograma marcado, assim como os termos “patamar”, “alcance” e “efeito pepita”, são análogos aqueles utilizados na Geoestatística clássica.

Conforme Baddeley, Rubak e Turner (2015), seja  $Z(\mathbf{u})$  uma função aleatória estacionária na localização  $\mathbf{u}$ , então o variograma de  $Z$  é

$$\gamma_Z(r) = \frac{1}{2} \mathbb{E} \left[ (Z(\mathbf{u}) - Z(\mathbf{v}))^2 \right], \quad (2.56)$$

em que  $\mathbf{u}$  e  $\mathbf{v}$  são localizações aleatórias, tais que  $\|\mathbf{u} - \mathbf{v}\| = r$ .

Apesar da similaridade das expressões (2.54) e (2.56), Baddeley, Rubak e Turner (2015) e Illian et al. (2008) salientam que elas diferem fortemente em sua definição. O variograma utilizado na Geoestatística caracteriza variáveis regionalizadas, ou seja, variáveis que podem assumir um valor em qualquer ponto da região de estudo. Exemplos de variáveis desse tipo incluem o teor de determinado nutriente no solo, os valores de precipitação de chuva em um estado do Brasil, entre outros. No processo pontual marcado, como mencionado anteriormente, as marcas ocorrem apenas nas localizações dos eventos e não estão definidas para as demais localizações na região de estudo. Assim, para o cálculo do variograma geoestatístico podem ser “escolhidos” diversos locais dentro da área de estudo sendo, portanto, possível obter diferentes variogramas dependendo da amostra que se faz. No caso do processo pontual marcado somente se pode calcular um variograma, já que as marcas só ocorrem nas localizações dos eventos.

Outra questão destacada Baddeley, Rubak e Turner (2015) é que a forma que um variograma marcado toma pode ser considerada bastante incomum para o caso geoestatístico. Espera-se, que em um fenômeno que pode ser modelado a partir da Geoestatística, a forma do variograma  $\gamma_Z(r)$  apresente valores pequenos para  $r$  pequeno e cresça conforme  $r$  aumenta até atingir um patamar. O caso do variograma marcado apresentado na Figura 2.11 é um exemplo do que seria uma situação incomum na Geoestatística.

## 2.6 Processos Pontuais Espaço-Temporais Marcados

Para introduzir a ideia de um processo pontual espaço-temporal marcado (PPETM)  $X_m$ , consideremos, conforme Iftimi, Cronie e Montes (2016), o processo pontual espaço temporal de base  $X$ , como apresentado na seção 2.4. De modo informal podemos atribuir as marcas  $m_i$  aos pontos de  $X$  para obter um PPETM  $X_m$ . Mais formalmente, um processo pontual espaço-

temporal marcado (PPETM)  $X_m$  pode ser considerado, de acordo com Iftimi, Cronie e Montes (2016), como um conjunto aleatório  $X_m = \{(\mathbf{s}_1, t_1, m_1), \dots, (\mathbf{s}_n, t_n, m_n)\}$  em que  $\mathbf{s}_i \in \mathbb{R}^d$ ,  $t_i \in \mathbb{R}$  e  $m_i \in \mathcal{M}$ , ou seja, o processo  $X_m$  ocorre em  $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$ . Aqui vamos considerar  $d = 2$  e  $\mathcal{M} = \mathbb{R}$ , de modo que temos o PPETM  $X_m$  em  $(\mathbb{R}^2 \times \mathbb{R}) \times \mathbb{R}$ .

### 2.6.1 Propriedades de primeira e segunda ordem

Assim como nos processos pontuais abordados até aqui, Iftimi, Cronie e Montes (2016) mostram que um PPETM também pode ser analisado utilizando funções intensidade. De acordo com os autores a função intensidade  $\lambda(\mathbf{s}, t, m)$  do processo  $X_m$  é dada por

$$\lambda(\mathbf{s}, t, m) = f_{(\mathbf{s}, t)}^{\mathcal{M}}(m) \lambda_g(\mathbf{s}, t), \quad (2.57)$$

em que  $f_{(\mathbf{s}, t)}^{\mathcal{M}}(\cdot)$  é a densidade da probabilidade condicional das marcas dos  $n$  pontos de  $X_m$ , dadas as suas localizações  $(\mathbf{s}, t)$ , e  $\lambda_b(\mathbf{s}, t)$  é a função intensidade do processo de base  $X$  (na seção 2.4 denotada apenas por  $\lambda(\mathbf{s}, t)$ ).

Considerado essa função de intensidade temos, a partir de Iftimi, Cronie e Montes (2016), a medida média<sup>1</sup> do processo  $X_m$  dado em  $B \times C$ , com  $B \subset \mathbb{R}^2 \times \mathbb{R}$  e  $C \subset \mathbb{R}$  dada por

$$\Lambda(B \times C) = \mathbb{E}[n(X_m \cap (B \times C))] = \int_{B \times C} \lambda(\mathbf{s}, t, m) dm ds dt. \quad (2.58)$$

No que se refere a homogeneidade, os autores apresentam, de modo análogo a Gabriel et al. (2013), as seguintes considerações:

- Se  $\lambda_b(\mathbf{s}, t) = \lambda_T(t)$  depende apenas de  $t \in \mathbb{R}$ , é dito que  $X_m$  é espacialmente homogêneo;
- Se  $\lambda_b(\mathbf{s}, t) = \lambda_S(\mathbf{s})$  depende apenas de  $\mathbf{s} \in \mathbb{R}^2$ ,  $X_m$  é dito espacialmente homogêneo;
- $X_m$  será dito espaço-temporalmente homogêneo, se o processo  $X$  é homogêneo, ou seja, se  $\lambda_b(\mathbf{s}, t) \equiv \lambda > 0$  e  $\lambda(\mathbf{s}, t, m) = f_{(\mathbf{s}, t)}^{\mathcal{M}} \lambda$ .  $X_m$  será não homogêneo em caso contrário.

<sup>1</sup> Os autores na realidade denotam por *medida de intensidade*, mas como apresentado anteriormente na página 17 a argumentação de Baddeley, Rubak e Turner (2015), de que esse termo pode causar alguma confusão, aqui será usado como *medida média*.

A partir da função intensidade pode-se definir, conforme Iftimi, Cronie e Montes (2016), a função de correlação entre pares, dada por

$$g((\mathbf{s}_2, t_1, m_1), (\mathbf{s}_2, t_2, m_2)) = \frac{f_{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)}^{\mathcal{M}}(m_1, m_2)}{f_{(\mathbf{s}_1, t_1)}^{\mathcal{M}}(m_1) f_{(\mathbf{s}_2, t_2)}^{\mathcal{M}}(m_2)} g_b((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) \quad (2.59)$$

em que  $g_b(\cdot)$  é função de correlação entre pares do processo  $X$  (na seção 2.4 denotada apenas por  $g(\cdot)$ ).

Nas seções anteriores foi mostrado que o processo de Poisson espaço-temporal é usado como referência para a ausência de interação (espacial, espaço-temporal, etc.). De modo análogo ao caso de um processo pontual espacial marcado (seção 2.5), um processo de Poisson  $X_m$  em  $(\mathbb{R}^2 \times \mathbb{R}) \times \mathbb{R}$  apresenta uma relação de independência entre os pontos e as marcas são i.i.d., isso, de acordo com os autores, implica que

$$g((\mathbf{s}_2, t_1, m_1), (\mathbf{s}_2, t_2, m_2)) = g_b((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) \equiv 1. \quad (2.60)$$

Tem-se a partir disso, que um PPETM  $Y$  com intensidade  $\lambda(\cdot)$  e  $g((\mathbf{s}_2, t_1, m_1), (\mathbf{s}_2, t_2, m_2)) > 1$  indica a existência de agrupamentos entre pontos de  $X_m$  localizados no entorno de  $(\mathbf{s}_1; t_1)$  e  $(\mathbf{s}_2, t_2)$ , com as marcas associadas  $m_1$  e  $m_2$ . Similarmente,  $g((\mathbf{s}_1, t_1, m_1); (\mathbf{s}_2, t_2, m_2)) < 1$  indica inibição.

A suposição de independência das marcas é, como mostrado até aqui, uma ferramenta comumente utilizada para as análises de processos pontuais marcados. No caso do PPETM, apoiados em Daley e Vere-Jones (2003) para o caso temporal, Iftimi, Cronie e Montes (2016) consideram as seguintes definições:

- $X_m$  tem marcas independentes se, dado o processo pontual não marcado  $Y_g$ , as marcas são variáveis aleatórias mutualmente independentes, tal que a distribuição de uma marca depende apenas da localização espaço-temporal do evento correspondente. Tem-se assim que  $f_{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_m, t_m)}^{\mathcal{M}}(m_1, \dots, m_n) = \prod_{i=1}^n f_{(\mathbf{s}_i, t_i)}^{\mathcal{M}}(m_i)$  para  $n \geq 1$ .
- Se as marcas do processo  $X_m$ , além de independentes, também forem identicamente distribuídas, então o processo  $Y$  terá a propriedade da rotulagem aleatória. Neste caso, tem-se que  $f_{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_m, t_m)}^{\mathcal{M}}(m_1, \dots, m_n) = \prod_{i=1}^n f^{\mathcal{M}}(m_i)$  para  $n \geq 1$ .

### 2.6.2 Estacionariedade de intensidade de segunda ordem reponderada

Esses conceitos relacionados com a função intensidade e a função de correlação entre pares, foi utilizado por Iftimi, Cronie e Montes (2016) para poder apresentar uma medida das propriedades de segunda ordem, denominada por eles de *função  $K$  não homogênea marcada*. No entanto, assim como no caso da função  $K$  espaço-temporal não homogênea, em que foi feita a suposição de que o processo é estacionário de intensidade de segunda ordem reponderada Baddeley, Møller e Waagepetersen (2000), a função proposta por Iftimi, Cronie e Montes (2016) necessita de uma suposição análoga, como descrita a seguir.

Baseados em Baddeley, Møller e Waagepetersen (2000), Iftimi, Cronie e Montes (2016) definem que um PPETM  $Y$  é estacionário de intensidade de segunda ordem reponderada (SOIRS – *Second-order intensity-reweighted stationary*), se a função de correlação entre pares existe e satisfaz

$$g((\mathbf{s}_1 + a, t_1 + b, m_1), (\mathbf{s}_2 + a, t_2 + b, m_2)) = g((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)), \quad (2.61)$$

para todo  $(a, b) \in \mathbb{R}^2 \times \mathbb{R}$ .

Os autores afirmam ainda que, se além de SOIRS, tem se que

$$g((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)) = \bar{g}(\|\mathbf{s}_1 - \mathbf{s}_2\|, |t_1 - t_2|, m_1, m_2), \quad (2.62)$$

ou seja, a função de correlação entre pares é dada por alguma função  $\bar{g}(\cdot)$  que espaço-temporalmente depende apenas das distâncias espaciais e das distâncias temporais, nos referimos a  $X_m$  como SOIRS *com isotropia* (SOIRSI).

### 2.6.3 Função $K$ não homogênea marcada

A função  $K$  não homogênea marcada é uma estrutura que busca descrever como pontos de  $X_m$ , com marcas em um subconjunto  $C \subseteq \mathcal{M}$ , interagem com pontos de  $X_m$ , com marcas em um subconjunto  $D \subseteq \mathcal{M}$ , dado que há uma separação espaço-temporal entre eles.

Dados esses conjuntos  $C$  e  $D$ , Iftimi, Cronie e Montes (2016) definem que para um PPETM SOIRS  $Y$ , com  $v(C), v(D) > 0$ , a função  $K$  não homogênea marcada é

$$K_{inhom}^{CD}(r, t) = \frac{1}{v(C)v(D)} \int_C \int_D \int_{\|\mathbf{s}\| \leq r} \int_{-t}^t g((0, 0, m_1), (\mathbf{s}, u, m_2)) ds du dm_2 dm_1. \quad (2.63)$$

Se  $Y$  é SOIRSI, a função  $K$  não homogênea marcada é

$$\begin{aligned} K_{inhom}^{CD}(r,t) &= \frac{1}{v(C)v(D)} \int_C \int_D \int_{\|s\| \leq r} \int_{-t}^t \bar{g}(\|s\|, |u|, m_2) ds du dm_2 dm_1 \\ &= \frac{1}{v(C)v(D)} \int_C \int_D \int_{-t}^t \int_0^r \omega_2 \bar{g}(p, q, m_1, m_2) p dp dq dm_2 dm_1. \end{aligned} \quad (2.64)$$

No caso do processo ser de Poisson em  $(\mathbb{R}^2 \times \mathbb{R}) \times \mathcal{M}$ , tem-se conforme Iftimi, Cronie e Montes (2016), que

$$K_{inhom}^{CD}(r,t) = 2tr^2 \omega_2 = 2tr^2 \pi. \quad (2.65)$$

Para efeitos de comparação, tem-se de acordo com os autores que

- Se  $K_{inhom}^{CD}(r,t) > 2tr^2 \omega_2$  há o indicativo de que os pontos de  $X_m$  com marcas em  $D$ , tem uma tendência a se agruparem em torno dos pontos com marcas em  $C$ .
- Quando  $K_{inhom}^{CD}(r,t) < 2tr^2 \omega_2$ , os pontos com marcas em  $D$  tem uma tendência a se afastar de pontos com marcas em  $C$ .

## 2.7 Análise de PPET usando o variograma marcado

Nas seções 2.4 e 2.5 foram apresentados conceitos sobre processos pontuais espaço-temporais e processos pontuais espaciais marcados, respectivamente. Nesta seção será apresentada a proposta de Stoyan et al. (2017) para análise de processos pontuais espaço-temporais a partir do variograma marcado.

A ideia consiste em transformar um processo pontual espaço-temporal em dois processos pontuais marcados, fazer a estimação dos variogramas marcados e ajustar a eles, um modelo gaussiano geoestatístico.

A estruturação desse procedimento está descrita a seguir.

### 2.7.1 Tempo e espaço como marcas

Seja o processo pontual espaço-temporal  $X = \{(s_1, t_2), (s_2, t_2), \dots\}$  definido na seção 2.4. Considerando que não existem pontos coincidentes e o processo ocorre em uma janela de tempo  $T = [0, \tau]$  e de espaço  $W$ , dada por um retângulo de lados  $a$  e  $b$ , Stoyan et al. (2017) afirmam que se pode transformar um processo pontual espaço-temporal em um processo pontual marcado de duas formas.

A primeira é reinterpretar o processo  $X$ , como um processo marcado  $X_T$  tomando os tempos  $t_i$  como marcas das localizações  $\mathbf{s}_i$ , ou seja,  $X_T = \{(\mathbf{s}_i, t(\mathbf{s}_i))\}$ . A segunda forma é reinterpretar o processo  $X$ , como um processo marcado  $X_W$  tomando as localizações  $\mathbf{s}_i$  como marcas dos tempos  $t_i$ , ou seja,  $X_W = \{(t_i, \mathbf{s}_i(t_i))\}$ . Os autores destacam que a distribuição de  $X_T$  e  $X_W$  depende das escolhas de  $T$  e  $W$ , o que leva a diferentes resultados na estimação do variograma, por exemplo. Stoyan et al. (2017) afirmam que as propriedades de invariância do processo  $X$  são transmitidas para  $X_T$  e  $X_W$ . Seja  $X$  um processo estacionário em que  $\mathbf{s}_i \in \mathbb{R}^2$  e  $t \in \mathbb{R}$  e

$$X \stackrel{d}{=} X^{s,t} \quad (2.66)$$

em que  $X^{s,t}$  é processo transladado  $X^{s,t} = \{(\mathbf{s}_1 + \mathbf{s}, t_1 + t), (\mathbf{s}_2 + \mathbf{s}, t_2 + t), \dots\}$ . Os autores apontam que se  $X$  é completamente estacionário, então  $X_T$  e  $X_W$  são estacionários como processos pontuais marcados.

### 2.7.2 O variograma marcado espacial

Para cada um dos processos  $X_T$  e  $X_W$ , Stoyan et al. (2017) consideram os respectivos variogramas marcados. De acordo com os autores, ambos descrevem a extensão das correlações entre os componentes espaciais e temporais, em particular a extensão do agrupamento espaço-temporal.

Para o caso do processo  $X_T$  Stoyan et al. (2017) definem o variograma marcado espacial  $\gamma_{sp}(r)$  por

$$\gamma_{sp}(r) = \frac{1}{2} \mathbb{E} \left[ (t(\mathbf{u}) - t(\mathbf{v}))^2 \mid \mathbf{u}, \mathbf{v} \in X_{space} \right] \quad (2.67)$$

em que  $t(\mathbf{s})$  é marca de tempo do ponto na localização  $\mathbf{s}$ , e  $N_{space}$  é a componente espacial do processo pontual espaço-temporal. Essa função tem basicamente a mesma estrutura da apresentada na equação (2.54) com a única diferença que as marcas são dadas pelos tempos.

A respeito do comportamento espaço-temporal do processo  $X$ , Stoyan et al. (2017) apontam que  $\gamma_{sp}(r)$  fornece uma valiosa informação e apresenta as seguintes características:

1. Se os pontos agregados no espaço tendem a ser também no tempo, então  $\gamma_{sp}(r)$  tem valores pequenos para  $r$  pequeno.
2. Com o aumento de  $r$ ,  $\gamma_{sp}(r)$  frequentemente aumenta e mostra um comportamento assintótico, convergindo em relação à um valor de patamar. A velocidade dessa convergência

está relacionada, com a velocidade no aumento da variabilidade dos tempos de aparecimento dos membros dos pares de pontos com o aumento da distância  $r$  entre os pontos. Isso significa que se a convergência em relação ao patamar é rápida, a extensão da dependência espaço-temporal existe para curtas distâncias. E se essa convergência é lenta, a extensão da dependência espaço-temporal existe para distâncias maiores.

3. O comportamento de  $\gamma_{sp}(r)$  próximo à origem pode ser caracterizado por um efeito pepita. De acordo com os autores, pode acontecer que, na mesma região do espaço, os clusters de pontos apareçam em momentos diferentes.

### 2.7.3 O variograma marcado temporal

Para o processo  $X_W$  Stoyan et al. (2017) definem o variograma marcado temporal  $\gamma_e(t)$  por

$$\gamma_e(t) = \frac{1}{2} \mathbb{E} \left[ (\mathbf{s}(l) - \mathbf{s}(k))^2 \mid l, k \in X_{time} \right] \quad (2.68)$$

em que  $\mathbf{s}(t)$  é marca espacial do ponto no tempo  $t$ , e  $X_{time}$  é a componente temporal do processo pontual espaço-temporal. Nesse caso as localizações são pontos em  $\mathbb{R}$  e as marcas estão em  $\mathbb{R}^d$ . Tem-se também que ao invés da distância espacial  $r$ , o que se considera é a distância temporal  $t$  e a diferença quadrática das marcas, tem-se a sua distância quadrática.

As propriedades de  $\gamma_e(t)$ , em relação ao processo pontual espaço-temporal  $X$  e apresentadas por Stoyan et al. (2017), são análogas ao caso de  $\gamma_{sp}(r)$ . Tem-se:

1. Se os pontos agregados no tempo tendem a ser também no espaço, então  $\gamma_e(t)$  tem valores pequenos para  $t$  pequeno.
2. De modo análogo ao comportamento de  $\gamma_{sp}(r)$  com  $r$  crescente, o variograma  $\gamma_e(t)$  mostra que com o aumento das diferenças de tempo  $t$ , as localizações dos pontos são mais dispersas. A interpretação da convergência de  $\gamma_e(t)$  em relação a um patamar é análoga ao caso de  $\gamma_{sp}(r)$ .
3. O comportamento de  $\gamma_e(t)$  próximo à origem pode ser caracterizado por um efeito pepita. Pode acontecer que, no mesmo intervalo de tempo, os clusters de pontos apareçam em lugares diferentes.

### 2.7.4 Estimação dos variogramas marcados $\gamma_{sp}(r)$ e $\gamma_{te}(t)$

Considerando que o processo pontual é estacionário, um estimador para  $\gamma_{sp}(r)$  apresentado por Stoyan et al. (2017) é

$$\widehat{\gamma}_{sp}(r) = \frac{\sum_{\mathbf{s}_1, \mathbf{s}_2 \in W}^{\neq} \frac{\frac{1}{2}(|t_1 - t_2|)^2 \kappa_{\varepsilon}(\|\mathbf{s}_1 - \mathbf{s}_2\| - r)}{2\pi r e_2(\mathbf{s}_1, \mathbf{s}_2)}}{\sum_{\mathbf{s}_1, \mathbf{s}_2 \in W}^{\neq} \frac{\kappa_{\varepsilon}(\|\mathbf{s}_1 - \mathbf{s}_2\| - r)}{2\pi r e_2(\mathbf{s}_1, \mathbf{s}_2)}}, \quad 0 < \varepsilon < r, \quad (2.69)$$

em que  $\kappa_{\varepsilon}$  é uma função kernel unidimensional com largura de banda  $\varepsilon$  e  $e_2(\mathbf{s}_1, \mathbf{s}_2)$  é um fator genérico para a correção de bordas. De acordo com Walder e Walder (2008), normalmente, funcões como o variograma marcado estimado s˜ao frequentemente abruptas e irregulares, alem de que e raro que se tenha pares de pontos com a mesma distˆancia exata  $r$ . Devido a isso e util entˆao utilizar estimadores baseados em kernel para obter funcões mais suaves.

Stoyan et al. (2017) afirmam, com base em resultados obtidos por meio de simulac˜ao, que o fator de correc˜ao do efeito de bordas pode ser dispensado e, portanto, n˜ao e necessario adotar formas, em geral, complicadas para a sua estimac˜ao. Os autores consideram entˆao, um estimador mais simples para  $\gamma_{sp}(r)$  dado por

$$\widehat{\gamma}_{sp}(r) = \frac{\sum_{\mathbf{s}_1, \mathbf{s}_2 \in W}^{\neq} \frac{1}{2}(|t_1 - t_2|)^2 \kappa_{\varepsilon}(\|\mathbf{s}_1 - \mathbf{s}_2\| - r)}{\sum_{\mathbf{s}_1, \mathbf{s}_2 \in W}^{\neq} \kappa_{\varepsilon}(\|\mathbf{s}_1 - \mathbf{s}_2\| - r)}. \quad (2.70)$$

Outro fato destacado por Stoyan et al. (2017) e que o estimador da equac˜ao (2.70) e robusto contra desvios do pressuposto de estacionariedade, mostrando que bons resultados podem ser obtidos mesmo que os dados apresentam falta de homogeneidade.

O estimador para  $\gamma_{te}(t)$  dado na equac˜ao (2.71) e analogo:

$$\widehat{\gamma}_{te}(t) = \frac{\sum_{t_1, t_2 \in T}^{\neq} \frac{1}{2}(\|\mathbf{s}_1 - \mathbf{s}_2\|)^2 \kappa_{\delta}(|t_1 - t_2| - t)}{\sum_{t_1, t_2 \in T}^{\neq} \kappa_{\delta}(|t_1 - t_2| - t)}, \quad (2.71)$$

em que a largura de banda  $\delta$  e o parˆametro de suavizac˜ao temporal.

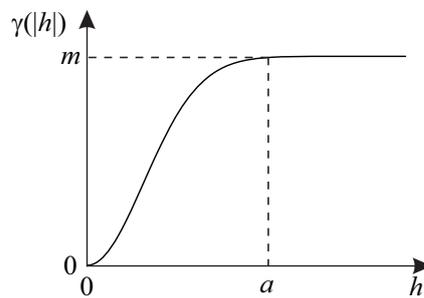
### 2.7.5 O modelo gaussiano

O modelo gaussiano geoestatıstico e definido, de acordo com, Montero, Fernandez-Aviles e Mateu (2015), como

$$\gamma(|h|) = m \left( 1 - \exp \left( -\frac{|h|^2}{a^2} \right) \right) \quad (2.72)$$

em que  $|h|$  é a distância entre duas localizações;  $a$  é o alcance da correlação, ou seja, a distância dentro da a qual as amostras são correlacionadas;  $m$  é o patamar, isto é, o valor do semivariograma que corresponde ao alcance  $a$ . Esses parâmetros podem ser visualizados na Figura 2.12 a seguir.

Figura 2.12 – Gráfico do modelo gaussiano teórico

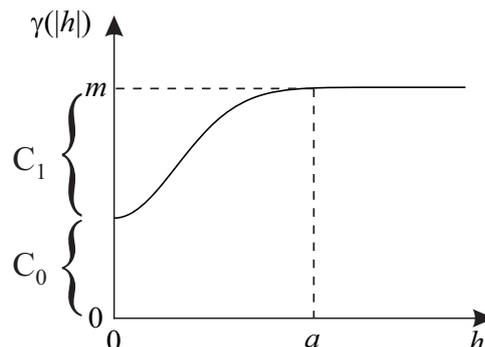


Fonte: Adaptado de Montero, Fernández-Avilés e Mateu (2015).

O valor de  $\gamma(0)$  é o efeito pepita e teoricamente é igual a zero. No entanto, na prática o que ocorre, segundo Camargo, Fucks e Câmara (2004), é que quando  $h \rightarrow 0$ ,  $\gamma(|h|)$  se aproxima de um valor positivo  $C_0$ . Este valor está relacionado com a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Assim, o efeito pepita representa a componente da variabilidade espacial que pode ser devida a erros de medição ou a variabilidade de pequena escala que não pode ser observada pela amostragem.

A partir disso, o valor do patamar é dado por  $m = C_0 + C_1$ , em que  $C_1$  é dito “contribuição”, e a forma do semivariograma gaussiano que se observa na prática é a que está apresentada na Figura 2.13 a seguir.

Figura 2.13 – Gráfico do modelo gaussiano teórico com efeito pepita maior que zero



Fonte: Adaptado de Camargo, Fucks e Câmara (2004).

Ao utilizar o modelo gaussiano, o objetivo de Stoyan et al. (2017) não foi fazer previsões a partir do modelo ajustado. O objetivos dos autores foi simplesmente utilizá-lo como ferramenta para obter estimativas desses parâmetros a fim de descrever a estrutura de correlação no processo pontual.

### 2.7.6 Considerações sobre os variogramas $\gamma_{sp}(r)$ e $\gamma_{te}(t)$

Um resultado apresentado por Stoyan et al. (2017) para os variogramas  $\gamma_{sp}(r)$  e  $\gamma_{te}(t)$  é o seu valor teórico sob a suposição de completa aleatoriedade espaço temporal, isto é, um processo de Poisson homogêneo. Considerando uma janela espacial  $W$  dada por um retângulo de lados  $a$  e  $b$ , e uma janela temporal  $T = [0, \tau]$ , tem-se

$$\gamma_{sp}(r) = \begin{cases} 0 & \text{para } r = 0 \\ c_{sp} & \text{para } r > 0 \end{cases} \quad \text{e} \quad \gamma_{te}(t) = \begin{cases} 0 & \text{para } t = 0 \\ c_{te} & \text{para } t > 0 \end{cases} \quad (2.73)$$

em que  $c_{sp} = \frac{\tau^2}{12}$  e  $c_{te} = \frac{a^2 + b^2}{12}$ .

Após fazer a estimação e ajustar o modelo gaussiano a dados simulados de PPET com diferentes estruturas e a dados reais, os autores Stoyan et al. (2017) concluíram que os variogramas estimados e ajustados apresentaram resultados consistentes com a estrutura dos dados. Outro aspecto destacado pelos autores é que utilizando o estimador proposto, os efeitos de borda e a falta de homogeneidade podem ser ignorados. Além disso, a forma da janela também é irrelevante, o que torna o estimador robusto para diferentes tipos de janela espacial e temporal.

O valor do patamar obtido pelo ajuste do modelo gaussiano para  $\gamma_{sp}(r)$  e  $\gamma_{te}(t)$  foi comparado com os valores de  $c_{sp}$  e  $c_{te}$ , respectivamente. Para os casos em que o valor do patamar se distanciou do esperado para um PPH, os autores argumentaram que isso se devia ao fato de que os dados dos padrões analisados estão longe de apresentar uma estrutura de um PPH. Contudo, não apresentaram uma forma de verificar isso utilizando os variogramas.

Neste trabalho, para os dados analisados essa verificação foi feita utilizando a abordagem das simulações de Monte Carlo para construção do envelope sob os modelos nulos de um PPH e de um PPNH. Esse procedimento está apresentado no capítulo 3.

## 2.8 Focos de calor e a potência radiativa do fogo (FRP)

Um tema que tem estado em foco é a questão das queimadas/incêndios. Nos últimos anos tem sido observada uma grande quantidade de casos em todo o mundo, principalmente em países como Austrália, Estados Unidos e Brasil. No caso do Brasil, por exemplo, só no ano 2020, até o mês de setembro foram registrados recordes de ocorrências de fogo na vegetação, principalmente na Amazônia e no Pantanal (DW Brasil, 2020).

Existe uma diferenciação entre os termos queimada e incêndios. De acordo com Gaspar (2012), a queimada é uma técnica utilizada basicamente por agricultores, com objetivos relacionados a limpar a área de cultivo, renovar a pastagem ou facilitar a colheita da cana-de-açúcar. Como aponta a autora, as cinzas decorrentes da queima da biomassa contribuem com a fertilidade do solo num primeiro momento, mas o fogo recorrente causa um empobrecimento do solo, pode acarretar em erosão, além de interferir na qualidade do ar e poder ocasionar também danos ao patrimônio público, como redes elétricas por exemplo.

No caso do incêndio o que se tem, conforme Gaspar (2012), é basicamente uma queimada descontrolada. De acordo com o Instituto Chico Mendes de Conservação da Biodiversidade - ICMBio (ICMBIO, 2016), aproximadamente 90% dos incêndios florestais nacionais tem origem antrópica, isto é, são causados pelo ser humano. As suas causas estão relacionadas, segundo Gaspar (2012), a pessoas descuidadas que jogam pontas de cigarro nas margens das estradas, ateiam fogo a lixões e ainda aqueles causados por balões, mas grande parte dos incêndios florestais tem motivos econômicos, sendo causados para ampliar áreas visando à criação de gado ou culturas agrícolas. Um exemplo disso, é conforme aponta Heal et al. (2019), que em 2019 na Amazônia brasileira, os incêndios foram três vezes mais comuns nas zonas de produção de carne, do que no restante do bioma.

As consequências dos incêndios podem ser observadas tanto em escala global como local. Segundo Schroeder et al. (2014), a queima de biomassa global nessas ocorrências de fogo na vegetação é um importante fator que afeta a física da atmosfera, influenciando assim no clima da Terra. Uma consequência disso, segundo os autores, pode ser a grande variabilidade interanual na atividade de queima de biomassa em escala regional, criando áreas sujeitas a incêndios florestais como resultado de condições climáticas mais secas e quentes.

A queima de biomassa também é, de acordo com Schroeder et al. (2014), um fator chave de perturbação em escalas regionais e de paisagem, podendo ser associada à uma rápida deterioração da qualidade do ar, bem como a efeitos duradouros nas propriedades da superfície,

incluindo mudança na cobertura da terra e alteração do escoamento. Ainda segundo os autores, a caracterização da queima de biomassa é crucial para permitir a quantificação de seus impactos nos biomas locais e nos *feedbacks* climáticos regionais e globais, além de contribuir para o desenvolvimento de estratégias de gestão para ajudar no combate aos incêndios florestais e para mitigar seus efeitos.

Como as observações no local das ocorrências não podem fornecer medições das emissões de fogo rotineiramente em todo o mundo, o sensoriamento remoto por satélite é frequentemente usado como um meio de analisar e avaliar as emissões de fumaça em escalas regionais e globais (WANG et al., 2018). No Brasil, o monitoramento dos focos de queimadas é realizado pelo Instituto Nacional de Pesquisas Espaciais - INPE, por meio de sensoriamento remoto por satélites e pode ser consultado livremente pela página da internet do Programa Queimadas (INPE, 2020). A informação capturada pelos satélites de monitoramento são os focos de calor, isto é, focos com temperatura acima de 47°C. O satélite de referência, ou seja, aquele cujos dados diários de focos detectados são usados para compor a série temporal ao longo dos anos, utilizado pelo INPE é o AQUA com o sensor MODIS, projetado pela Agência Espacial Norte-Americana (NASA).

O satélite AQUA está em órbita a uma altitude de 710 km e de acordo com informações do INPE (INPE, 2020), trabalhos de validação de campo indicam que uma frente de fogo com cerca de 30 metros de extensão por 1 metro de largura, ou maior, será detectada como foco de calor. Entretanto, a relação foco *versus* queimada não é direta nas imagens de satélite, sendo que um foco indica a existência de fogo em um elemento de resolução da imagem, isto é um píxel, que no caso do satélite AQUA mede 1 km × 1 km. Neste píxel pode haver uma ou várias frentes de fogo ativo distintas, mas a indicação será de um único foco. Pode ocorrer ainda que o evento de fogo na vegetação seja muito extenso, podendo ser detectado em alguns píxeis vizinhos, de modo que vários focos estarão associados assim, a um único grande evento.

Outro fato a considerar é o erro na localização dos focos de queima, que é de até aproximadamente 1 km das coordenadas indicadas, considerando o satélite AQUA. Além disso, há alguns fatores que prejudicam ou impedem a detecção dos focos calor. São eles: frentes de fogo com menos de 30 m; fogo apenas no chão de uma floresta densa, sem afetar a copa das árvores; nuvens cobrindo a região (nuvens de fumaça não atrapalham); queimada de pequena duração, ocorrendo entre o horário das imagens disponíveis; fogo ocorrendo em uma encosta de montanha, enquanto que o satélite só observou o outro lado (INPE, 2020).

O INPE salienta que, apesar desses fatores, que o sistema de Queimadas detecta a ocorrência de fogo, sendo um dado por si só válido e de extrema importância. De acordo com o instituto, os focos de calor não devem ser considerados como medida absoluta da ocorrência de fogo, que certamente é maior do que a indicada pelos focos, mas considerando o modo regular de detecção e tomando-se apenas um satélite como referência, é possível constatar tendências espaciais e temporais nas ocorrências de fogo na vegetação.

No que se refere à estimação da biomassa queimada na vegetação, um dos métodos que tem sido muito utilizados é derivado de uma quantidade chamada Potência Radiativa do Fogo – do inglês *Fire Radiative Power* - FRP (FREEBORN et al., 2008; MATAVELI et al., 2015; ROBERTS et al., 2005; WOOSTER; ZHANG, 2004; WOOSTER et al., 2005). Essa medida é, de acordo com Wooster e Zhang (2004), a taxa na qual a energia é emitida pelo fogo na forma de radiação, e é expressa em Joules por segundo, isto é, em *watts*.

A FRP é proporcional à taxa de consumo de biomassa, e a integração temporal da FRP, ao longo do período de ocorrência do fogo fornece uma estimativa da energia radiativa do fogo, que em experimentos de laboratório (COSTA; FONSECA, 2017; FREEBORN et al., 2008; PEREIRA et al., 2012; WOOSTER, 2002; WOOSTER et al., 2005), e em incêndios florestais (KONOVALOV et al., 2014; LI et al., 2018) mostrou haver uma relação linear com a quantidade total de biomassa consumida pelo fogo. A relação entre a FRP e taxa de biomassa queimada é dada pela equação:

$$TBQ = \beta \times FRP \quad (2.74)$$

em que  $TBQ$  é a taxa de biomassa queimada em  $kg.s^{-1}$ , a FRP é dada em  $MW$  e o coeficiente multiplicativo  $\beta$  está em  $kg.MJ^{-1}$ .

Considerando assim essa associação entre a biomassa queimada e a FRP, pode-se então utilizá-la como um indicador da severidade de uma queimada, e uma observação ao longo do tempo, para uma dada região espacial pode fornecer além do padrão espaço-temporal dos focos de calor, informações sobre como a intensidade das queimadas varia de um ano para outro por exemplo.

Detalhes técnicos de como os sensores dos satélites capturam as informações da FRP podem ser obtidos em Costa (2012), Freeborn et al. (2008), Ichoku e Kaufman (2005), Wooster (2002).

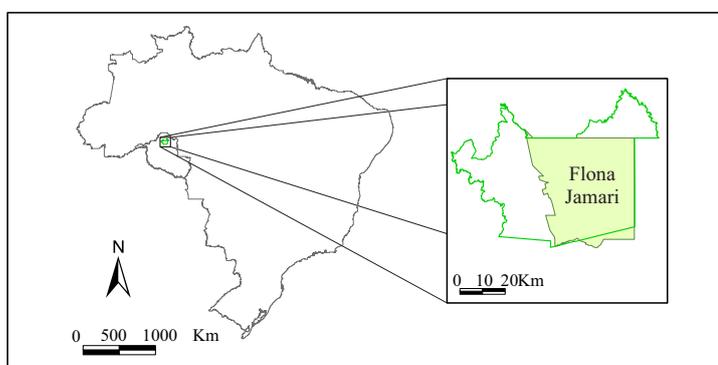
### 3 MATERIAL E MÉTODOS

#### 3.1 O conjunto de dados

##### 3.1.1 Dados reais

Os dados analisados nesse estudo correspondem a focos de calor, observados nos anos de 2018 e 2019 no município de Itapuã do Oeste - RO. O bioma do município é o amazônico e dos 4081 km<sup>2</sup> da sua área, aproximadamente 2119,32 km<sup>2</sup> são cobertos pela Floresta Nacional do Jamari (DIPLOMA LEGAL DE CRIAÇÃO: Decreto nº 90.224 de 25 de setembro de 1984).

Figura 3.1 – Localização do município de Itapuã do Oeste - RO e a Flona Jamari



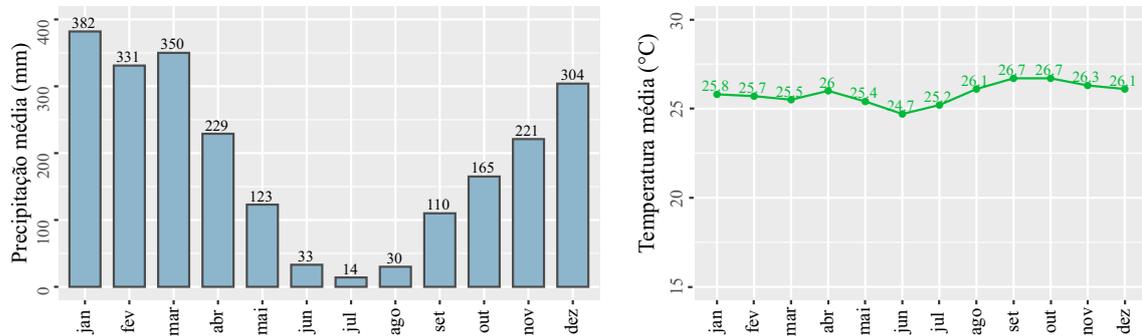
Fonte: Adaptado de Brasil. Ministério da Agricultura, Pecuária e Abastecimento. Serviço Florestal Brasileiro (2020)

O clima de Itapuã do Oeste é considerado tropical úmido ou subúmido pela classificação de Köppen e Geiger, possuindo curta época seca e temperatura média anual de 25,8°C, sendo setembro o mês mais quente, com temperatura média de 26,7°C e junho o mês com a menor temperatura média de 24,7°C (Figura 3.2). A pluviosidade média anual é de 2292 mm, em que julho é o mês mais seco com 14 mm e janeiro o mais chuvoso, com uma média de 382 mm (CLIMATE-DATA, 2020).

Os dados foram obtidos junto INPE, a partir da página da internet do Programa Queimadas (INPE, 2020), considerando os registros do satélite de referência AQUA. Os dados correspondem a 216 focos de calor e os respectivos valores de FRP (em megawatts - MW), sendo 111 focos observados em 2018 e 95 em 2019. As informações utilizadas em cada foco são as coordenadas geográficas, o tempo da ocorrência formado pela data e hora, e os valores de FRP associados a cada foco.

Para a análise foi necessário converter as coordenadas geográficas para o sistema UTM (Universal Transversa de Mercator), que é um sistema de coordenadas cartesianas.

Figura 3.2 – Gráfico da precipitação média e da temperatura média mensais para Itapuã do Oeste - RO.



Fonte: Adaptado de Climate-Data (2020)

As informações de data e hora, foram transformadas para dias, sendo que a data foi convertida para dia do ano e a hora foi calculada como fração do dia.

### 3.1.2 Dados simulados

Além da análise com dados reais, também foi feito o uso de dados simulados para o método proposto nesta tese, que está apresentado na seção 4.2. Esses dados simulados, constituem-se em padrões pontuais espaço-temporais marcados, sendo o processo pontual de base estacionário apresentando as estruturas de inibição, completamente aleatório, infecção (agrupamento), e um processo não estacionário, que é o processo de Poisson não homogêneo. As marcas foram simuladas para serem i.i.d. e independentes dos pontos do processo de base, sendo retiradas de duas distribuições de probabilidade diferentes: uma normal  $N(\mu = 50, \sigma = 20)$  e uma exponencial  $\text{Exp}(\lambda = 1/50)$ .

Foram considerados cenários com 100 e 500 pontos, distribuídos na mesma região espacial dos dados reais, no intervalo temporal  $[0, 365]$  dias.

Esses dados foram simulados para verificar a correspondência entre o resultado teórico esperado e a forma gráfica do variograma marcado proposto para a análise do PPETM, sob a hipótese nula de rotulagem aleatória. Para cada cenário, o gráfico foi construído a partir da média de 100 cálculos do variograma.

## 3.2 Análise do PPET usando o variograma marcado

Na análise dos dados a partir da proposta de Stoyan et al. (2017), para cada um dos variogramas  $\gamma_{sp}(r)$  e  $\gamma_e(t)$ , foi feita a estimação e o ajuste do modelo gaussiano.

Posteriormente, para investigar os desvios do variograma estimado em relação ao esperado para o caso da completa aleatoriedade, foram construídos os envelopes a partir de 199 simulações de um PPH, com o mesmo número de pontos e na mesma região espaço-temporal dos dados.

Outro modelo nulo que também foi considerado para essas análises foi o PPNH. Isso deve-se ao fato de que se o variograma apresenta um patamar muito menor do que esperado para um PPH, então uma variabilidade muito pequena entre as localizações e tempos pode formar agrupamentos que não são devidos a um efeito de segunda ordem. Isto é, o processo pode ser não estacionário, possuindo assim uma intensidade de primeira ordem não homogênea, de modo que se tem um efeito de tendência nas ocorrências dos focos de calor e não de dependência.

No caso da simulação de um PPNH, a intensidade não homogênea  $\lambda(\mathbf{s}, t)$  pode ser introduzida como uma função das coordenadas ou um objeto *array* tridimensional, com as intensidades em cada ponto da região espaço-temporal (que foi o caso utilizado aqui). A construção desse objeto *array* pode ser feita a partir da estimação da intensidade do processo observado utilizando as funções kernel. A descrição desse processo é feita em detalhes em Gabriel et al. (2013).

Se o variograma estimado não permanecer dentro do envelope, mesmo considerando o PPNH como modelo nulo, teremos então que o comportamento de agrupamento observado não foi devido simplesmente a um efeito de tendência, mas sim de dependência.

### 3.3 Software

Todas as análises foram executadas utilizando funções desenvolvidas no ambiente computacional *software* R (R Core Team, 2020). As bibliotecas (pacotes) utilizadas foram:

- *sp* (BIVAND et al., 2008; PEBESMA; BIVAND, 2005) , *maptools* (BIVAND; LEWINKOH, 2020), *rgdal* (BIVAND; KEITT; ROWLINGSON, 2020)— Conversão de coordenadas e estruturas de mapas;
- *lubridate* (GROLEMUND; WICKHAM, 2011) — Conversão de datas;
- *stpp* (GABRIEL et al., 2013) — Simulação de dados espaço-temporais, cálculo dos variogramas  $\gamma_{sp}(r)$  e  $\gamma_e(t)$ ;
- *spatstat* (BADDELEY; TURNER, 2005) — Simulação de dados espaciais, análise exploratória;

- *gridExtra* (AUGUIE, 2017), *ggplot2* (WICKHAM, 2016), *plot3D* (SOETAERT, 2019), *lattice* (SARKAR, 2008), *RColorBrewer* (NEUWIRTH, 2014) – Construção de gráficos;
- *kedd* (GUIDOUM, 2015) – Cálculo usando função kernel;
- *gdata* (WARNES et al., 2017) – Operações com matrizes;
- *geoR* (RIBEIRO Jr et al., 2020) – Ajuste do modelo gaussiano ao variograma.

## 4 RESULTADOS E DISCUSSÃO

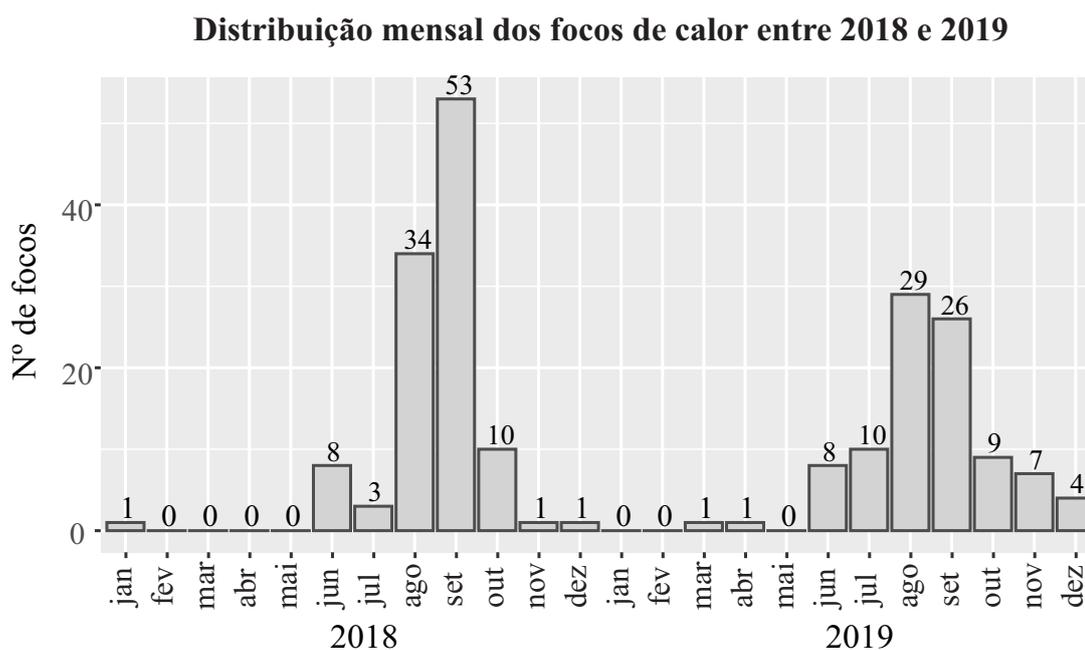
Nesse capítulo os resultados estão divididos em duas partes. Na primeira (seção 4.1) estão apresentados os resultados das análises dos dados de focos de calor, no que se refere a análise descritiva e análise espaço-temporal do PPET sem as marcas usando o variograma marcado a partir da abordagem de (STOYAN et al., 2017). Na segunda parte (seção 4.2) está apresentado o desenvolvimento da análise baseada no variograma marcado espaço-temporal para um PPETM, com a aplicação a dados simulados e também aos dados de focos de calor.

### 4.1 Análises relativas às componentes espacial e temporal dos dados de focos de calor

#### 4.1.1 Análise descritiva

Considerando primeiramente a distribuição temporal dos dados, na Figura 4.1 está apresentado o número de casos por mês durante o período de estudo. Pode-se notar um comportamento sazonal no número de focos de calor, sendo que para ambos anos de 2018 e 2019, no período de junho a outubro é onde está concentrada a grande maioria das ocorrências. É possível observar que, dentro de cada ano, há um comportamento não uniforme na distribuição dos focos, sendo os meses de agosto e setembro, aqueles com os maiores números de ocorrências.

Figura 4.1 – Distribuição mensal dos focos de calor em Itapuã do Oeste - RO entre 2018 e 2019

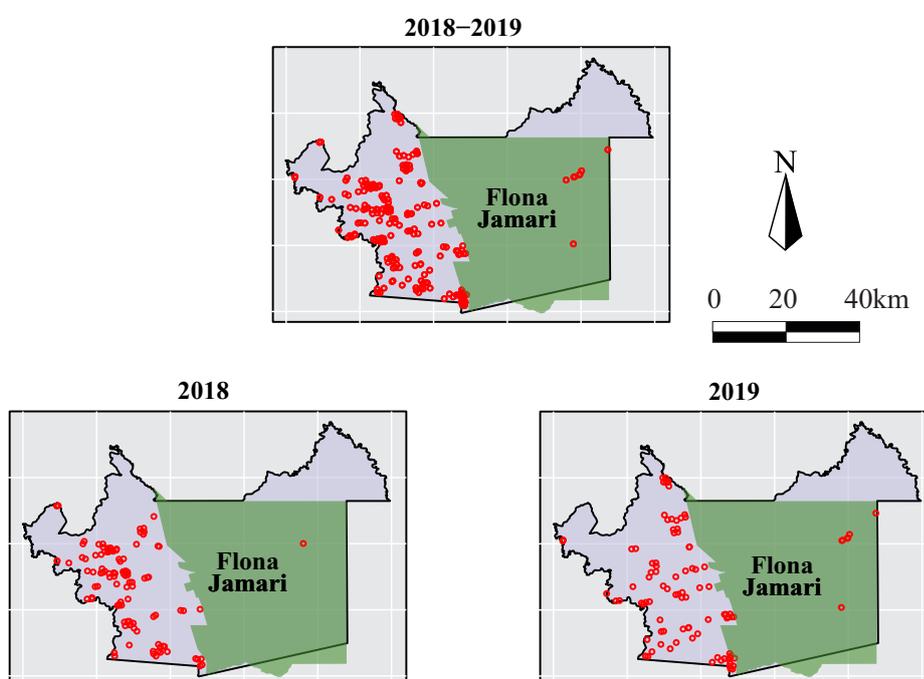


Fonte: Do autor (2020)

Este período de alta incidência dos focos de calor coincide com as épocas de menor intensidade de chuvas (Figura 3.2), sendo esse período em que se observa também uma baixa umidade do ar e altas temperaturas atmosféricas. Esses fatores, de acordo com Corrêa (2007) são propícios para ocorrência de fogo em áreas naturais.

No que se refere à distribuição espacial dos focos de calor, na Figura 4.2 temos as localizações de cada um dos focos registrados para o período de 2018 a 2019, bem como para cada um dos anos em separado.

Figura 4.2 – Distribuição espacial dos focos de calor em Itapuã do Oeste - RO entre 2018 e 2019. As unidades dos mapas são dadas em quilômetros.

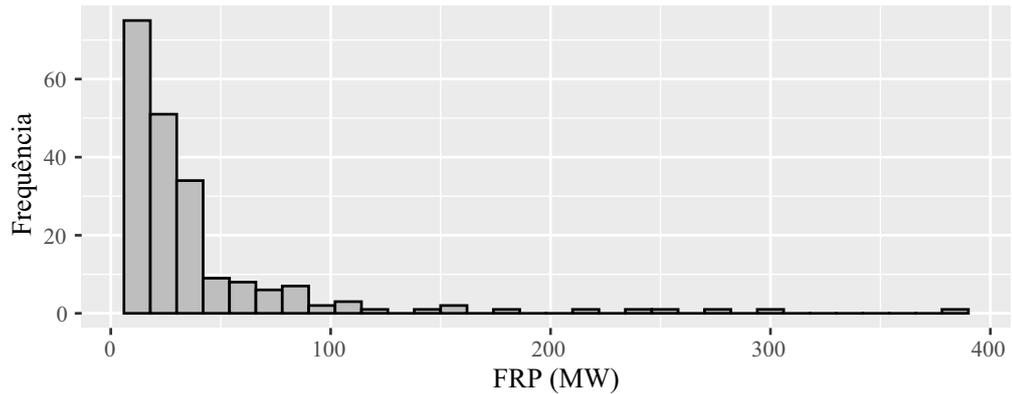


Fonte: Do autor (2020)

Pode-se notar, que a grande maioria dos focos foi observada fora da área da Floresta Nacional do Jamari. Esse é um resultado importante, pois reforça a importância das unidades de conservação ambiental, no que se refere ao combate e prevenção de incêndios e a preservação da biodiversidade. Conforme apontam Ramos, Fonseca e Morello (2016), a capacidade que as unidades de conservação possuem de controlar o uso e ocupação do solo e a pressão demográfica humana, tendem a minimizar ou impedir a ocorrência de incêndios no seu interior. E na Amazônia, mesmo muitas unidades de conservação tendo pouca estrutura e não possuindo brigadas, mostraram-se eficazes em conter incêndios e desmatamento (NEPSTAD et al., 2006).

Considerando agora a distribuição dos valores de FRP associados a cada um dos focos de calor, observou-se que a maioria dos focos apresentou baixos valores, sendo o valor da mediana igual a 22,35 MW, o terceiro quartil igual a 38,92 MW e apenas 6,8% apresentam valor superior a 100 MW. Na Figura 4.3 a seguir está apresentado o histograma para os valores de FRP.

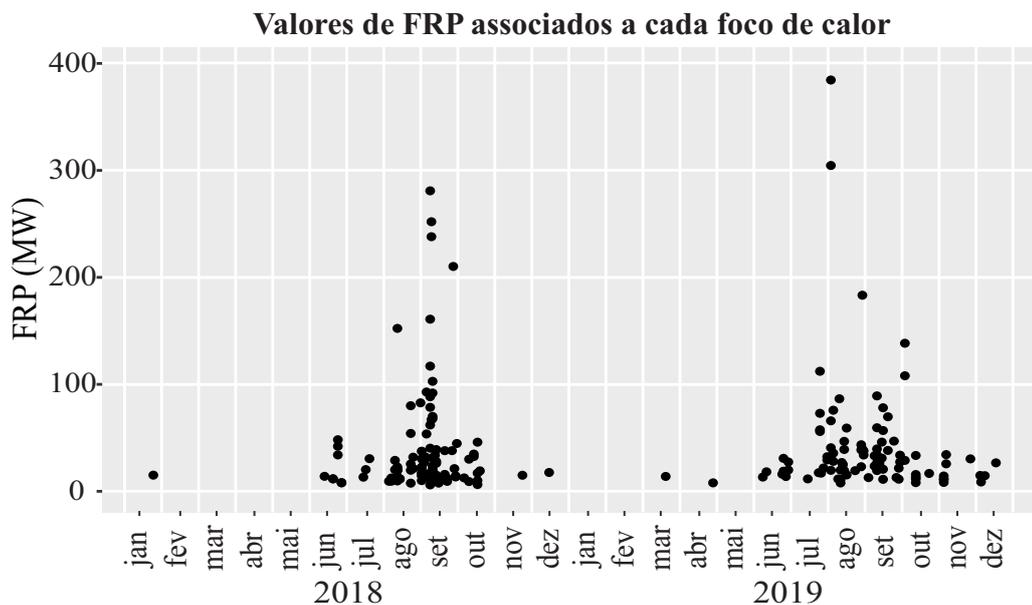
Figura 4.3 – Histograma dos valores de FRP associados a cada foco de calor



Fonte: Do autor (2020)

No que se refere à distribuição temporal dos valores de FRP, podemos observar a partir da Figura 4.4, que nos períodos com baixa intensidade de focos de calor, os valores de FRP tendem a ser baixos e, a partir do momento que essa intensidade aumenta, tende a aparecer focos com valores mais elevados de FRP.

Figura 4.4 – Distribuição dos valores de FRP associados a cada foco de calor em Itapuã do Oeste - RO entre 2018 e 2019

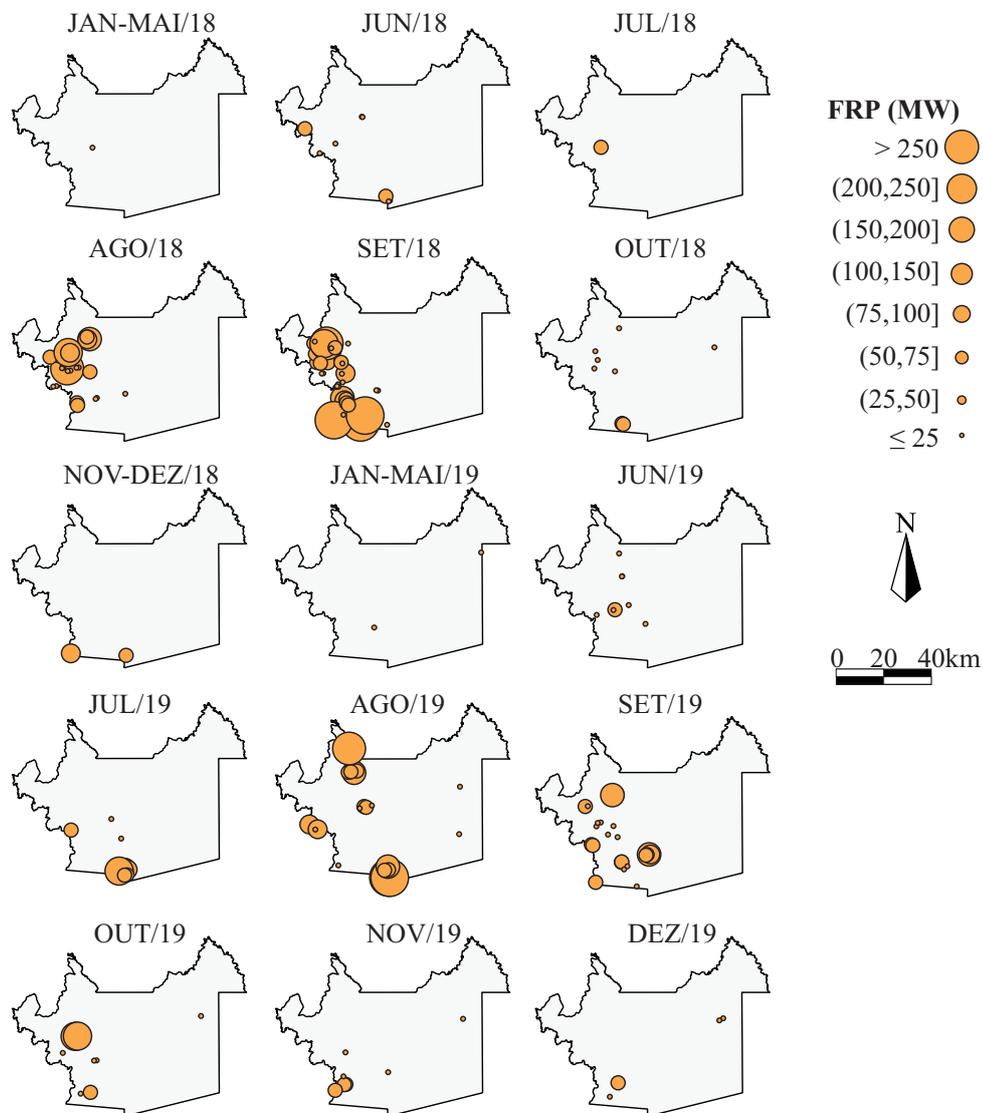


Fonte: Do autor (2020)

Comparando a Figura 4.4 com a Figura 4.1, pode-se observar que em cada um dos anos 2018 e 2019, os maiores valores de FRP foram observados nos meses de agosto e setembro, nos quais as frequências de focos de calor observadas foram as mais elevadas. Ou seja, nos períodos em que se observou o maior número de focos de calor, nos períodos mais secos, é onde ocorreram as queimadas mais severas.

Com relação à distribuição espacial da FRP, na Figura 4.5 está apresentada a evolução a focos de calor com os valores de FRP, para diferentes períodos ao longo dos anos 2018 e 2019.

Figura 4.5 – Distribuição dos espacial dos valores de FRP associados a cada foco de calor em Itapua do Oeste - RO entre 2018 e 2019



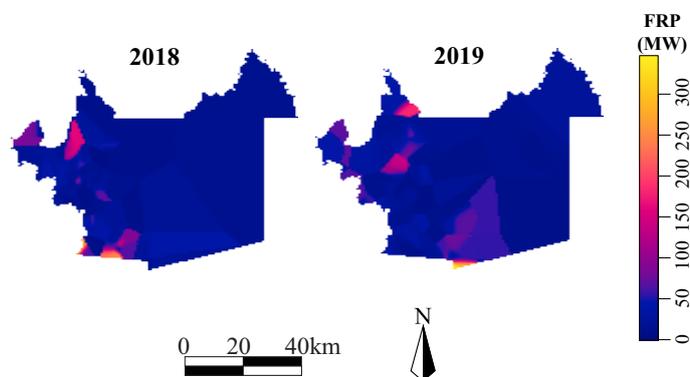
Fonte: Do autor (2020)

A partir de uma análise visual da Figura 4.5 pode-se observar que a distribuição dos focos não ocorre de maneira uniforme. Nos períodos de menor intensidade, os focos ocorreram de modo mais disperso, enquanto nos períodos de maior intensidade, o padrão foi mais agrupado.

No caso da FRP, foi observado para alguns períodos uma similaridade entre focos próximos uns dos outros, como por exemplo, nos meses de agosto e outubro de 2018, agosto, setembro e novembro de 2019, de modo que tem-se valores pequenos próximos a valores pequenos, e o mesmo comportamento para valores mais altos. Em outros períodos no entanto, como setembro de 2018, junho e outubro de 2019, pode-se notar focos com valores mais altos tendo outros com valores mais baixos no seu entorno.

Considerando a distribuição dos focos de calor, também foi observado que de 2018 para 2019 houve uma “migração” nas localizações dos focos (ver também a Figura 4.2). Em certas regiões onde houve uma maior (menor) incidência no ano 2018, em 2019 o número observado foi menor (maior). Resultado similar foi observado em termos da FRP, conforme apresentado na Figura 4.6, em que está apresentada para os dois anos, as marcas espacialmente suavizadas pelo estimador de Nadaraya-Watson.

Figura 4.6 – Valores de FRP espacialmente suavizados pelo estimador de Nadaraya-Watson para os focos de calor em Itapuã do Oeste - RO entre 2018 e 2019



Fonte: Do autor (2020)

Pode-se observar que as regiões com maiores valores de FRP no ano 2018 foram aquelas que apresentaram os menores valores no ano seguinte. Uma razão para isso, é que de um ano para outro não houve reincidência em uma área que foi queimada. Esse resultado vai ao encontro dos resultados de Barbosa (2001), em que a maioria das áreas que são queimadas anualmente, são de áreas que não queimaram no ano anterior, o que indica uma maior quantidade de biomassa para queimar em períodos seguintes devido ao acúmulo de combustível.

#### 4.1.2 Análise do PPET usando o variograma marcado

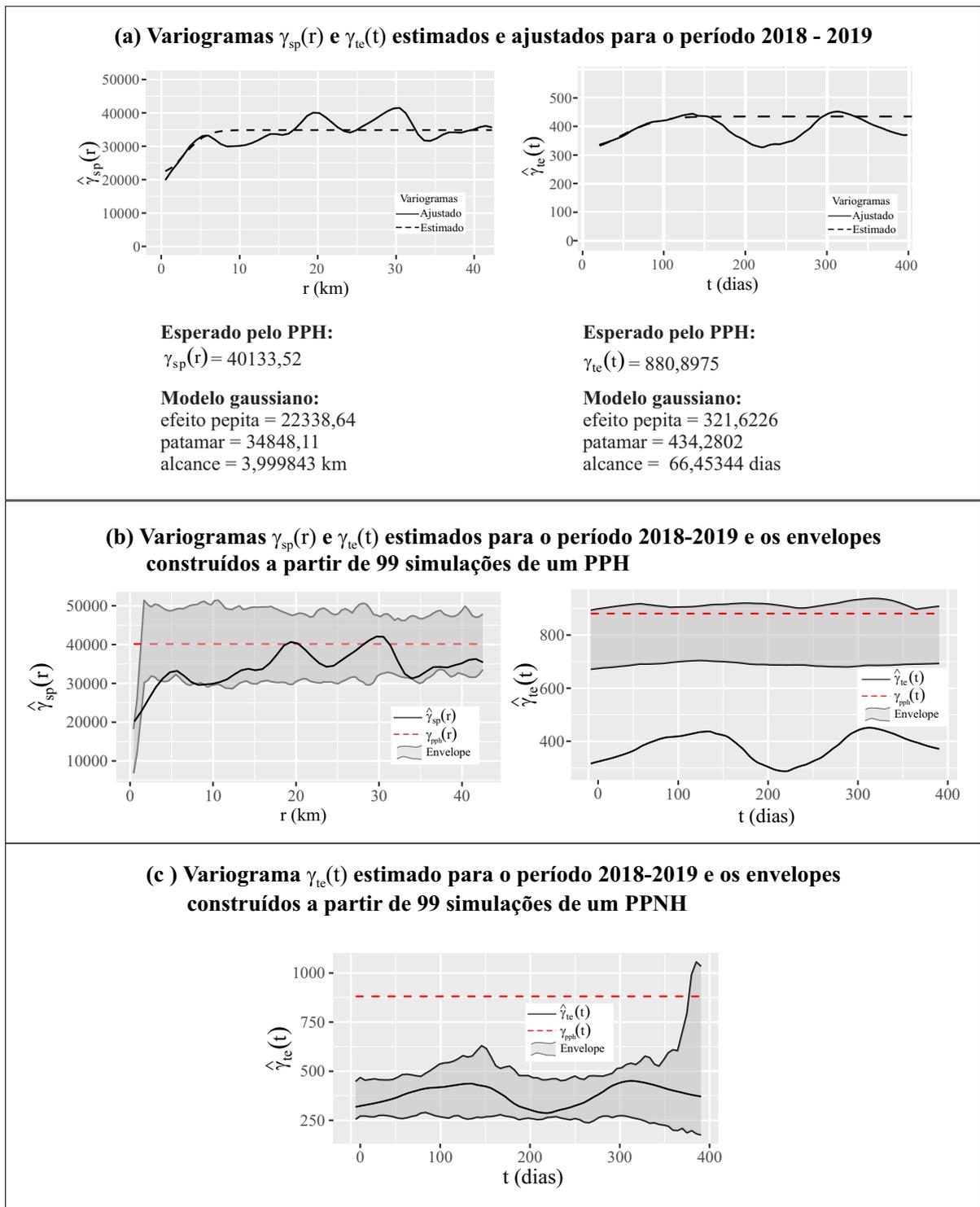
Os resultados para a análise do processo pontual espaço-temporal não marcado para os dados dos focos de calor, a partir da abordagem proposta por Stoyan et al. (2017), e também considerando os envelopes de simulação estão apresentados na Figura 4.7 a seguir.

A Figura 4.7 (a) apresenta os variogramas marcados  $\gamma_{sp}(r)$  e  $\gamma_e(t)$  estimados e ajustados pelo modelo gaussiano para os focos de calor em Itapuã do Oeste - RO entre 2018 e 2019, juntamente com os respectivos valores dos parâmetros do modelo gaussiano ajustado e o valor teórico esperado do variograma para o caso de um PPH. Pelo variograma  $\gamma_{sp}(r)$  observa-se que o valor do patamar do modelo gaussiano corresponde a aproximadamente 87% do esperado para o caso de um PPH. Observa-se também um alcance de, aproximadamente, 4 quilômetros, indicando que existe uma estrutura de correlação entre os tempos das ocorrências até essa distância e depois o padrão passa a ser aleatório. Esse comportamento também pode ser observado na Figura 4.7 (b) em que para distâncias de, aproximadamente, 1 a 4 quilômetros,  $\hat{\gamma}_{sp}(r)$  fica fora do envelope construído sob o modelo nulo do PPH e se mantém no seu interior a partir disso. Como o valor do efeito pepita está abaixo do patamar, isso indica que pontos próximos no espaço ocorreram próximos no tempo, podendo assim haver alguma forma de agrupamento dos focos de calor.

No caso de  $\gamma_e(t)$ , na Figura 4.7 (a), o modelo gaussiano apresenta um alcance de aproximadamente 66 dias, ou seja, para esse intervalo de tempo, o modelo indica uma estrutura de correlação entre as ocorrências dos focos de incêndios. Observa-se, contudo, que o patamar do modelo gaussiano é, aproximadamente, metade do esperado para um PPH, de modo que mesmo para grandes intervalos de tempo, a variabilidade nesses intervalos entre os focos é muito menor do que se esperaria em um PPH, o que indica uma estrutura de agrupamentos no tempo. De fato, observando a Figura 4.7 (b) tem-se que  $\hat{\gamma}_e(t)$  permanece sempre fora do envelope, o que leva a rejeição da hipótese do processo ser Poisson homogêneo.

Para esse último caso, foi realizada ainda a construção dos envelopes utilizando 199 simulações de um processo de Poisson não homogêneo cuja intensidade  $\lambda(s, t)$  atribuída foi dada por um *array*, com cada elemento sendo o valor da intensidade espaço-temporal do processo observado. Na Figura 4.7 (c) pode-se verificar que o gráfico de  $\hat{\gamma}_e(t)$  permanece sempre dentro

Figura 4.7 – Resultados da análise espaço-temporal para o período 2018-2019 utilizando o modelo gaussiano e os envelopes de simulação.



Fonte: Do autor (2020)

do envelope, indicando assim que o padrão observado pode ser modelado como um processo de Poisson não homogêneo. Isso significa, que a estrutura de agrupamento ocorre por variações no valor médio do processo e não por dependência entre os focos.

Observa-se de, modo geral, que o padrão espaço-temporal dos focos de calor para o período de estudo, apresenta um comportamento mais de tendência do que de dependência. Uma justificativa para isso pode ser que, temporalmente, o período em que os focos são observados tem um padrão de sazonalidade, ocorrendo em maior quantidade na época mais seca. Já espacialmente, os casos tendem a ocorrer nos locais em que não há proteção ambiental, sendo o ser humano o principal agente a iniciar queimadas (ICMBIO, 2016).

Observa-se ainda, que nas análises espaço-temporais utilizando os variogramas  $\gamma_{sp}(r)$  e  $\gamma_{te}(t)$ , o uso dos envelopes de simulação são importantes para complementar a análise proposta por Stoyan et al. (2017). Uma vez que, quando há desvios do variograma em relação ao valor esperado sob a completa aleatoriedade para grandes distâncias, a utilização do modelo gaussiano simplesmente não permite verificar se o padrão observado realmente exibe dependência entre as ocorrências, ou se há apenas um efeito de tendência no processo.

Apesar de mostrar que a análise do PPET fazendo o uso do variograma marcado apresenta resultados interessantes, Stoyan et al. (2017), no entanto, não justificam a escolha do modelo gaussiano, quais as suas implicações em relação ao uso de outros modelos válidos como o esférico ou exponencial, por exemplo. No caso dessa tese, buscou-se apenas complementar a metodologia acrescentando o uso dos envelopes de simulação, portanto, por não ser o foco do trabalho, também não foi discutido o mérito da escolha do modelo gaussiano, de modo que futuros trabalhos podem ser desenvolvidos para abordar essa questão.

## 4.2 Proposta de análise de um PPETM utilizando o variograma marcado

Para o caso de um PPETM estacionário, a proposta dessa tese é utilizar o variograma marcado como ferramenta de análise, de modo que o procedimento consista basicamente em duas partes: uma análise marginal, em que as marcas serão analisadas considerando no processo  $X$ , ora apenas a componente espacial e ora apenas a componente temporal e a segunda análise com as duas componentes conjuntamente. Para ambos os casos será assumido que não há pontos coincidentes no processo, ou seja, para todo  $(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j) \in X$ , tem-se  $(\mathbf{s}_i, t_i) \neq (\mathbf{s}_j, t_j)$ . O processo de análise está detalhado a seguir.

### 4.2.1 Análise marginal do PPETM

Sejam  $\{(\mathbf{s}_1, t_1, m_1), \dots, (\mathbf{s}_n, t_n, m_n)\}$  os pontos do PPETM  $X$  em  $(\mathbb{R}^2 \times \mathbb{R}) \times \mathcal{M}$  (lembrando que aqui será considerado  $\mathcal{M} = \mathbb{R}$ ). Vamos separar  $X$  em dois processos  $X_S$  e  $X_T$  tais

que

$$X_S = \{(\mathbf{s}_1, m_1), \dots, (\mathbf{s}_n, m_n)\}, \text{ em que } (\mathbf{s}_i, m_i) \in \mathbb{R}^2 \times \mathcal{M} \quad (4.1)$$

e

$$X_T = \{(t_1, m_1), \dots, (t_n, m_n)\}, \text{ em que } (t_i, m_i) \in \mathbb{R} \times \mathcal{M}. \quad (4.2)$$

Ambos os processos serão analisados a partir do variograma marcado, sendo que, para  $X_S$  tem-se

$$\gamma_S(r) = \frac{1}{2} \mathbb{E} \left[ (m(\mathbf{s}_i) - m(\mathbf{s}_j))^2 \mid \|\mathbf{s}_i - \mathbf{s}_j\| = r \right] \quad (4.3)$$

exatamente como na equação (2.54), e para  $X_T$  tem-se

$$\gamma_T(t) = \frac{1}{2} \mathbb{E} \left[ (m(t_i) - m(t_j))^2 \mid |t_i - t_j| = t \right]. \quad (4.4)$$

A interpretação dos resultados desses variogramas é análoga ao apresentado na seção 2.5.2. No entanto, os resultados mostrados por  $\gamma_S(r)$ , por exemplo, vão exibir de um modo geral a estrutura de dependência das marcas ao longo do espaço, sem considerar as variações no tempo, e de modo análogo para  $\gamma_T(t)$ .

Para a estimação dos variogramas marcados, a estrutura de  $\hat{\gamma}_S(r)$  e  $\hat{\gamma}_T(t)$ , será análoga a de  $\hat{\gamma}_{sp}(r)$  e  $\hat{\gamma}_{te}(t)$ , utilizada por Stoyan et al. (2017), de modo que

$$\hat{\gamma}_S(r) = \frac{\sum_{\mathbf{s}_1, \mathbf{s}_2 \in W}^{\neq} \frac{1}{2} (|m_1 - m_2|)^2 \kappa_\varepsilon(\|\mathbf{s}_1 - \mathbf{s}_2\| - r)}{\sum_{\mathbf{s}_1, \mathbf{s}_2 \in W}^{\neq} \kappa_\varepsilon(\|\mathbf{s}_1 - \mathbf{s}_2\| - r)}. \quad (4.5)$$

e

$$\hat{\gamma}_T(t) = \frac{\sum_{t_1, t_2 \in T}^{\neq} \frac{1}{2} (|m_1 - m_2|)^2 \kappa_\delta(|t_1 - t_2| - t)}{\sum_{t_1, t_2 \in T}^{\neq} \kappa_\delta(|t_1 - t_2| - t)}, \quad (4.6)$$

onde,  $W \subset \mathbb{R}^2$ ,  $T \subset \mathbb{R}$ ,  $\kappa$  é uma função kernel e  $\varepsilon$ ,  $\delta$  são as larguras de banda espacial e temporal, respectivamente.

E por fim, tem-se a construção dos envelopes de simulação, considerando o modelo nulo de rotulagem aleatória, a partir de 199 permutações das marcas e com nível de significância  $\alpha = 0,05$  para cada um dos casos.

#### 4.2.2 Análise conjunta do PPETM

Como mostrado anteriormente, para o caso puramente espacial, o variograma é uma ferramenta que existe tanto para a geoestatística, como para processos pontuais marcados, contudo dadas as especificidades desses processos, cada variograma tem as suas características.

No caso de dados espaço-temporais, o variograma também é uma ferramenta utilizada na geoestatística, de modo a levar em consideração ambas componentes espacial e temporal, e é dado conforme Montero, Fernández-Avilés e Mateu (2015) para um processo estacionário por

$$\gamma(\mathbf{h}, u) = \frac{1}{2} \mathbb{E} \left[ (Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2 \right], \quad (4.7)$$

em que  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$  e  $u = t_i - t_j$ .

Será apresentado aqui então, a partir de uma suposição de estacionariedade, uma expressão para o variograma marcado  $\gamma_{ST}(r, t)$  de um PPETM, e o seu estimador  $\hat{\gamma}_{ST}(r, t)$ . Define-se então, o variograma marcado para um PPETM estacionário  $X_m$ , dado no produto cartesiano  $(\mathbb{R}^2 \times \mathbb{R}) \times \mathbb{R}$ , por

$$\gamma_{ST}(r, t) = \frac{1}{2} \mathbb{E} \left[ (m(\mathbf{s}_i, t_i) - m(\mathbf{s}_j, t_j))^2 \mid \mathbf{s}_i, \mathbf{s}_j \in X_{space}; t_i, t_j \in X_{time} \right], \quad (4.8)$$

em que  $\|\mathbf{s}_i - \mathbf{s}_j\| = r$ ,  $|t_i - t_j| = t$ ,  $X_{space}$  e  $X_{time}$  representam, respectivamente, as componentes puramente espacial e temporal do processo.

Tendo essa estrutura,  $\hat{\gamma}_{ST}(r, t)$  graficamente é uma superfície, ao invés de uma curva, como nos casos puramente espacial ou temporal. Assim, dado um PPETM  $X_m$ , esperada-se em analogia a Stoyan et al. (2017) que:

- Se as marcas de pontos com uma distância espaço-temporal pequena forem semelhantes, para pequenos valores de  $r$  e  $t$ ,  $\gamma_{ST}(r, t)$  tem valores pequenos. Pode ser, no entanto, que marcas de valores muito distintos ocorram a pequenas distâncias espaço-temporais, o que levaria a grandes valores de  $\gamma_{ST}(r, t)$ , para  $r$  e  $t$  pequenos.
- O comportamento de  $\gamma_{ST}(r, t)$  próximo à origem (efeito pepita), denotado por  $\gamma_{ST}(0+, 0+)$ , reflete a agregação espaço-temporal de pontos próximos com marcas diferentes, ou seja, um valor alto de  $\gamma_{ST}(0+, 0+)$  é um indicativo de que pontos próximos tendem a ter marcas significativamente diferentes.

No caso de  $\gamma_{ST}(0+, 0+)$  ser constante para  $r, t > 0$ , tem-se então um efeito pepita puro, ou seja, as marcas são i.i.d. de modo a não existir nenhuma estrutura de correlação espaço-temporal entre elas.

Considerando esse último cenário, teremos que

$$\gamma_{ST}(r, t) = \begin{cases} 0 & \text{se } r, t = 0 \\ \sigma_m^2 & \text{se } r, t > 0 \end{cases} \quad (4.9)$$

em que,  $\sigma_m^2$  é a variância das marcas.

**Prova:** Se as marcas são i.i.d. e não dependem das localizações dos pontos, então tem-se  $\mathbb{E}[m(\mathbf{s}_k, t_k)] = \mu_m$  e  $\text{Var}[m(\mathbf{s}_k, t_k)] = \sigma_m^2$ , para todo  $k = 1, \dots, n$ . Considerando  $\gamma_{ST}(r, t)$  dado na equação (4.8) temos:

$$\begin{aligned} \gamma_{ST}(r, t) &= \frac{1}{2} \mathbb{E} \left[ (m(\mathbf{s}_i, t_i) - m(\mathbf{s}_j, t_j))^2 \right] \\ &= \frac{1}{2} \mathbb{E} \left[ m(\mathbf{s}_i, t_i)^2 + m(\mathbf{s}_j, t_j)^2 - 2m(\mathbf{s}_i, t_i)m(\mathbf{s}_j, t_j) \right] \\ &= \frac{1}{2} \left\{ \mathbb{E}[m(\mathbf{s}_i, t_i)^2] + \mathbb{E}[m(\mathbf{s}_j, t_j)^2] - 2\mathbb{E}[m(\mathbf{s}_i, t_i)]\mathbb{E}[m(\mathbf{s}_j, t_j)] \right\} \\ &= \frac{1}{2} \left\{ \text{Var}[m(\mathbf{s}_i, t_i)] + \mathbb{E}[m(\mathbf{s}_i, t_i)]^2 + \text{Var}[m(\mathbf{s}_j, t_j)] + \mathbb{E}[m(\mathbf{s}_j, t_j)]^2 \right. \\ &\quad \left. - 2\mathbb{E}[m(\mathbf{s}_i, t_i)]\mathbb{E}[m(\mathbf{s}_j, t_j)] \right\} \\ &= \frac{1}{2} \left\{ \sigma_m^2 + \mu_m^2 + \sigma_m^2 + \mu_m^2 - 2\mu_m^2 \right\} \\ &= \sigma_m^2. \end{aligned} \quad (4.10)$$

- Com o aumento de  $r$  e  $t$ , o valor de  $\gamma_{ST}(r, t)$  também aumenta convergindo assintoticamente para um valor constante (patamar), ou seja,

$$\lim_{r, t \rightarrow \infty} \gamma_{ST}(r, t) \equiv \sigma_m^2. \quad (4.11)$$

Pode ser no entanto que o valor de  $\gamma_{ST}(r, t)$  diminua em alguns intervalos de  $r$  e  $t$ . A velocidade da convergência para o patamar está relacionada com a velocidade do declínio da correlação espaço-temporal das marcas com o aumento das distâncias entre os pontos.

Considerando assim, um PPETM estacionário dado em  $(\mathbb{R}^2 \times \mathbb{R}) \times \mathbb{R}$ , será considerado aqui o seguinte estimador para  $\widehat{\gamma}_{ST}$ :

$$\hat{\gamma}_{ST}(r, t) = \frac{\sum_{\mathbf{s}_i, \mathbf{s}_j \in W; t_i, t_j \in T}^{\neq} \frac{1}{2} \frac{(m_i - m_j)^2 \kappa_{\varepsilon}(\|\mathbf{s}_i - \mathbf{s}_j\| - r) \kappa_{\delta}(|t_i - t_j| - t)}{w_{ij}^{(s)} w_{ij}^{(t)}}}{\sum_{\mathbf{s}_i, \mathbf{s}_j \in W; t_i, t_j \in T}^{\neq} \frac{\kappa_{\varepsilon}(\|\mathbf{s}_i - \mathbf{s}_j\| - r) \kappa_{\delta}(|t_i - t_j| - t)}{w_{ij}^{(s)} w_{ij}^{(t)}}} \quad (4.12)$$

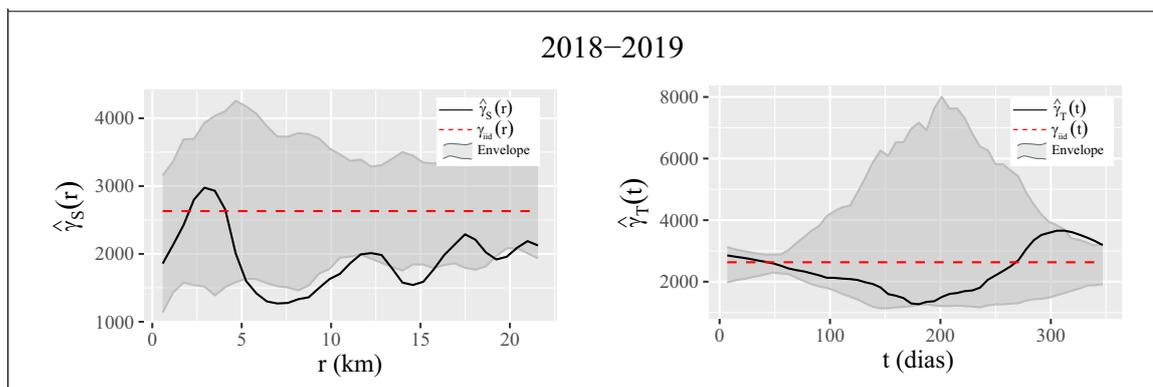
em que  $\sum_{\mathbf{s}_i, \mathbf{s}_j \in W; t_i, t_j \in T}^{\neq}$  indica que a soma é feita sobre todos os pares de pontos  $(\mathbf{s}_i, t_i) \neq (\mathbf{s}_j, t_j)$ , tais que  $\mathbf{s}_i, \mathbf{s}_j \in W \subset \mathbb{R}^2$  e  $t_i, t_j \in T \subset \mathbb{R}$ ;  $\kappa_{\varepsilon}$  e  $\kappa_{\delta}$  são funções kernel com larguras de banda  $\varepsilon$  e  $\delta$ , respectivamente e  $w_{ij}^{(s)}$  e  $w_{ij}^{(t)}$  são os fatores de correção do efeito de bordas isotrópicos espacial e temporal, respectivamente.

Após o cálculo de  $\hat{\gamma}_{ST}(r, t)$ , para verificar desvios em relação ao modelo nulo de rotulagem aleatória realiza-se de modo análogo ao caso da análise marginal, a construção dos envelopes de simulação. Como nesse caso graficamente os cálculos de  $\hat{\gamma}_{ST}(r, t)$  retornam superfícies e não curvas, para facilitar a visualização, os resultados serão apresentados num plano, de modo que as regiões em que  $\hat{\gamma}_{ST}(s, t)$  estiver dentro do envelope serão representadas pela cor branca, quando estiver acima do envelope a cor será vermelha, e quando  $\hat{\gamma}_{ST}(s, t)$  estiver abaixo do envelope, a cor utilizada será o azul.

### 4.2.3 Resultado da análise marginal

Considerando a análise marginal do processo pontual espaço-temporal marcado, os variogramas  $\hat{\gamma}_S(r)$  e  $\hat{\gamma}_T(t)$ , bem como os envelopes obtidos a partir de 199 permutações sob o modelo nulo de rotulagem aleatória, para os anos de 2018 a 2019 estão apresentados na Figura 4.8.

Figura 4.8 – Variogramas marginais  $\hat{\gamma}_S(r)$  e  $\hat{\gamma}_T(t)$  estimados, e os envelopes obtidos a partir de 199 permutações sob o modelo nulo de rotulagem aleatória, para os anos de 2018 a 2019.



Fonte: Do autor (2020)

Observando os resultados de ambos variogramas, pode-se verificar que para pequenas distâncias no espaço e no tempo,  $\hat{\gamma}_S(r)$  e  $\hat{\gamma}_T(t)$  não saem do envelope. Isso indica que os valores de FRP não estão correlacionados para focos de calor observados próximos no espaço e no tempo, isto é, o valor da FRP em um foco independe de um foco vizinho.

Para  $\hat{\gamma}_S(r)$  pode-se notar também que para distâncias entre 5 km e 12 km, aproximadamente, a variabilidade nos valores de FRP é mais baixa do que o esperado para o caso da rotulagem aleatória, indicando que existe uma maior similaridade entre eles. Para  $\hat{\gamma}_T(t)$ , observa-se que o variograma ficou acima do envelope, indicando uma variabilidade maior do que o esperado para o caso da rotulagem aleatória, a partir de 310 dias aproximadamente. Isso indica que, de um ano para outro, os valores de FRP tendem a ser mais diferentes entre si. Esse resultado vai ao encontro do que foi mostrado na Figura 4.6, indicando que regiões queimadas em um ano, não foram no outro.

Essa análise marginal permite verificar, para cada uma das componentes espacial e temporal, qual é a estrutura de variação das marcas ao longo do tempo e do espaço de modo que, para alguns distâncias de interesse (como para pontos próximos no espaço/tempo ou de um ano para outro), pode-se observar a existência ou não de uma estrutura de correlação entre os focos de calor. Contudo, ao observar apenas uma das componentes espacial ou temporal, o cálculo do variograma condensa a outra componente, perdendo informação no que se refere a diferentes escalas de distâncias. Assim, pode acontecer que algumas regiões em que observa-se uma variabilidade diferente da esperada para um caso completamente aleatório, ao olhar para o variograma marginal ela pode não aparecer.

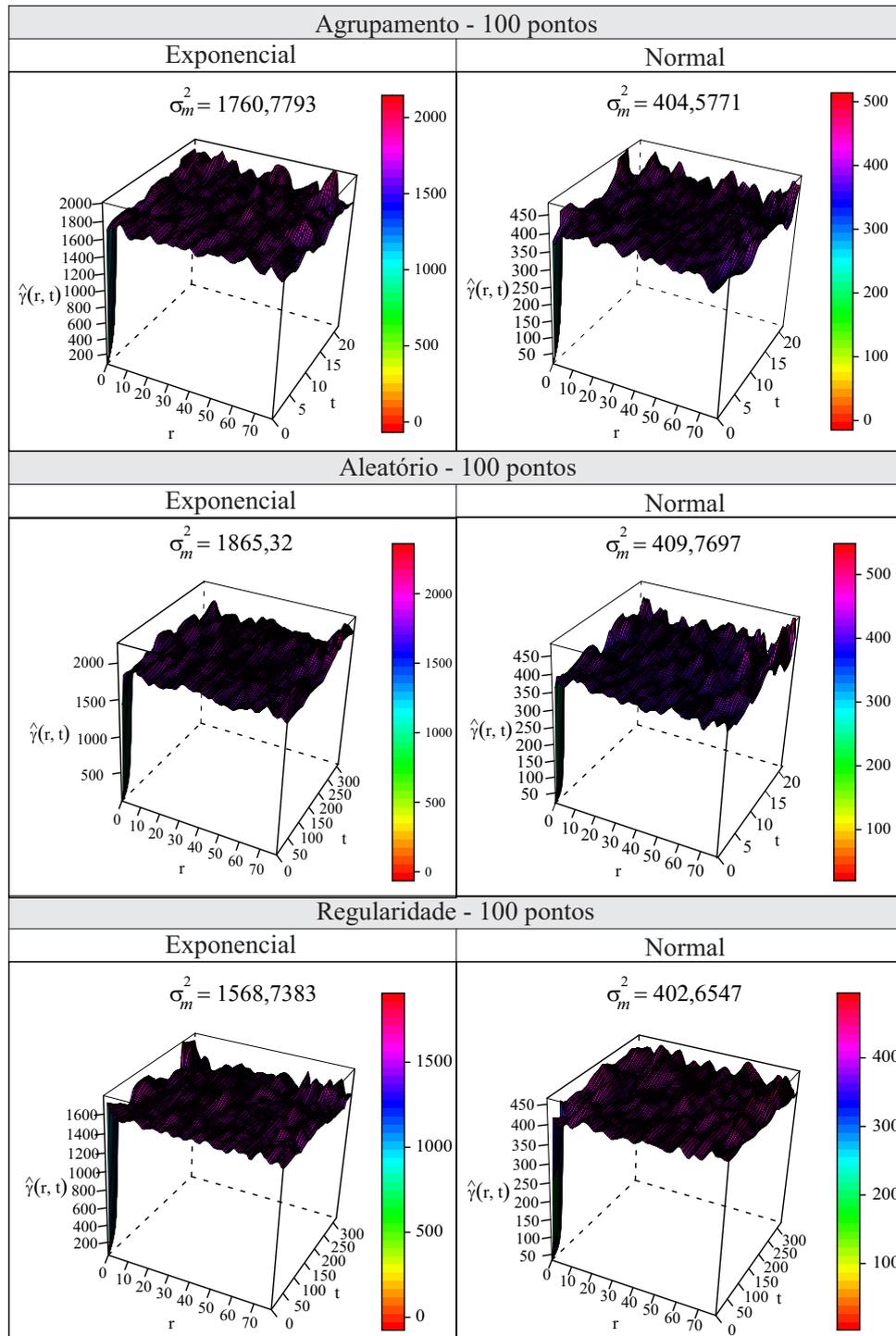
No tópico a seguir, está apresentada a análise conjunta do processo pontual marcado. Veremos que regiões onde os variogramas marginais apontam para a aleatoriedade, no caso espaço-tempo não são.

#### 4.2.4 Resultado da análise conjunta

Considerando os três padrões pontuais simulados de agrupamento, aleatoriedade e regularidade, pode-se ver pelas Figuras 4.9 e 4.10 que em todos os casos, o valor do variograma próximo à origem ( $\hat{\gamma}_{ST}(0+,0+)$ ) é próximo de zero, depois aumenta rapidamente e passa a flutuar em torno do valor da variância das marcas  $\sigma_m^2$ , para  $r,t > 0$ .

Para o caso do processo de Poisson não homogêneo, pode-se ver na Figura 4.11, que o variograma médio apresentou resultado similar aos casos anteriores. Apesar do valor do efeito

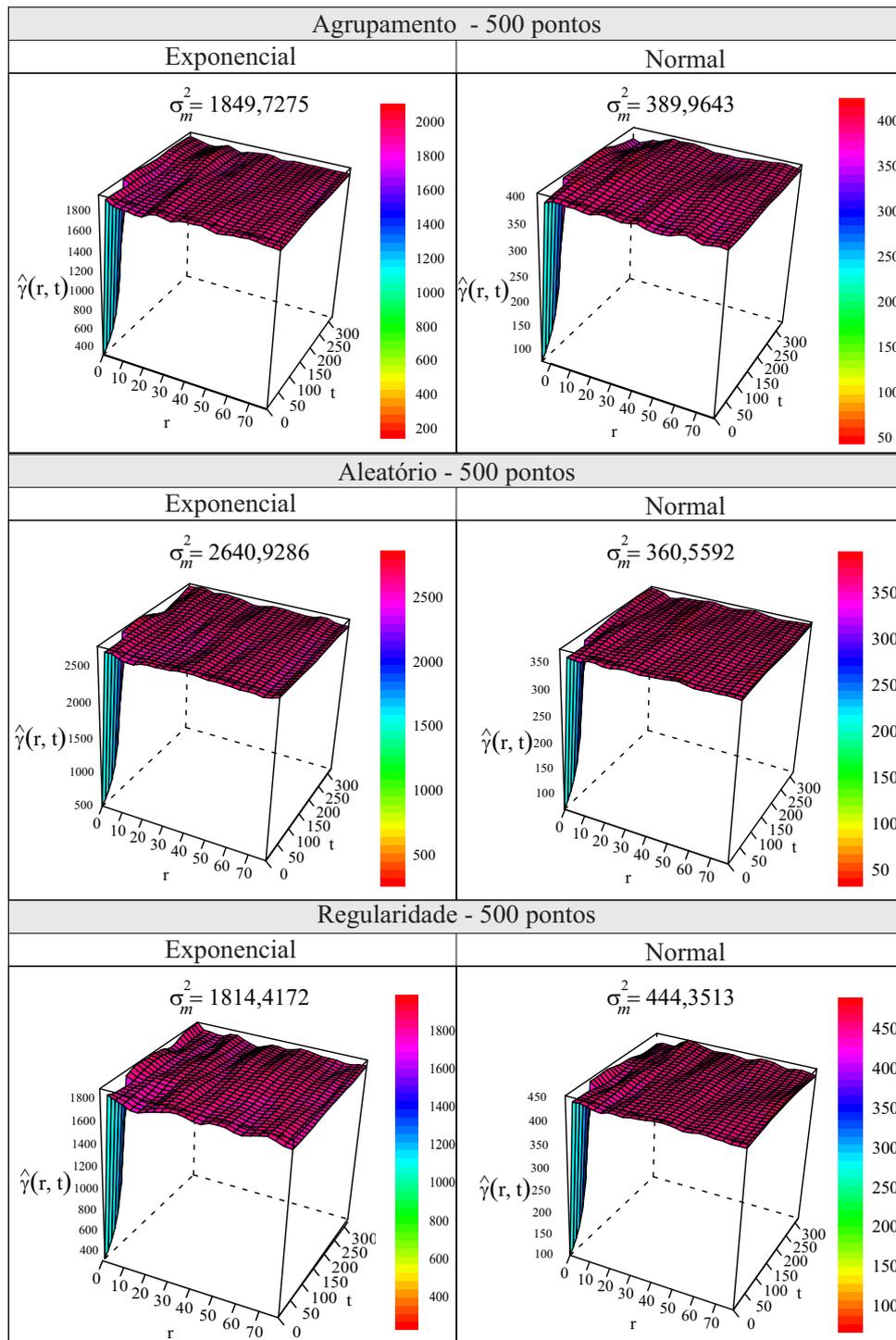
Figura 4.9 – Variogramas médios obtidos para o modelo nulo de rotulagem aleatória para 100 permutações das marcas em diferentes estruturas espaço-temporais.



Legenda: Esquerda: marcas provenientes de 100 observações de uma distribuição exponencial com taxa  $\lambda = 1/50$ . Direita: marcas provenientes de 100 observações de uma distribuição normal com média  $\mu = 50$  e desvio padrão  $\sigma = 20$ .

Fonte: Do autor (2020)

Figura 4.10 – Variogramas médios obtidos para o modelo nulo de rotulagem aleatória para 100 permutações das marcas em diferentes estruturas espaço-temporais.

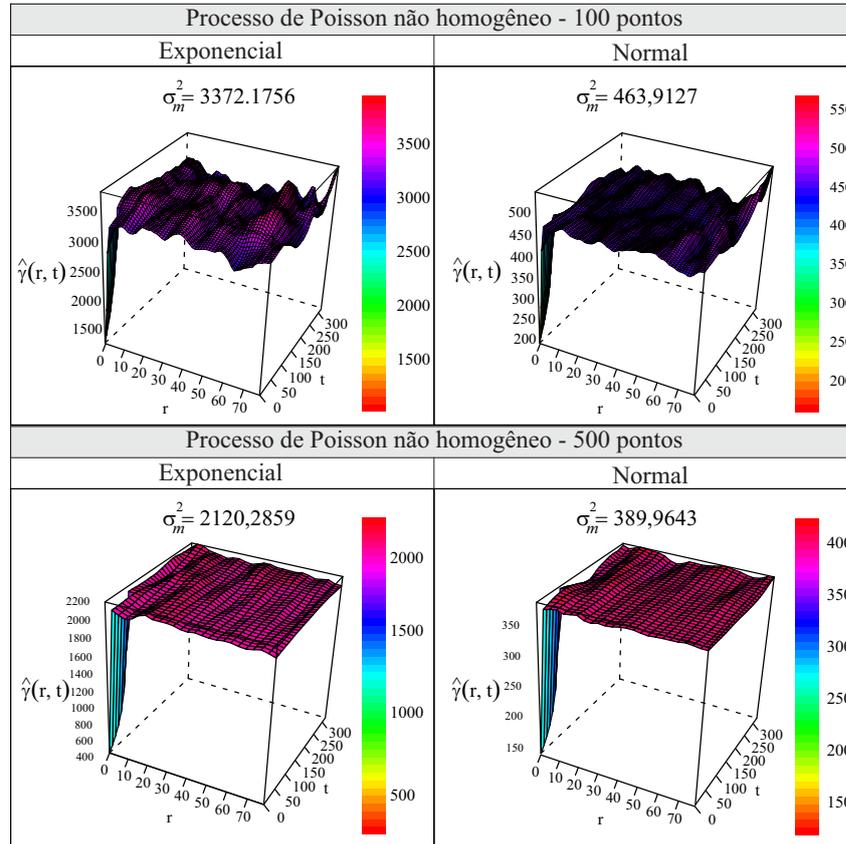


Legenda: Esquerda: marcas provenientes de 500 observações de uma distribuição exponencial com taxa  $\lambda = 1/50$ . Direita: marcas provenientes de 500 observações de uma distribuição normal com média  $\mu = 50$  e desvio padrão  $\sigma = 20$ .

Fonte: Do autor (2020)

pepita não ter ficado próximo de zero, a forma dos variogramas foi a mesma: rápido crescimento a partir do efeito pepita e flutuação em torno do valor da variância das marcas para  $r, t > 0$ .

Figura 4.11 – Variogramas médios obtidos para o modelo nulo de rotulagem aleatória para 100 permutações das marcas em um PPNH espaço-temporal.



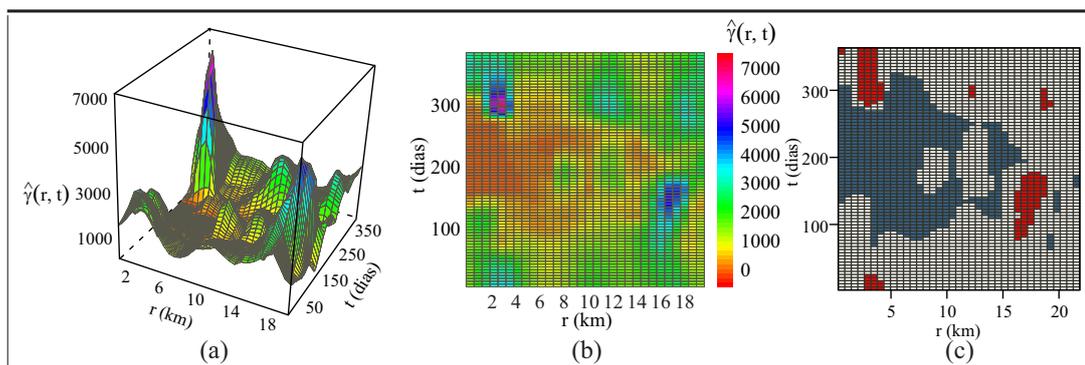
Legenda: Esquerda: marcas provenientes de 100 observações de uma distribuição exponencial com taxa  $\lambda = 1/50$ . Direita: marcas provenientes de 100 observações de uma distribuição normal com média  $\mu = 50$  e desvio padrão  $\sigma = 20$ .

Fonte: Do autor (2020)

Esses resultados empíricos mostram a compatibilidade entre os cálculos de  $\hat{\gamma}_{ST}(s, t)$  e o resultado teórico da equação 4.9.

Considerando os dados dos focos de calor e da FRP, os resultados da análise estão na Figura 4.12.

Figura 4.12 – Resultado de  $\hat{\gamma}_{ST}(r,t)$  aplicado aos dados de focos de calor



Legenda: Variograma  $\hat{\gamma}_{ST}(r,t)$  estimado para os anos de 2018 a 2019 ((a) e (b)) e (c) as regiões em que  $\hat{\gamma}_{ST}(r,t)$  fica fora do envelope (vermelho e azul) construído a partir de 199 simulações sob o modelo nulo de rotulagem aleatória.

Fonte: Do autor (2020)

Na Figura 4.12, os gráficos (a) e (b) possuem essencialmente a mesma informação, sendo o primeiro, o variograma apresentado como superfície tridimensional e o segundo, como superfície de níveis. O gráfico (c) é um plano que representa as regiões em que  $\hat{\gamma}_{ST}(r,t)$  ficou fora do envelope construído a partir de 199 simulações sob o modelo nulo de rotulagem aleatória e uma taxa de erro  $\alpha = 0,05$ . A cor azul indica que  $\hat{\gamma}_{ST}(r,t)$  ficou abaixo do envelope e a cor vermelha indica que ficou acima. As regiões em que  $\hat{\gamma}_{ST}(r,t)$  permaneceu dentro do envelope estão na cor branca.

Pode-se notar que para focos separados entre 3 km e 5 km, a variabilidade nos valores da FRP é mais alta do que o esperado pela rotulagem aleatória para até aproximadamente 20 dias, bem como para tempos maiores do que 280 dias. Isso indica que para focos próximos no espaço, mas ocorrendo de um ano para outro, os valores de FRP tendem a ser diferentes. Isso confirma os resultados apresentados anteriormente nas Figuras 4.6 e 4.8. Além disso, observa-se que para curtas distâncias no espaço e para períodos de tempo a partir de 150 dias, o variograma fica abaixo do envelope, indicando uma variabilidade muito baixa nos valores das marcas. Esse comportamento pode ser devido ao fato de que o período entre as estações mais secas apresenta um baixo número de focos de calor e, assim, não há muitos pontos para o cálculo de  $\hat{\gamma}_{ST}(r,t)$ . A partir desses resultados, percebe-se que o variograma também apontou o padrão de sazonalidade no comportamento da intensidade dos valores de FRP.

Por fim, temos que para pequenos intervalos de tempo, somente se observa um desvio da aleatoriedade para pequenas distâncias espaciais, não havendo assim correlação entre os valores

de FRP de pontos próximos no tempo, mas afastados no espaço. Isso indica que a quantidade de biomassa queimada em uma região, independe do quanto foi queimado no mesmo dia ou em dias próximos, de um local longe dali.

Observa-se que o uso do variograma  $\gamma_{ST}(r,t)$  apresentou intervalos no tempo e no espaço, que ficaram fora do envelope construído sob a hipótese de rotulagem aleatória, que no caso puramente espacial ( $\gamma_p(r)$ ) e temporal ( $\gamma_t(t)$ ) não ocorreu. Isto é,  $\hat{\gamma}_{ST}(r,t)$  foi capaz de identificar uma estrutura de correlação para a variabilidade das marcas em várias escalas espaço-temporais, o que não ocorreu ao considerar os processos separadamente.

De modo geral, associando o valor da FRP à quantidade de biomassa queimada, os resultados mostram que focos próximos no tempo e no espaço tendem a apresentar valores muito distintos entre si, de modo que regiões onde houve uma grande quantidade de biomassa queimada, os focos no seu entorno apresentaram valores mais baixos. Cada foco de calor é representado como um ponto, mas sabemos que a extensão do fogo na vegetação pode atingir grandes distâncias. Assim, esse comportamento de alta variabilidade em pontos próximos, pode-se justificar no fato de que se um incêndio/queimada atingiu grandes proporções, então consumiu grande quantidade de biomassa, e as ocorrências no seu entorno tiveram baixas taxas de queima, por não haver mais uma grande quantidade de biomassa para ser queimada.

Apesar das informações dos dados de focos de calor, juntamente com a FRP apresentarem imprecisões oriundas das limitações dos sensores dos satélites, o seu uso pode mostrar tendências espaço-temporais importantes para a compreensão do comportamento dos incêndios e queimadas. Verificou-se nesse estudo, por exemplo, que regiões não queimadas em um ano tendem a queimar no ano seguinte em função do acúmulo de biomassa. Esse resultado contribui, por exemplo, para uma melhor compreensão do comportamento do fenômeno, bem como na tomada de decisão sobre estratégias de enfrentamento e alocação de recursos.

É importante destacar também que os resultados obtidos referem-se à uma análise retrospectiva, de modo que o padrão observado para os dados reais não necessariamente irá se repetir para tempos futuros, ou que irá ocorrer do mesmo modo para outras regiões que possam ser tomadas para o estudo.

## 5 CONCLUSÕES

A análise de um PPET utilizando os variogramas  $\gamma_{sp}(r)$  e  $\gamma_e(t)$ , acrescida do uso dos envelopes de simulação complementa a análise proposta por Stoyan et al. (2017), uma vez que quando há desvios do variograma em relação ao valor esperado pela a completa aleatoriedade, o seu uso possibilita verificar se o padrão observado realmente exhibe dependência entre as ocorrências, ou se há apenas um efeito de tendência no processo.

O variograma marcado espaço-temporal  $\gamma_{ST}(r,t)$  proposto nesta tese possibilita ganhos em relação a uma análise puramente espacial ou temporal, já que é possível identificar uma estrutura de correlação para a variabilidade das marcas em várias escalas espaço-temporais, o que pode não ocorrer ao considerar os processos separadamente.

Nesta tese foi verificada a estrutura que o variograma marcado espaço-temporal apresenta sob o modelo nulo de rotulagem aleatória. Um trabalho futuro que pode ser realizado a partir do que foi apresentado, inclui verificar a forma que o variograma espaço-temporal assume sob modelos que apresentem uma estrutura, em que as marcas são geradas de modo dependente ao processo pontual de base. O desafio relacionado a isso se deve tanto ao aspecto teórico, como de implementação computacional.

## REFERÊNCIAS

- ARAGÓ, P. et al. Spatial point process modeling applied to the assessment of risk factors associated with forest wildfires incidence in castellón, spain. **European journal of forest research**, Springer, v. 135, n. 3, p. 451–464, 2016.
- ASSUNÇÃO, R.; MAIA, A. A note on testing separability in spatial-temporal marked point processes. **Biometrics**, Wiley Online Library, v. 63, n. 1, p. 290–294, 2007.
- AUGUIE, B. **gridExtra: Miscellaneous Functions for "Grid"Graphics**. [S.l.], 2017. R package version 2.3. Disponível em: <<https://CRAN.R-project.org/package=gridExtra>>.
- BADDELEY, A.; BÁRÁNY, I.; SCHNEIDER, R. Spatial point processes and their applications. **Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004**, Springer, p. 1–75, 2007.
- BADDELEY, A.; RUBAK, E.; TURNER, R. **Spatial point patterns: methodology and applications with R**. [S.l.]: Chapman and Hall/CRC, 2015.
- BADDELEY, A.; TURNER, R. spatstat: An R package for analyzing spatial point patterns. **Journal of Statistical Software**, v. 12, n. 6, p. 1–42, 2005. Disponível em: <<http://www.jstatsoft.org/v12/i06/>>.
- BADDELEY, A. J.; MØLLER, J.; WAAGEPETERSEN, R. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. **Statistica Neerlandica**, Wiley Online Library, v. 54, n. 3, p. 329–350, 2000.
- BARBOSA, R. I. **Savanas da Amazônia: emissão de gases do efeito estufa e material particulado pela queima e decomposição da biomassa acima do solo, sem a troca do uso da terra, em Roraima, Brasil**. Tese (Doutorado) — Instituto Nacional de Pesquisas da Amazônia, 2001.
- BEISBART, C.; KERSCHER, M. Luminosity-and morphology-dependent clustering of galaxies. **The Astrophysical Journal**, IOP Publishing, v. 545, n. 1, p. 6, 2000.
- BIVAND, R.; KEITT, T.; ROWLINGSON, B. **rgdal: Bindings for the 'Geospatial' Data Abstraction Library**. [S.l.], 2020. R package version 1.5-16. Disponível em: <<https://CRAN.R-project.org/package=rgdal>>.
- BIVAND, R.; LEWIN-KOH, N. **maptools: Tools for Handling Spatial Objects**. [S.l.], 2020. R package version 1.0-2. Disponível em: <<https://CRAN.R-project.org/package=maptools>>.
- BIVAND, R. S. et al. **Applied spatial data analysis with R**. [S.l.]: Springer, 2008. v. 747248717.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Serviço Florestal Brasileiro. **Floresta Nacional do Jamari (RO)**. Brasília, DF, 2020. Disponível em: <

- CÂMARA, G.; CARVALHO, M. S. In: DRUCK, S. et al. (Ed.). **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2003. cap. Análise espacial de eventos. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/>>.
- CAMARGO, E. C. G.; FUCKS, S. D.; CÂMARA, G. Análise espacial de superfícies. **Análise espacial de dados geográficos. Planaltina: Embrapa Cerrados**, p. 79–122, 2004.
- CHIU, S. N. et al. **Stochastic geometry and its applications**. [S.l.]: John Wiley & Sons, 2013.
- CLIMATE-DATA. **Itapuã do Oeste Clima**. 2020. Disponível em: <<https://pt.climate-data.org/america-do-sul/brasil/rondonia/itapua-do-oeste-312427/>>.
- COMAS, C.; MATEU, J. Modelling forest dynamics: a perspective from point process methods. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 49, n. 2, p. 176–196, 2007.
- CORRÊA, S. d. C. **A influência dos sistemas climáticos sobre os incêndios florestais: estudo de caso: evento de incêndio ocorrido em setembro de 2005 no Jardim Botânico de Brasília**. Tese (Doutorado) — Universidade de Brasília, 2007.
- COSTA, B. S. C. d.; FONSECA, E. L. d. The use of fire radiative power to estimate the biomass consumption coefficient for temperate grasslands in the atlantic forest biome. **Revista Brasileira de Meteorologia**, SciELO Brasil, v. 32, n. 2, p. 255–260, 2017.
- COSTA, B. S. C. da. **Uso da energia radiativa do fogo para estimativa do coeficiente de consumo de biomassa por queima de vegetação campestre**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, 2012.
- COX, D. R.; ISHAM, V. **Point processes**. [S.l.]: Routledge, 2018.
- CRESSIE, N. **Statistics for Spatial Data (Wiley Series in Probability and Statistics)**. Revised edition. [S.l.]: Wiley-Interscience, 1993.
- CRONIE, O.; LIESHOUT, M. N. M. van. Summary statistics for inhomogeneous marked point processes. **Annals of the Institute of Statistical Mathematics**, Springer, v. 68, n. 4, p. 905–928, 2016.
- CRONIE, O.; SÄRKKÄ, A. Some edge correction methods for marked spatio-temporal point process models. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 7, p. 2209–2220, 2011.
- DALEY, D. J.; VERE-JONES, D. **An introduction to the theory of point processes**. 2nd ed. ed. [S.l.]: Springer, 2002. Volume 1.
- DALEY, D. J.; VERE-JONES, D. **An introduction to the theory of point processes: volume I: general theory and structure**. [S.l.]: Springer Science & Business Media, 2003.
- DIGGLE, P. A kernel method for smoothing point process data. **Applied statistics**, JSTOR, p. 138–147, 1985.
- DIGGLE, P. J. **Statistical analysis of spatial and spatio-temporal point patterns**. [S.l.]: Chapman and Hall/CRC, 2013.

DIGGLE, P. J. et al. Second-order analysis of space-time clustering. **Statistical methods in medical research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 4, n. 2, p. 124–136, 1995.

DW Brasil. **Amazônia e Pantanal têm recorde de queimadas em outubro**. 2020. Disponível em: <<https://p.dw.com/p/3kjMv>>.

FREEBORN, P. H. et al. Relationships between energy release, fuel mass loss, and trace gas and aerosol emissions during laboratory biomass fires. **Journal of Geophysical Research: Atmospheres**, Wiley Online Library, v. 113, n. D1, 2008.

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. et al. stpp: an r package for plotting, simulating and analyzing spatio-temporal point patterns. **Journal of Statistical Software**, v. 53, n. 2, p. 1–29, 2013.

GASPAR, L. Queimadas no brasil. **Pesquisa Escolar Online**, Fundação Joaquim Nabuco, 2012. Disponível em: <<http://basilio.fundaj.gov.br/pesquisaescolar>>.

GERVINI, D. Independent component models for replicated point processes. **Spatial Statistics**, Elsevier, v. 18, p. 474–488, 2016.

GROLEMUND, G.; WICKHAM, H. Dates and times made easy with lubridate. **Journal of Statistical Software**, v. 40, n. 3, p. 1–25, 2011. Disponível em: <<http://www.jstatsoft.org/v40/i03/>>.

GUIDOUM, A. **kedd: Kernel estimator and bandwidth selection for density and its derivatives**. [S.l.], 2015. R package version 1.0.3. Disponível em: <<http://CRAN.R-project.org/package=kedd>>.

HEAL, A. et al. Revealed fires three times more common in amazon beef farming zones. **The Guardian**, v. 10, 2019.

HOLDEN, L.; SANNAN, S.; BUNGUM, H. A stochastic marked point process model for earthquakes. **Natural Hazards and Earth System Science**, v. 3, n. 1/2, p. 95–101, 2003.

ICHOKU, C.; KAUFMAN, Y. J. A method to derive smoke emission rates from modis fire radiative energy measurements. **IEEE transactions on Geoscience and Remote Sensing**, IEEE, v. 43, n. 11, p. 2636–2649, 2005.

ICMBIO. **Mais de 90% dos incêndios têm ação humana, diz PrevFogo**. Instituto Chico Mendes de Conservação da Biodiversidade, Brasília - DF, 2016. Disponível em: <<https://www.icmbio.gov.br/portal/ultimas-noticias/20-geral/8327-mais-de-90-dos-incendios-tem-acao-humana-diz-prevfogo>>.

IFTIMI, A.; CRONIE, O.; MONTES, F. The second-order analysis of marked spatio-temporal point processes, with an application to earthquake data. **arXiv preprint arXiv:1611.04808**, 2016.

ILLIAN, J. et al. **Statistical analysis and modelling of spatial point patterns**. [S.l.]: John Wiley & Sons, 2008. v. 70.

INPE. **Programa Queimadas**. 2020. Disponível em: <<http://queimadas.dgi.inpe.br/queimadas/bdqueimadas>>.

KONOVALOV, I. et al. Constraining CO<sub>2</sub> emissions from open biomass burning by satellite observations of co-emitted species: a method and its application to wildfires in Siberia. **Atmospheric Chemistry and Physics**, European Geosciences Union, v. 14, p. 10383–10410, 2014.

LAWSON, A.; ZHOU, H. Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. **Preventive Veterinary Medicine**, Elsevier, v. 71, n. 3-4, p. 141–156, 2005.

LI, F. et al. Investigation of the fire radiative energy biomass combustion coefficient: A comparison of polar and geostationary satellite retrievals over the conterminous United States. **Journal of Geophysical Research: Biogeosciences**, Wiley Online Library, v. 123, n. 2, p. 722–739, 2018.

LLOYD, C. D. **Local models for spatial analysis**. [S.l.]: CRC Press, 2010.

MATAVELI, G. A. V. et al. Relação entre o ângulo de visada e a estimativa da potência radiativa do fogo. **Boletim de Ciências Geodésicas**, SciELO Brasil, v. 21, n. 2, p. 371–388, 2015.

MOHLER, G. O. et al. Self-exciting point process modeling of crime. **Journal of the American Statistical Association**, Taylor & Francis, v. 106, n. 493, p. 100–108, 2011.

MØLLER, J.; DÍAZ-AVALOS, C. Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 37, n. 1, p. 2–25, 2010.

MONTERO, J.-M.; FERNÁNDEZ-AVILÉS, G.; MATEU, J. **Spatial and spatio-temporal geostatistical modeling and kriging**. [S.l.]: John Wiley & Sons, 2015.

NADARAYA, E. A. On estimating regression. **Theory of Probability & Its Applications**, SIAM, v. 9, n. 1, p. 141–142, 1964.

NADARAYA, E. A. **Nonparametric Estimation of Probability Densities and Regression Curves**. 1. ed. [S.l.]: Springer Netherlands, 1989. (Mathematics and its Applications (Soviet Series) 20).

NEPSTAD, D. et al. Inhibition of Amazon deforestation and fire by parks and indigenous lands. **Conservation Biology**, Wiley Online Library, v. 20, n. 1, p. 65–73, 2006.

NEUWIRTH, E. **RColorBrewer: ColorBrewer Palettes**. [S.l.], 2014. R package version 1.1-2. Disponível em: <<https://CRAN.R-project.org/package=RColorBrewer>>.

OGATA, Y. Space-time point-process models for earthquake occurrences. **Annals of the Institute of Statistical Mathematics**, Springer, v. 50, n. 2, p. 379–402, 1998.

PAWLAS, Z. Empirical distributions in marked point processes. **Stochastic Processes and their Applications**, Elsevier, v. 119, n. 12, p. 4194–4209, 2009.

PEBESMA, E.; BIVAND, R. S. S classes and methods for spatial data: the sp package. **R news**, v. 5, n. 2, p. 9–13, 2005.

- PENTTINEN, A.; STOYAN, D.; HENTTONEN, H. M. Marked point processes in forest statistics. **Forest science**, Oxford University Press, v. 38, n. 4, p. 806–824, 1992.
- PEREIRA, G. et al. Determinação e modelagem da taxa de consumo de biomassa queimada. **Revista Brasileira de Meteorologia**, v. 27, n. 1, p. 13–22, 2012.
- PEREZ, A. M.; WARD, M. P.; CARPENTER, T. E. Control of a foot-and-mouth disease epidemic in argentina. **Preventive veterinary medicine**, Elsevier, v. 65, n. 3-4, p. 217–226, 2004.
- PODUR, J.; MARTELL, D. L.; CSILLAG, F. Spatial patterns of lightning-caused forest fires in ontario, 1976–1998. **Ecological Modelling**, Elsevier, v. 164, n. 1, p. 1–20, 2003.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.
- RAMOS, R. M.; FONSECA, R. L.; MORELLO, T. F. Unidades de conservação e proteção contra incêndios florestais: relação entre focos de calor e ações articuladas pelas brigadas contratadas. **Biodiversidade Brasileira**, v. 6, n. 2, p. 135–148, 2016.
- RIBEIRO Jr, P. J. et al. **geoR: Analysis of Geostatistical Data**. [S.l.], 2020. R package version 1.8-1. Disponível em: <<https://CRAN.R-project.org/package=geoR>>.
- RIPLEY, B. D. The second-order analysis of stationary point processes. **Journal of applied probability**, Cambridge University Press, v. 13, n. 2, p. 255–266, 1976.
- ROBERTS, G. et al. Retrieval of biomass combustion rates and totals from fire radiative power observations: Application to southern africa using geostationary seviri imagery. **Journal of Geophysical Research: Atmospheres**, Wiley Online Library, v. 110, n. D21, 2005.
- SARKAR, D. **Lattice: Multivariate Data Visualization with R**. New York: Springer, 2008. ISBN 978-0-387-75968-5. Disponível em: <<http://lmdvr.r-forge.r-project.org>>.
- SÄRKKÄ, A.; RENSHAW, E. The analysis of marked point patterns evolving through space and time. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 3, p. 1698–1718, 2006.
- SCHABENBERGER, O.; GOTWAY, C. A. Statistical methods for spatial data analysis. Chapman & Hall/CRC, 2005.
- SCHOENBERG, F. P. Multidimensional residual analysis of point process models for earthquake occurrences. **Journal of the American Statistical Association**, Taylor & Francis, v. 98, n. 464, p. 789–795, 2003.
- SCHOENBERG, F. P. Testing separability in spatial-temporal marked point processes. **Biometrics**, Wiley Online Library, v. 60, n. 2, p. 471–481, 2004.
- SCHROEDER, W. et al. Integrated active fire retrievals and biomass burning emissions using complementary near-coincident ground, airborne and spaceborne sensor data. **Remote Sensing of Environment**, Elsevier, v. 140, p. 719–730, 2014.
- SNYDER, D. L.; MILLER, M. I. **Random Point Processes in Time and Space**. 2. ed. [S.l.]: Springer-Verlag New York, 1991. (Springer Texts in Electrical Engineering). ISBN 978-1-4612-7821-4, 978-1-4612-3166-0.

- SOETAERT, K. **plot3D: Plotting Multi-Dimensional Data**. [S.l.], 2019. R package version 1.3. Disponível em: <<https://CRAN.R-project.org/package=plot3D>>.
- STOYAN, D.; PENTTINEN, A. Recent applications of point process methods in forestry statistics. **Statistical Science**, JSTOR, p. 61–78, 2000.
- STOYAN, D. et al. Mark variograms for spatio-temporal point processes. **Spatial statistics**, Elsevier, v. 20, p. 125–147, 2017.
- STOYAN, D.; WÄLDER, O. On variograms in point process statistics, ii: models of markings and ecological interpretation. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 42, n. 2, p. 171–187, 2000.
- TURNER, R. Point patterns of forest fire locations. **Environmental and ecological statistics**, Springer, v. 16, n. 2, p. 197–223, 2009.
- VERE-JONES, D.; DENG, Y. A point process analysis of historical earthquakes from north china. **Earthquake Research in China**, v. 2, n. 2, p. 165–181, 1988.
- WÄLDER, K.; WÄLDER, O. Analysing interaction effects in forests using the mark correlation function. **iForest-Biogeosciences and Forestry**, SISEF-Italian Society of Silviculture and Forest Ecology, v. 1, n. 1, p. 34, 2008.
- WÄLDER, O.; STOYAN, D. On variograms in point process statistics. **Biometrical Journal**, Wiley Online Library, v. 38, n. 8, p. 895–905, 1996.
- WANG, J. et al. Mitigating satellite-based fire sampling limitations in deriving biomass burning emission rates: Application to wrf-chem model over the northern sub-saharan african region. **Journal of Geophysical Research: Atmospheres**, Wiley Online Library, v. 123, n. 1, p. 507–528, 2018.
- WARNES, G. R. et al. **gdata: Various R Programming Tools for Data Manipulation**. [S.l.], 2017. R package version 2.18.0. Disponível em: <<https://CRAN.R-project.org/package=gdata>>.
- WATSON, G. S. Smooth regression analysis. **Sankhyā: The Indian Journal of Statistics, Series A**, JSTOR, p. 359–372, 1964.
- WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<https://ggplot2.tidyverse.org>>.
- WIEGAND, T.; MOLONEY, K. A. **Handbook of spatial point-pattern analysis in ecology**. [S.l.]: CRC press, 2013.
- WOOSTER, M.; ZHANG, Y. Boreal forest fires burn less intensely in russia than in north america. **Geophysical Research Letters**, Wiley Online Library, v. 31, n. 20, 2004.
- WOOSTER, M. J. Small-scale experimental testing of fire radiative energy for quantifying mass combusted in natural vegetation fires. **Geophysical Research Letters**, Wiley Online Library, v. 29, n. 21, p. 23–1, 2002.

WOOSTER, M. J. et al. Retrieval of biomass combustion rates and totals from fire radiative power observations: Frp derivation and calibration relationships between biomass consumption and fire radiative energy release. **Journal of Geophysical Research: Atmospheres**, Wiley Online Library, v. 110, n. D24, 2005.

ZHANG, T.; ZHUANG, Q. On the local odds ratio between points and marks in marked point processes. **Spatial Statistics**, Elsevier, v. 9, p. 20–37, 2014.

## APÊNDICE A – Implementação em R do variograma marcado espaço-temporal $\hat{\gamma}_{ST}(r,t)$

```

varioST<-function(x,y,t,m,jan,ps=0.25,pt=0.25){
#-----#

# x: vetor de coordenadas x
# y: vetor de coordenadas y
# t: vetor de tempos t
# jan: janela de observacao espacial (objeto owin do spatstat)
# m: vetor das marcas associadas a cada ponto (x,y,t)
# ps: proporcao da distancia maxima no espaço para o calculo do variograma
# pt: proporcao da distancia maxima no tempo para o calculo do variograma

#-----#
# Carregando os pacotes necessarios
#-----#

library(kedd)
library(gdata)
library(KernSmooth)
library(spatstat)
library(stpp)
library(lattice)
library(gridExtra)
library(RColorBrewer)

#-----#
# Construindo a matriz de distancias espaciais entre os pontos (x,y)
#-----#

N=length(t)
Distance.S<-matrix(0,N,N)

for (i in 1:N){
  for (j in 1:N){
    Distance.S[i,j]=sqrt((x[i]-x[j])^2+(y[i]-y[j])^2)
  }
}

#-----#
# Construindo a matriz de distancias entre os tempos t
#-----#

Distance.T<-matrix(0,N,N)

for (i in 1:N){
  for (j in 1:N){
    Distance.T[i,j]=abs(t[i]-t[j])
  }
}

```

```

#-----#
# Calculando 1/2*(mi-mj)^2 para todos os pares de pontos
#-----#

HalfSquareDiff<-matrix(0,N,N)

for (i in 1:N){
  for (j in 1:N){
    HalfSquareDiff[i,j]=0.5*(m[i]-m[j])^2
  }
}

#-----#
# Obter a correcao do efeito de bordas espacial
#-----#

borda<-edge.Ripley(ppp(x,y, window = jan), r=Distance.S)
bord.S<-c(borda)

#-----#
# Obter a correcao do efeito de bordas temporal
#-----#

borda.corr<-matrix(0,N,N)

for (i in 1:N) {
  for (j in 1:N) {
    if ((t[i]-abs(t[i]-t[j]))>=range(t)[1])&
        t[i]+abs(t[i]-t[j])<=range(t)[2]){
      borda.corr[i,j]<-1
    } else borda.corr[i,j]<-0.5
  }
}

bord.T<-c(borda.corr)

#-----#
# Determinar a largura de banda espacial
#-----#

h <- mse2d(as.points(x,y), cbind(rev(jan$bdry[[1]]$x),rev(jan$bdry[[1]]$y)),
  nsmse = 1000, range = 100)
hs <- h$h[which.min(h$mse)]

#-----#
# Determinar a largura de banda temporal
#-----#

dst.T<-c(Distance.T)
ht<-density(t,kernel = "epanechnikov")$bw

```

```

#-----#
# Distancias espaciais e temporais para o variograma ser calculado
#-----#

max.xy<-max(jan$ xrange[2], jan$ yrange[2])
max.t<-range(t)[2]-range(t)[1]

d.S<-seq(0, max.xy*ps, by=max.xy/150)
d.T<-seq(0, max.t*pt, by=max.t/150)

#-----#
# Calculo do variograma
#-----#

gama<-matrix(0, length(d.S)-1, length(d.T)-1)
mc<-c(HalfSquareDiff)
dst.S<-c(Distance.S)

for (k in 1:length(d.S)-1) {
  for (p in 1:length(d.T)-1) {
    gama[k,p]<-sum(
      (mc*(kernel.fun(x = (dst.S-d.S[k+1])/hs,
        kernel = "epanechnikov")$kx)*
        (kernel.fun(x = (dst.T-d.T[p+1])/ht,
        kernel = "epanechnikov")$kx))/(bord.S[k]*bord.T[p]))/
      (sum((kernel.fun(x = (dst.S-d.S[k+1])/hs,
        kernel = "epanechnikov")$kx)*
        (kernel.fun(x = (dst.T-d.T[p+1])/ht,
        kernel = "epanechnikov")$kx)
        )/(bord.S[k]*bord.T[p]))
    )
  }
}

#-----#
# Organizar os valores para apresentar como grafico 3D
#-----#

marcas<-as.vector(gama)
grid<-merge(x=d.S[2:length(d.S)], y=d.T[2:length(d.T)])

pontos3D<-cbind(grid,marcas)

trellis.par.set("axis.line", list(col=NA, lty=0.1, lwd=0.5))
variograma <- wireframe(pontos3D$marcas ~ pontos3D$x*pontos3D$y,
data=pontos3D,
xlab = "r (km)",
ylab = "t (dias)",
zlab = expression(hat(gamma)(r,t)),
drape = TRUE,
colorkey = TRUE,

```

```

scales = list(arrows=FALSE,cex=0.55,
tick.number =7,
z = list(arrows=F),
distance =c(1, 1, 1.2)),
light.source = c(10,5,5),
col.regions = rainbow(200, s = 1, v = 1,
start = 1,
end = max(1,100 - 1)/100,
alpha = 1),
screen = list(z =-30, x=-60,y =5),
cuts = 40)

plot(variograma)

#-----#
# Grafico de nivel
#-----#

levelplot(pontos3D$marcas ~ pontos3D$x*pontos3D$y, data=pontos3D,
          xlab = "r",
          ylab = "t",
          main="",
          distance =c(1, 0.5, 1.1),
          light.source = c(10,5,5),
          col.regions = rainbow(100, s = 1, v = 1,
                                start = 1,
                                end = max(1,100 - 1)/100,
                                alpha = 1),
          cuts = 40,
          panel = function(...){
            inc.T<-(d.T[2]-d.T[1])/2
            inc.S<-(d.S[2]-d.S[1])/2
            panel.levelplot(...)
            panel.abline(h = c(d.T[2:length(d.T)]-inc.T,
                               max(d.T)+inc.T),
                          col="#333329")
            panel.abline(v = c(d.S[2:length(d.S)]-inc.S,
                               max(d.S)+inc.S),
                          col="#333329")
          }
          )

list(gama=gama,ds=d.S[2:length(d.S)],dt=d.T[2:length(d.T)])
}

```

**APÊNDICE B – Construção dos envelopes para  $\hat{\gamma}_{ST}(r,t)$  a partir de 199 simulações do processo pontual sob a hipótese nula de rotulagem aleatória e taxa de erro  $\alpha = 0,05$**

```

#-----#
# Numero de simulacoes a ser executado
#-----#

nsim=199

#-----#
# Array para acomodar os calculos do variograma de todas as simulacoes
#-----#

variog<-array(data = 0,c(length(a$ds),length(a$dt),nsim))

#-----#
# Cálculo dos variogramas para cada uma das simulacoes
#-----#

for (k in 1:nsim) {
    m<-sample(m)
    vario<-varioST(x,y,t,m,jan,ps=0.25,pt=0.25)
    variog[,k]<-vario$gama
}

#-----#
# Ordenar os valores obtidos para construir os envelopes
#-----#

vario.ord<-array(data = 0,c(length(variog[,1,1]),length(variog[1,,1]),199))

for(i in 1:length(variog[,1,1])){
    for(j in 1:length(variog[1,,1])){
        vario.ord[i,j,]<-variog[1,2,order(variog[1,2,])]
    }
}

#-----#
# Obter os quintos maiores e menores valores do variograma para os dados
# simulados
#-----#

env.max5<-matrix(nrow = length(variog[,1,1]),
ncol = length(variog[1,,1]))

env.min5<-matrix(nrow = length(variog[,1,1]),
ncol = length(variog[1,,1]))

env.max5<-vario.ord[,,(length(vario.ord[1,1,])-5)]
env.min5<-vario.ord[, ,6]

```

```

#-----#
# Obter os valores do variograma dos dados que ficam acima, dentro e
# abaixo do envelope
#-----#

gama.diff<-matrix(0,ncol = length(env.max5[1,]),
nrow = length(env.max5[,1]))

for (i in 1:length(env.max5[,1])) {
  for (j in 1:length(env.max5[1,])) {
    if (is.nan(env.max5[i,j])| is.nan(a$gama[i,j])){
      gama.diff[i,j]<-0
    } else {
      if ((env.max5[i,j]-a$gama[i,j])<0){
        gama.diff[i,j]<-1
      }
      if ((env.max5[i,j]-a$gama[i,j])>=0 &
          (env.min5[i,j]-a$gama[i,j])<=0){
        gama.diff[i,j]<-0
      }
      if ((env.min5[i,j]-a$gama[i,j])>0){
        gama.diff[i,j]<-(-1)
      }
    }
  }
}

#-----#
# Construcao do grafico:
#   vermelho - acima do envelope
#   branco - dentro do envelope
#   azul - abaixo do envelope
#-----#

marcas.sup1<-as.vector(gama.diff)
grid.sup1<-merge(x=a$ds,y=a$dt)

pontos3D.sup1<-cbind(grid.sup1,marcas.sup1)

color.ramp.length <- 40

negative.length <- round(abs(min(marcas.sup1))/diff(range(marcas.sup1))
*color.ramp.length)

positive.length <- color.ramp.length - negative.length

cols <- c(colorRampPalette(c("steelblue4", "white"))(negative.length),
colorRampPalette(c("white", "red"))(positive.length))

```

```
levelplot (pontos3D.sup1$marcas.sup1 ~ pontos3D.sup1$x*pontos3D.sup1$y,
data=pontos3D.sup1,
xlab = "r (km)",
ylab = "t (dias)",
main=" ",
distance =c(1, 0.5, 1.1),
light.source = c(10,5,5),
col.regions =cols,
cuts =2,
panel = function(...){
inc.T<-(a$dt[2]-a$dt[1])/2
inc.S<-(a$ds[2]-a$ds[1])/2
panel.levelplot(...)
panel.abline(h = c(a$dt-inc.T,max(a$dt)+inc.T),
col="#333329")
panel.abline(v = c(a$ds-inc.S,max(a$ds)+inc.S),
col="#333329")
}
)
```