VICTOR FERREIRA DA SILVA

PROPOSTA DE UM TESTE PARA VERIFICAR A HIPÓTESE DE INDEPENDÊNCIA ENTRE PONTOS E MARCAS QUANTITATIVAS DE UM PROCESSO PONTUAL MARCADO EM REDE LINEAR

LAVRAS–MG
2023

VICTOR FERREIRA DA SILVA

PROPOSTA DE UM TESTE PARA VERIFICAR A HIPÓTESE DE INDEPENDÊNCIA ENTRE PONTOS E MARCAS QUANTITATIVAS DE UM PROCESSO PONTUAL MARCADO EM REDE LINEAR

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

Prof. Dr. João Domingos Scalon
Orientador

LAVRAS–MG
2023

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

Ferreira da Silva, Victor.
Proposta de um teste para verificar a hipótese de independência entre pontos e marcas quantitativas de um processo pontual marcado em rede linear / Victor Ferreira da Silva. – 2023.
75 p. : il.

Orientador: João Domingos Scalon.
Tese (Doutorado) - Universidade Federal de Lavras, 2023.
Bibliografia.

1. Redes Lineares. 2. Processos Pontuais em Rede. 3. Manejo Florestal. I. Scalon, João Domingos. IV. Título.

VICTOR FERREIRA DA SILVA

PROPOSTA DE UM TESTE PARA VERIFICAR A HIPÓTESE DE INDEPENDÊNCIA ENTRE PONTOS E MARCAS QUANTITATIVAS DE UM PROCESSO PONTUAL MARCADO EM REDE LINEAR
A TEST PROPOSAL FOR VERIFICATION THE HYPOTHESIS OF INDEPENDENCE BETWEEN POINTS AND QUANTITATIVE MARKS OF A MARKED POINT PROCESS IN A LINEAR NETWORK

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

APROVADA em 20 de Janeiro de 2023.

Prof. Dr. Alex de Oliveira Ribeiro	UFLA
Prof. Dr. Elias Silva de Medeiros	UFGD
Prof. Dr. Fernando Luiz Pereira de Oliveira	UFOP
Prof. Dr. Paulo Henrique Sales Guimarães	DES- UFLA

Prof. Dr. João Domingos Scalon
Orientador

LAVRAS – MG
2023

A Deus que me deu forças, saúde e paciência para chegar até aqui.
A minha família que me ajudou e me apoiou pelas mãos todos esses anos e por todas as intempéries. Aos amigos e mentores que fiz e tive por toda essa jornada.
Humildemente e com toda gratidão, dedico.

AGRADECIMENTOS

Agradeço a Deus, por me dar saúde paciência, forças e sabedoria para continuar caminhando mesmo quando exausto e enfraquecido.

Minha mãe, Delfina e as ouras duas Maria Madalena e Maria do Carmo por serem o espelho de dedicação e carinho e o zelo de continuar vivendo e lutando para mostrar para as mesmas o quanto são capazes de criar bem um filho. A minha irmã e meu pai por estarem sempre ao meu lado, e meus primos por estarem comigo sem e sendo mais que irmãos.

Aos professores Scalon, Fernando, Elias, Paulo Henrique, Deive, Joel e Daniel, principalmente, pela paciência, dedicação, ensinamentos e conselhos. Aos professores do Programa de PósGraduação em Estatística e Experimentação Agropecuária, do Instituto de Ciências Exatas e Tecnológicas (ICET), da UFLA, pelos ensinamentos de qualidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos. Aos meus amigos do Departamento de Estatística (DES/ICET) por fazerem os meus dias na UFLA mais interessantes e sempre de bom humor.

A Kelly, Édipo, Henrique e Júlia por ser uma das pessoas especial que me ajudou muito a ver a vida de uma forma melhor e serem amigos par a vida toda apesar de qualquer distância. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Por fim, agradecer a todos que de forma direta ou indireta fizeram parte dessa caminhada.

Aproveite a caminhada, pois o fim quando menos se espera já chegou(...)
(Scalon, J. D.,2016)

RESUMO

Para analisar processos pontuais marcados é necessário supor que as observações (isto é, as marcas) e as localizações (isto é, os pontos) sejam dependentes. Existem vários métodos propostos para testar a validade dessa suposição no plano. Entretanto, não existem disponíveis na literatura testes estatísticos com esse propósito para o caso das observações estarem localizados em uma rede (ex. rios, ruas, etc). Nesta tese é proposta uma abordagem formal para testar essa suposição. O teste é baseado em duas características (E(t) e V(t)) de processos pontuais marcados estacionários e isotrópicos no plano que foram propostas por Schlather, Ribeiro e Diggle (2004). Essas quantidades são funções da distância *t* entre pontos e denotam a esperança e variância condicional, respectivamente, de uma marca dado que existe um outro ponto do processo pontual a uma distância *t*. Nesta tese, adaptamos o teste de independência entre marcas quantitativas e pontos proposto por Schlather, Ribeiro e Diggle (2004) para o caso de redes lineares. O teste é avaliado através de simulações Monte Carlo e aplicado em dados reais do diâmetro à altura do peito (DAP) de árvores da espécie *Heura Crepitans* localizadas nas margens dos rios que cortam a Fazenda *Canary*, sediada no estado do Acre. Os resultados demonstram que o teste proposto pode ser útil na detecção de dependência entre marcas quantitativas e pontos de processos pontuais marcados em redes lineares.

Palavras-chave: Redes Lineares. Teste de Independência. Processos Pontuais marcados. Marcas Quantitativas. Simulação Monte Carlo. Rede de Rios. Manejo Florestal.

ABSTRACT

To analyze marked point processes it is necessary to assume that observations (that is, marks) and locations (that is, points) are dependents. There are several available methods to test the validity of this assumption in the plan. However, there are no statistical tests available in the literature with this purpose in case of the observations are located in a network (e.g. rivers, streets, etc). In this thesis, it is proposed a formal approach to test this assumption. The test is based on two characteristics (E(t) and V(t)) of stationary and isotropic marked point processes in the plane that were proposed by Schlather, Ribeiro e Diggle (2004). These quantities are functions of the distance *t* between points and denote the expectation and conditional variance, respectively, of a mark given that there is another point of the point process at a distance *t*. In this thesis, we adapted the independence test between quantitative marks and points proposed by Schlather, Ribeiro e Diggle (2004) for the case of linear networks. The test was evaluated through Monte Carlo simulations and applied in actual data of the diameter at breast height (DBH) of trees of the species *Heura Crepitans* located on the banks of the rivers that cross the Fazenda *Canary* located in the state of Acre, Brazil. The results show that the proposed test may be useful in detecting dependence between quantitative marks and points of marked point processes in linear networks.

Keywords: Linear Network. Independence Test. Marked Point Process. Quantitative Marks. Monte Carlo Simulation. River Network. Forest Management

LISTA DE FIGURAS

Figura 1 - Kernel planar.....22
Figura 2 - Função K espacial.....25
Figura 3 - Variograma empírico para distribuição de variabilidade Abetos Noruegueses 30
Figura 4 - Gráficos de E(r) e V(r) contra a distância r 33
Figura 5 - Rede formada.....35
Figura 6 - Estimador kernel linear.....38
Figura 7 - Ilustração kernel linear.....39
Figura 8 - (a) Área (b) Rios (c) APP.....47
Figura 9 - Área de estudo 49
Figura 10 – Rede leste 53
Figura 11 - Rede linear Oeste com as ocorrências 54
Figura 12 - Alisamento Kernel Leste em tons de cinza.....55
Figura 13 - Alisamento Kernel Leste.....56
Figura 14 - Alisamento kernel Leste 57
Figura 15 - Função K de Ripley na rede leste 58
Figura 16 - Teste de independência para rede Leste.....61

LISTA DE TABELAS

Tabela 1 - Desempenho do teste em 500 realizações 59

LISTA DE QUADROS

Quadro 1- Critérios 40

SUMÁRIO

1 INTRODUÇÃO **14**

2 REFERENCIAL TEÓRICO **17**

2.1 Estatística espacial 17

2.2 Processos pontuais 18

2.3 Processos pontuais espaciais não marcados 19

2.3.1 Efeitos de primeira ordem 20

2.3.2 Efeitos de segunda ordem 22

2.3.3 Análise contra a hipótese de completa aleatoriedade espacial (CAE) 25

2.4 Processos pontuais espaciais marcados 27

2.4.1 Propriedades de primeira e segunda ordem 28

2.4.2 Função de correlação de marcas 28

2.4.3 Função *K* ponderada por marcas 29

2.4.4 Variograma de marcas 30

2.5 Independência entre marcas e pontos no plano 30

2.5.1 As funções descritoras E e V 31

2.5.2 Estimadores para as funções descritoras E e V 31

2.5.3 Independência entre marcas e pontos no plano 32

2.6 Processos pontuais não marcados em redes 33

2.6.1 Definições básicas de uma rede 34

2.6.2 Métricas de distância em rede 35

2.6.3 Efeitos de primeira ordem em um processo pontual em rede 36

2.6.4 Efeitos de segunda ordem em um processo pontual em rede 39

2.7 Processo pontual marcado em redes lineares 40

2.7.1 Propriedades de primeira e segunda ordem 41

2.7.2 A função *K* ponderada por marcas em rede 41

3.1 Estimadores de E(r) e V(r) para redes lineares 43

3.2 Método de Monte Carlo para testar a independência entre pontos e marcas em uma rede linear 45

4 MATERIAIS E MÉTODOS **47**

4.1 Materiais 47

4.2 Métodos 49
4.3 Softwares 51
5 RESULTADOS E DISCUSSÃO **52**
5.1 Análise de primeira ordem 52
5.2 Análise de segunda ordem 57
5.3 Independência entre pontos e marcas em redes lineares 58
5.3.1 Resultados da simulação 59
5.3.2 Aplicação em dados reais 60
6 CONSIDERAÇÕES FINAIS **62**
REFERÊNCIAS **65**
APÊNDICE A – SCRIPT **68**

1 INTRODUÇÃO

Existem muitos dados que são referenciados e, portanto, passíveis de serem analisados através de métodos da estatística espacial. Por exemplo, em inventários de florestas nativas, em geral, coleta-se informações sobre a localização das árvores e dos diâmetros à altura do peito (DAP) dessas árvores. Os pesquisadores geralmente estão interessados em estudar esses dados para entender a natureza e a força de possíveis interações entre a localização das árvores e o volume da madeira (obtido a partir do DAP).

Uma abordagem útil para analisar esses dados é utilizar métodos desenvolvidos tendo como base a teoria de processos pontuais marcados, onde as marcas são as medidas de DAP e os pontos são as localizações das árvores. Para utilizar esses métodos é absolutamente necessário que os pontos e as marcas sejam dependentes.

Entretanto, uma suposição comumente usada para análise desses dados é que as marcas e os pontos são independentes. O principal benefício dessa suposição é que os dados podem ser analisados utilizando métodos geoestatísticos (ex. variograma e krigagem) que são muito mais desenvolvidos e difundidos que os métodos baseados na teoria de processos pontuais marcados.

Embora seja uma suposição conveniente, a independência entre marcas e pontos pode não valer na prática em muitas situações. Por exemplo, em dados de florestas nativas, o DAP das árvores (isto é, as marcas) pode depender da relativa posição das árvores (ou seja, os pontos) devido à competição por luz e/ou nutrientes entre árvores vizinhas. Nessa situação, qualquer análise do DAP usando geoestatística que desconsidera possíveis dependências entre pontos e marcas deve ser, no mínimo, questionada. Nesse sentido, métodos estatísticos para detectar a dependência entre marcas e pontos é de fundamental importância para garantir a qualidade das análises estatísticas, consequentemente, das implicações científicas dessas análises.

Assim, várias métodos estatísticos têm sido propostos para detectar a independência entre marcas e pontos, incluindo o uso da função de correlação de marca/variograma (WÄLDER; STOYAN, 1996), a função J (LIESHOUT, 2006), as funções E e V (SCHLATHER; RIBEIRO; DIGGLE, 2004), a função K (GUAN, 2006), a estatística do Qui-quadrado (GUAN; AFSHARTOUS, 2007) e a estatística de Kolmogorov-Smirnov (ZHANG, 2014), para citar apenas alguns.

Todos os métodos mencionados no parágrafo anterior foram desenvolvidos para dados referenciados no espaço. Entretanto, existem situações em que os dados espaciais ocorrem em uma rede, como é o caso de árvores que são localizadas nas margens de rios ou o caso de acidentes de trânsitos que ocorrem em ruas e estradas. Como métodos baseados na teoria de

processos pontuais marcados em redes lineares ainda estão em desenvolvimento, os pesquisadores aplicam métodos da geoestatística para redes lineares que são amplamente desenvolvidos e, inclusive, contando com disponibilidade de *softwares* para análise de dados. Essa prática, como ocorre no plano, deve ser questionada, uma vez que não existe uma garantia de independência entre pontos e marcas e, consequentemente, os variogramas produzidos poderão ser incorretos.

Nesse sentido, métodos estatísticos para detectar a dependência entre marcas e pontos em redes lineares também é de fundamental importância para garantir a qualidade das análises. Até o momento, não existe disponível na literatura qualquer método estatístico para testar a hipótese de independência entre marcas e pontos de processos pontuais marcados em redes lineares. Esta tese vem preencher essa lacuna. Para tal, propõe-se um método para testar a hipótese nula de independência entre marcas e pontos de processos pontuais marcados em redes lineares, tendo como base a teoria proposta por (SCHLATHER; RIBEIRO; DIGGLE, 2004).

Schlather, Ribeiro e Diggle (2004) introduzem duas funções para caracterizar processos pontuais marcados com o objetivo de avaliar se os valores das marcas podem ser modelados por um campo aleatório independente do processo pontual não marcado (ou seja, os locais). Mais especificamente, E(h) e V(h) representam a esperança e a variância condicional de uma marca, respectivamente, dado que existe outro ponto do processo pontual a uma distância h. Eles também propuseram um teste formal de Monte Carlo para independência entre marcas e pontos para uma classe de marcas gaussianas estacionárias baseado no desvio das estimativas de E(h) e Vh) de funções constantes pois, sob a hipótese nula de independência entre marcas e pontos, E(h) e V(h) deveriam ser constantes.

O principal objetivo desta tese é adaptar o teste proposto por Schlather, Ribeiro e Diggle (2004) para o caso de processos pontuais marcados em redes lineares, avaliá-lo através de simulações Monte Carlo e aplicá-lo em um caso real de dados disponíveis em uma rede linear.

Deste modo, esta tese terá a seguinte organização:

a) capítulo 2 - Referencial Teórico: será apresentada uma revisão de literatura contendo uma breve introdução sobre a estatística espacial e as diferentes áreas de estudo; uma abordagem da teoria de processos pontuais espaciais não marcados e marcados; uma apresentação da teoria de processos pontuais em redes lineares não marcados e marcados; apresentação de testes de independência entre marcas e pontos em configurações pontuais marcadas no plano utilizanndo as funções E(h) e V(h);

b) capítulo 3 - Independência entre marcas e pontos em processos pontuais marcados em redes lineares: será apresentado o desenvolvimento teórico dos

estimadores E(r) e V(r) para processos pontuais marcados em redes lineares; proposto o novo método para análise da independência entre marcas e pontos para processos pontuais marcados em redes lineares; construção de envelopes de simulação Monte Carlo para testar a hipótese de independência entre marcas e pontos em redes lineares;

c) capítulo 4 - Material e Métodos: será feita uma descrição da base de dados, métodos, *softwares* e funções em *R* utilizados nas análises;

d) capítulo 5 - Resultados e discussão: serão apresentados os resultados obtidos, discussões a respeito dos métodos utilizados e seus desempenhos;

e) capítulo 6 - Considerações Finais: serão feitas considerações sobre os aspectos positivos e negativos dos métodos propostos e sugestões para futuros trabalhos;

f) referências - serão apresentadas todas as referências bibliográficas utilizadas na tese;

g) apêndice - serão apresentados os *scripts* das funções em R utilizadas nas análises de dados.

2 REFERENCIAL TEÓRICO

Neste capítulo será apresentado o estado da arte do processo pontual em rede linear com suas principais e tradicionais referências. Um ponto de suporte e partida para a pesquisa que foi desenvolvida neste trabalho.

2.1 Estatística espacial

Em muitas situações os pesquisadores deparam com observações que foram coletadas com suas respectivas coordenadas geográficas. Nesse tipo de dado, uma pergunta principal é se a ocorrência dessas observações são correlacionadas (interagem entre si) ou não. Caso as observações sejam correlacionadas (dependentes), o uso de métodos da estatística usual podem ser inapropriados. Nesse cenário, recomenda-se utilizar métodos da estatística espacial.

Conforme Bailey e Gatrell (1995), a estatística espacial pode ser definida como um conjunto de técnicas que busca descrever os padrões existentes em dados que são espacialmente localizados, onde a informação da localização dos dados é colocada explicitamente na análise. Segundo Cressie (1993), a estatística espacial pode ser dividida em três subáreas, sendo elas:

a) Geoestatística: São dados de superfície contínua, mas coletados em alguns de seus pontos, dado que o fenômeno está distribuído continuamente sobre uma superfície, como, por exemplo, teor de um determinado nutriente no solo, ou profundidade de um lago;
b) 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, bairros, quarteirões etc, em que não se sabe exatamente onde os eventos ocorrem mas dispõe-se de um valor agregado a uma área, que comumente será representada pontualmente por um centroide;
c) Processos pontuais: São dados em que a informação é a própria localização do evento. Localizações da ocorrência de crimes, epidemias e plantas de uma determinada espécie são alguns exemplos desses tipos de dados. Cada localização pode estar associada a alguma variável marca (uma variável aleatória atribuída a cada ocorrência). São exemplos: tipos de crimes, altura da planta, etc.).

A inferência utilizada em dados espaciais é realizada por meio de métodos e processos estocásticos que podem ser definidos como uma coleção ou família de variáveis aleatórias definidas no mesmo espaço de probabilidade e indexadas por um conjunto de índices. Um

processo espacial é um caso especial de um processo estocástico cuja indexação é a posição do evento.

Schabenberger e Gotway (2005) descreve um processo estocástico espacial como uma coleção de variáveis aleatórias indexadas por um conjunto $\boldsymbol{D} \subset \mathrm{R}^d$ contendo coordenadas espaciais $\mathrm{s} = [s_1,s_2,...,s_d]'$. Usualmente, para um processo no plano $d = 2$, as coordenadas de longitude e latitude são definidas como $\mathbf{s} = [x,y]'$. Se a dimensão d do conjunto de índices do processo estocástico for maior que um, o processo estocástico é, muitas vezes referido como um campo aleatório. A notação usual para um processo estocástico espacial é dado por

$$\{Z(\mathbf{s}) : \mathbf{s} \in \boldsymbol{D} \subset \mathrm{R}^d\},$$

em que $Z(\boldsymbol{s})$, representam os eventos nas localizações $\boldsymbol{s}$.

A estatística espacial apresenta uma ampla diversidade de métodos, divididas em classes, que dependem da natureza estocástica dos dados. Para cada umas dessas divisões existem métodos específicos para realização da análise estatística. Este trabalho focará nos métodos estatísticos para análise de processos pontuais no espaço bidimensional e em redes lineares.

2.2 Processos pontuais

Segundo Diggle (2003), um processo pontual pode ser visto como um processo estocástico (mecanismo gerador) que gera um conjunto de eventos contáveis (x_i,y_j) no plano. Assim, os dados provenientes de um processo pontual são pontos localizados no espaço e identificados por coordenadas do fenômeno em estudo. Em geral, o objetivo da análise está em determinar se a ocorrência dos pontos na área em estudo apresenta algum padrão espacial, como, por exemplo, agrupamentos. Entre as áreas de aplicações, incluem Ciências Florestais, Ecologia, Segurança Pública, Epidemiologia, etc.

Diggle (2013) e Schabenberger e Gotway (2005) afirmam que os dados de um processo pontual podem ser obtidos por amostragem ou mapeamento. Quando amostrados, apenas um subconjunto dos eventos gerados pelo processo estocástico é observado e considerado na análise. Se todos os eventos gerados pelo processo estocástico forem observados e considerados na análise, os dados do processo pontual são denominados de dados mapeados ou mapa de pontos.

No presente trabalho será considerado apenas mapa de pontos.

Em geral, os processos pontuais podem ser divididos em duas categorias:

a) não marcados: Considera apenas a localização do evento, como, por exemplo, localização de árvores em uma floresta ou localização das ocorrências de uma doença em uma região;

b) marcados: Considera uma variável aleatória associada a cada evento (localização ponto do processo) como, por exemplo, altura ou espécie das uma árvores.

Além disso, os processos pontuais marcados e não-marcados podem ser divididos em:

a) espaciais: Considera os eventos (pontos) ocorrendo em qualquer posição do espaço bidimensional (plano) ou tridimensional;

b) redes lineares: Considera os eventos (pontos) ocorrendo em uma rede linear.

Como apresentado em 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 o autor, são úteis no estudo teórico de processos pontuais e na inferência estatística sobre padrões pontuais.

Apesar do enfoque desta tese estar na análise das propriedades de primeira e segunda ordem em mapas de processos pontuais marcados em redes lineares, este trabalho inicia com a apresentação teórica dos processos pontuais espaciais. Isso é necessário tendo em vista que existe uma relação próxima entre os dois processos, e o processo pontual espacial não marcado é mais familiar para os pesquisadores.

2.3 Processos pontuais espaciais não marcados

Formalmente, segundo Diggle (2013), um processo pontual espacial não marcado é um mecanismo estocástico que gera um conjunto contável finito ou infinito de pontos (denominados eventos) $\{s_i : i = 1,2,...\}$ tais que $\mathbf{s}_i \in S \subset R^2$, ou seja, cada variável aleatória denota a localização de um evento específico no espaço bidimensional.

Um processo pontual espacial não marcado fundamental para as análises é o processo de Poisson homogêneo, também conhecido como modelo de completa aleatoriedade espacial (CAE). Segundo Schabenberger e Gotway (2005), a CAE ocorre se o padrão pontual atende aos seguintes critérios:

a) O número médio de eventos por unidade de área, ou seja, a intensidade $\lambda(\boldsymbol{s})$ é homogênea em S;

b) O número de eventos em duas sub-regiões sem sobreposição A_1 e A_2 são independentes;

c) O número de eventos em qualquer sub-região segue uma distribuição Poisson.

Pode-se observar que sob a suposição de CAE, os eventos distribuem-se uniformemente e independentemente ao longo do domínio. Observa-se também que esse processo não exibe qualquer estrutura espacial, ou seja, não apresenta efeitos de primeira e segunda ordem. Por essas características, a CAE serve como hipótese nula para muitas investigações estatísticas em padrões pontuais.

Dois conceitos que são importantes para a análise estatística de um processo espacial pontual e que estão intimamente relacionados aos efeitos de primeira e segunda ordem são os conceitos de estacionariedade e isotropia. Um processo é dito estacionário ou homogêneo se os efeitos de primeira e segunda ordem são invariantes sob a translação, ou seja, a intensidade de primeira ordem $\lambda(\mathbf{s}) = \lambda$ não varia no espaço e a intensidade de segunda ordem $\lambda_2(s_i,s_j) = \lambda_2(s_i - s_j)$ depende apenas da distância relativa entre os pontos. Cressie (1993) denomina de "movimento invariante" aquele processo que é tanto estacionário como isotrópico. Nesse caso, a covariância depende apenas da distância entre dois eventos e, portanto, os efeitos direcionais são ausentes, assim tem-se que:

$$\lambda_2(\mathbf{s}_i,\mathbf{s}_j) = \lambda_2(t) \tag{2.1}$$

em que $t = \|s_i - s_j\|$ é a distância euclidiana entre dois pontos, e como consequência da independência de um processo de CAE, $\lambda_2(t) = \lambda^2$.

Tendo em vista o que foi apresentado anteriormente, pode-se observar que importantes aspectos do comportamento de um processo pontual espacial podem ser caracterizados a partir das análises dos efeitos de primeira e segunda ordem.

2.3.1 Efeitos de primeira ordem

Os efeitos de primeira ordem, também conhecidos como globais ou de larga escala, descrevem quando o valor esperado (média) varia dentro do espaço. O objetivo da análise está em estimar a intensidade do processo, ou seja, o número de eventos por unidade área.

As propriedades de primeira ordem do processo pontual espacial podem ser definidas pela função intensidade dada por

$$\lambda(\mathbf{s}) = \lim_{|d\mathbf{s}|\to 0} \left\{ \frac{\mathbb{E}[N(d\mathbf{s})]}{d\mathbf{s}} \right\} \tag{2.2}$$

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

Um estimador para intensidade de um processo pontual espacial estacionário (intensidade constante em toda a região de estudo) é dado pela razão entre o número observado de eventos (*n*) e a área de estudo *A*. Um estimador (global) para a intensidade é dado por:

$$\hat{\lambda} = n/|A| \tag{2.3}$$

Caso exista alguma razão para supor que o processo pontual espacial não seja estacionário (a intensidade não é constante em toda a região), pode-se usar um estimador não paramétrico do tipo kernel.

Este estimador baseia-se na relação entre as estimativas de densidade e intensidade e utiliza a contagem de todos os eventos dentro de uma região de influência, ponderando-os pela distância de cada evento à localização de interesse (CÂMARA; MONTEIRO; MEDEIROS, 2003).

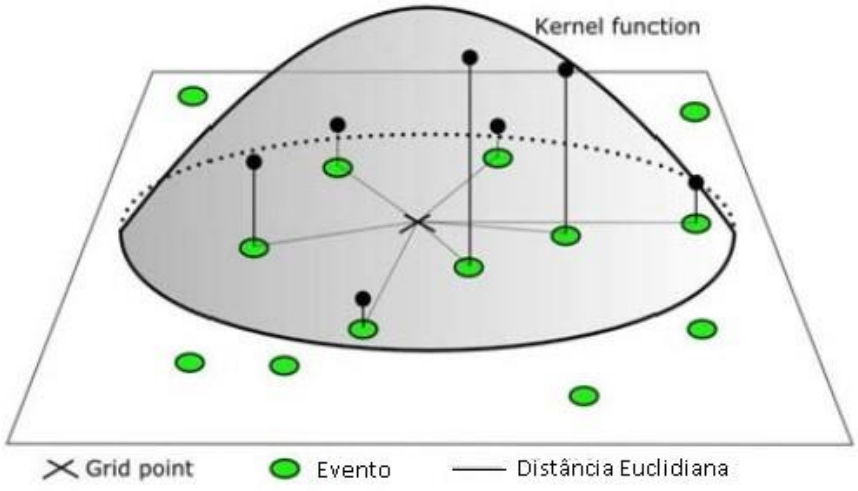
Neste caso, considere *s* uma localização qualquer em uma área *A* e $s_1,\ldots,s_n$ são as localizações dos *n* eventos observados nesta área. Um estimador kernel de intensidade $\lambda(s)$ na posição *s*, com correção do efeito de borda, proposto por Diggle (1985), pode ser expresso por:

$$\hat{\lambda}(s) = \frac{1}{p_h(s)} \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{s-s_i}{h}\right), \tag{2.4}$$

em que *h* é denotado como raio de influência ou largura de banda. Neste trabalho adota-se o termo *bandwith* (largura de banda) para o mesmo, em que $p_h(s) = \int_A h^{-2}$.

A função kernel espacial pode ser vista geometricamente como:

Figura 1 – Kernel planar



Legenda: Esquema de kernel no plano onde as alturas dos pontos representam sua intensidade.
Fonte: Adaptado de Câmara et al. (1995)

Pode-se observar na figura anterior que $k\left(\frac{s-s_i}{h}\right)$, é um fator de correção do efeito de bordas e *K*(.) é uma função de densidade de probabilidade. Segundo Schabenberger e Gotway (2005), uma das funções de densidade mais aplicada é a kernel quadrática dada por:

$$k(t) = \begin{cases} \frac{3}{4}(1-t^2), & \text{se } |t| \leq 1 \\ 0, & \text{caso contrário.} \end{cases} \qquad (2.5)$$

De acordo com Schabenberger e Gotway (2005), a escolha de *K*(.) tem menos importância do que a escolha da largura de banda. Se *h* é pequena, as estimativas quase não apresentam viés mas apresentarão variabilidade alta. Com o aumento da largura de banda *h*, a estimativa é suavizada, variando menos, entretanto, ficando mais tendenciosa. Existem várias propostas (ex. erro quadrático médio como validação cruzada) para definir o valor "adequado" para *h*, conforme podem ser visto em Diggle (2013) e Schabenberger e Gotway (2005).

2.3.2 Efeitos de segunda ordem

Os efeitos de segunda ordem, ainda de acordo Câmara e Carvalho (2004), são aqueles denominados locais ou de pequena escala e representam a dependência espacial do processo que é proveniente da estrutura de correlação espacial. Na análise dos efeitos de segunda ordem,

procura-se responder se os eventos exibem algum grau de relação entre eles, ou seja, se estão mais próximos ou mais distantes entre si do que seria esperado no caso da completa aleatoriedade espacial.

As propriedades de segunda ordem podem ser formalmente definidas, conforme Diggle (2013), pela função intensidade de segunda ordem dada por:

$$\lambda_2(s_i, s_j) = \lim_{|ds_i|,|ds_j|\to 0} \frac{E[N(ds_i)N(ds_j)]}{|ds_i||ds_j|}, \tag{2.6}$$

em que $N(ds_i)$ e $N(ds_j)$ representam o número de eventos dentro pequenos círculos centrados nas localizações s_i e s_j com áreas iguais a $|ds_i|$ e $|ds_j|$, respectivamente.

Uma medida fortemente relacionada a essa propriedade é a intensidade condicional $\lambda_c(s_i, s_j) = \frac{\lambda_2(s_i,s_j)}{\lambda(s_j)}$, a qual corresponde a intensidade no ponto s_i condicional a informação de que existe um evento em s_j.

Outra medida é a função de correlação de pares, a qual é baseada no produto de densidade de segunda ordem e, pode ser expressa por

$$g(s_i, s_j) = \frac{\lambda_2(s_i,s_j)}{\lambda(s_i)\lambda(s_j)}. \tag{2.7}$$

Existem na literatura vária funções descritoras (ex. *F*, *G*, *J*, *K*, *L*, etc.) que foram propostas para caracterizar as propriedades de segunda ordem de um processo pontual espacial não marcado conforme pode ser visto em Diggle (2013). Uma das funções mais conhecidas e utilizadas é a função *K*.

A função *K* de Ripley (1976) e Ripley (1977) também conhecida como "função de segunda ordem reduzida" está inteiramente ligada à interação dos diferentes eventos do processo pontual em análise.

Segundo Cressie (1993), a função *K* tem como vantagem detectar o padrão espacial pontual em diferentes escalas de distâncias, simultaneamente. A função *K* para um processo pontual estacionário e isotrópico é definida por Ripley (1976) e Ripley (1977) como sendo

$$K(t) = \lambda^{-1}E[N(t)], \tag{2.8}$$

em que $E[.]$ é a esperança, $N(t)$ é o número de eventos encontrados em uma distância t e λ é a intensidade de primeira ordem do processo pontual.

Com intenção de estabelecer uma relação entre $K(t)$ e $\lambda_2(t)$, segundo Diggle (2013), assume-se que o processo seja ordenado. Assim, não deverão ocorrer múltiplos eventos sobrepostos, ou seja, $P\{N(ds) > 1\}$ é uma ordem de grandeza menor que $|ds|$. Significando que

$E[N(ds)] \sim P\{N(ds) = 1\}$no sentido que a relação destas duas quantidades tende para 1 quando $|ds_i| \rightarrow 0$. De maneira análoga, $E\left[N(ds_i)N\left(ds_j\right)\right] \sim P\left\{N(ds_i) = N\left(ds_j\right) = 1\right\}$. Observando essas condições, o número esperado de eventos dentro da distância t de um evento arbitrário pode ser obtido integrando a função intensidade condicional sobre um disco com centro na origem e raio t. Assim,

$$K(t) = \lambda^{-1} \int_0^{2\pi} \int_0^t \lambda_c(s \mid o) s ds d\theta. \tag{2.9}$$

Considerando que, $\lambda_2(s) = \lambda^2$, consequentemente,

$$K(t) = 2\pi\lambda^{-2} \int_0^t \lambda_2(s) s ds. \tag{2.10}$$

Conforme definido anteriormente, a função $K(t)$ representa o número esperado de eventos dentro de uma distância t de um evento arbitrário. Assim, para um processo sob a hipótese de CAE, $\lambda_2(s) = \lambda^2$ e, portanto, a equação (2.11) se reduz a:

$$K(t) = 2\pi\lambda^{-2} \int_0^t \lambda_2(s) s ds \tag{2.11}$$

$$= 2\pi\lambda^{-2}\lambda^2 \frac{t^2}{2}$$

$$= \pi t^2. \tag{2.12}$$

O estimador mais simples para λ utilizado na função K univariada definida na equação (2.8) é o número observado de eventos por unidade de área, ou seja, $\hat{\lambda} = n/|A|$. Segundo Diggle (2013), considerando que $E(t) = E[N(t)]$, segue que a interpretação é a mesma, ou seja, o número esperado de eventos dentro de uma distância t de um evento arbitrário. Assim, um estimador para $E(t)$ pode ser expresso por

$$\widehat{E}(t) = n^{-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} I(t_{ij} \leq t), \tag{2.13}$$

em que I (.) representa a função indicadora e $t_{ij} = \|s_i - s_j\|$.

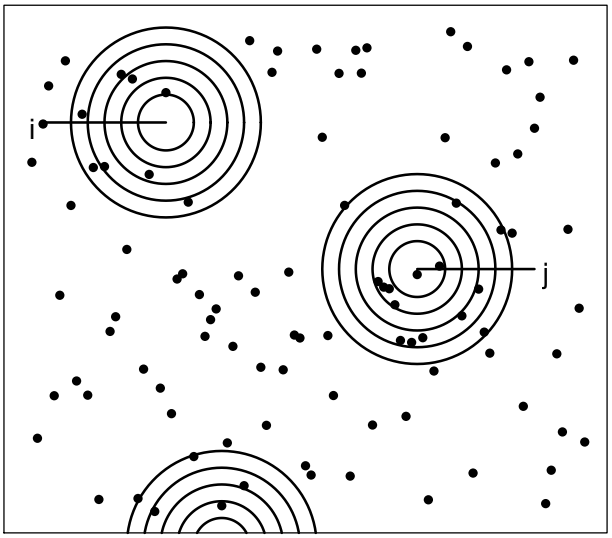
Portanto, um estimador não tendencioso (corrigido para efeito de bordas) para a função *K*, proposto por Ripley (1976), é dado por:

$$\hat{K}(t) = \frac{|A|}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{I(t_{ij})}{w_{ij}} \right). \quad (2.14)$$

em que e w_{ij} é um fator de correção para o efeito de borda que representa a proporção da circunferência ao redor de um evento s_i, passando sobre o evento s_j que está dentro da região A.

Deve-se observar que *K*(*t*) é uma função acumulada, pois são traçados círculos com origem posicionadas sobre os eventos até a uma distância máxima *t* e, então, são computados todos os eventos dentro desses círculos. Uma ilustração da estimação da função *K* pode ser observada na Figura 2.2.

Figura 2 – Função K espacial



Legenda: Diagrama demostrativo da função K e sua varredura ao longo do plano.
Fonte: Adaptado de Bailey e Gatrell (1995)

Uma vez obtida as estimativas da função *K* para diversas distâncias *t*, elas podem ser utilizadas para testar a hipótese nula de completa aleatoriedade espacial.

2.3.3 Análise contra a hipótese de completa aleatoriedade espacial (CAE)

Para a análise da interação entre os eventos de uma configuração pontual espacial usualmente, compara-se a função *K* estimada com a função *K* teórica sob a suposição de hipótese nula de CAE para um conjunto de distâncias, conforme visto nos capítulos anteriores. Para avaliar se as diferenças entre essas duas funções é estatisticamente significante para uma particular distância, pode-se utilizar tanto testes de hipóteses formais quanto procedimentos gráficos conhecidos como envelopes de confiança. Em ambos os casos, utilizam-se simulações

de Monte Carlo como vista em Scalon e Silva (2006) e Diggle (2013) e adaptada para redes lineares em Baddeley, Rubak e Turner (2015).

Conforme descrito em Scalon e Silva (2006), para conduzir um teste de hipótese formal contra a hipótese nula de CAE, é necessário definir uma medida de discrepância entre a função K estimada $\hat{K}(t)$ e a teórica nula $K(t)$. Essa medida pode ser obtida tanto pela distância de Cramer-Von Mises quanto pela distância de Kolmogorov, que podem ser definidas, respectivamente, por:

$$u_i = \int_0^{t_0} \left(\hat{K}_i(t) - K(t)\right)^2 dt \tag{2.15}$$

$$u_i = \max_{0 \le t_0} \left|\hat{K}_{i(t)} - K(t)\right|, \text{ para } u_i: i = 2, \dots, s, \tag{2.16}$$

em que a distância t_0 é escolhida em relação ao tamanho da área finita.

Sobretudo, as equações (2.15) e (2.16) não possuem distribuições amostrais conhecidas. Assim, Diggle e Milne (1983) propõe o uso da abordagem de Monte Carlo para realização do teste de hipótese, da seguinte forma:

a) Calcula-se o valor de u_1 para os dados observados;

b) Simular padrões de pontos espaciais independentes S sob hipótese nula de CAE com a mesma intensidade que o padrão observado;

c) Calcula-se u_i para cada padrão aleatório espacial simulado ($i = 2,...S$);

d) Comparar o valor de u_1 com os valores ($i = 2,...S$) para os padrões simulados. Se u_1 estiver localizado entre os maiores de $u_2, u_3, \dots, u_S$, indicará afastamento da hipótese de CAE.

Suponha $u_1 = u_{(j)}$, para alguns $j \in \{1,...,s\}$, então rejeita-se a hipótese de aleatoriedade espacial se $valor - p = \left(\frac{s+1-j}{s}\right) \le \alpha$, em que α é o nível de significância.

Alternativamente ao teste formal, o método de Monte Carlo também pode ser utilizado para a construção de envelopes de simulação sob a hipótese de CAE utilizando-os pelos seguintes passos:

a) Calcula-se o valor de $\hat{K}_1$ para o processo pontual observado para um conjunto de distâncias;

b) Calcula-se $\hat{K}_i$: $i = 2,3,...,s$ para cada $s-1$ simulações independentes e identicamente distribuídas sob a hipótese de CAE;

c) São construídos os envelopes de simulações superior e inferior para um conjunto de distâncias, em que são considerados os valores máximos e mínimos de $\hat{K}_i$, definidos como:

$$U = \max\{\hat{K}_i\}, i = 2,...,s.\ L = \min\{\hat{K}_i\}, i = 2,...,s; \quad (2.17)$$

d) constrói-se um gráfico, em que os valores de $\hat{K}_i$, U e L são colocados no eixo das coordenadas e as distâncias no eixo das abcissas.

Se para qualquer distância a linha de $\hat{K}_1$ estiver fora dos envelopes de simulação, há evidências estatísticas de existência de efeitos de segunda ordem no processo pontual marcado e, portanto, rejeitando a hipótese nula de completa aleatoriedade espacial.

2.4 Processos pontuais espaciais marcados

Penttinen, Stoyan e Henttonen (1992) definem um processo pontual marcado como um processo estocástico $\Psi = [s_i;m(s_i)]$ que gera os eventos (s_i) e as marcas associadas aos eventos $m(s_i)$. Uma realização de Ψ é um conjunto de posições junto às marcas associadas, e essas marcas podem ser quantitativas ou qualitativas.

As marcas qualitativas são variáveis categóricas discretas, que descrevem diferentes tipos de eventos, tais como espécies de plantas. O processo é chamado de bivariado se for apenas dois tipos de marcas; caso contrário, é multivariado.

As marcas quantitativas são valores reais que descrevem, por exemplo, altura da árvore, diâmetro de células. Claramente, as marcas quantitativas podem ser reduzidas a marcas qualitativas, dividindo as marcas em classes discretas, como "pequeno", "médio" e "grande".

O processo pontual marcado é definido por meio de um processo estacionário e isotrópico, semelhante à ideia de processos pontuais não marcados. Illian et al. (2008) definem um processo $\Psi = [s_1;m(s_1)],[s_2;m(s_2),...]$ juntamente com um processo transladado $\Psi_s = [s_1+s;m(s_1)],[s_2+s;m(s_2),...]$. Isso significa que, no processo pontual marcado transladado, as marcas permanecem as mesmas, somente os pontos são transladados. Assim, o processo pontual marcado Ψ é dito estacionário, ou seja, os efeitos de primeira e segunda ordem foram constantes em toda a região W estudada. A definição de isotropia é análoga a processos pontuais não marcados, visto anteriormente. Considera-se o processo Ψ estacionário, o conceito de isotropia, se além de ser estacionário, no processo Ψ a covariância dos pontos s_i e s_j dependa apenas da distância euclidiana entre eles, e não da direção.

2.4.1 Propriedades de primeira e segunda ordem

Um processo pontual marcado é descrito em termos de propriedades de primeira e segunda ordem. No caso estacionário, são utilizados dois tipos diferentes de características de primeira ordem: a que diz respeito aos eventos, isto é, a intensidade λ, e a outra que descreve as marcas (ILLIAN et al., 2008). Os estimadores globais e locais para as propriedades de primeira ordem para processos pontuais marcados são obtidos a partir de extensões diretas dos estimadores de processos pontuais não marcados.

As características de segunda ordem são úteis para caracterizar não apenas a variabilidade da distribuição dos eventos, mas também a variabilidade das marcas e, ainda, descrever correlações entre marcas e pontos (WÄLDER; STOYAN, 1996). Estimadores das propriedades de segunda ordem, em geral, dependem do tipo de marca.

Se as marcas forem qualitativas, o interesse é estudar as posições relativas dos diferentes tipos de pontos, verificar se existe agregação ou repulsão dentro do tipo e entre os tipos.

Já para marcas quantitativas, a análise avalia se as marcas variam continuamente e verificam a independência entre as marcas, isto é, até que ponto a interação entre as marcas tem influência sobre estas. Por exemplo, as marcas de pontos vizinhos podem tender a ser menores (maiores) do que a marca média, isto é, pode haver inibição ou estimulação das marcas. Outra questão diz respeito à semelhança numérica das marcas de eventos vizinhos. Isso significa que esses pontos tendem a ter marcas semelhantes, mas o comportamento oposto também pode ser observado, ou seja, pontos específicos podem dominar os pontos em sua vizinhança e, portanto, ter uma marca grande, enquanto os outros pontos têm marcas pequenas.

As principais funções para analisar as propriedades de segunda ordem em processos pontuais marcados são as funções: correlação marcada, K marcada, variograma marcado, U e V. Essas funções são apresentadas a seguir.

2.4.2 Função de correlação de marcas

A função de correlação de marcas (ou marcada), proposta por Wälder e Stoyan (1996), é uma medida da força e do sentido da dependência entre as marcas e pontos separados por uma distância h. Assumindo que um processo pontual marcado é estacionário e isotrópico então a função de correlação marcada é definida por:

$$\rho_f(h) = \frac{E(f(m_i,m_j)\times h)}{\mu^2} \tag{2.18}$$

em que $(f(m_i,m_j))$ é chamada de função teste; $\mu = E(m_i)$; $h = |s_i - s_j|$.

A interpretação da função de correlação marcada é:

a) $\rho_f(h) > 1$ indica correlação positiva, caracterizando atração entre as marcas;

b) $\rho_f(h) = 1$ indica independência entre a marcas;

c) $\rho_f(h) < 1$ indica correlação negativa, caracterizando repulsão entre as marcas.

2.4.3 Função *K* ponderada por marcas

A função *K* ponderada por marcas, de acordo com Penttinen, Stoyan e Henttonen (1992), é uma extensão da função *K* de Ripley, na qual para cada par de eventos existe uma ponderação por uma função teste $f(m_1,m_2)$ de suas respectivas marcas. Um estimador sem viés, que corrige o efeito de borda para a função K_f marcada é dado por

$$\widehat{\boldsymbol{K}}_f(\mathrm{h}) = \frac{|W|}{n^2}\sum_{k=1,l=1k\neq 1}^{n}\sum^{n} f(m(sk),m(sl))1(||sk - sl|| \leq \mathrm{h})\delta(sk,sl), \tag{2.19}$$

em que n é o número de pontos marcados na região de estudo W de área $|W|$; f é uma função que atribui um peso a um par dos pontos s_k e s_l dependendo do tipo de suas marcas $m(s_k)$ e $m(s_l)$; 1(.) é a função que é igual a 1 se o argumento for verdadeiro e 0 caso contrário; $\delta(s_k,s_l)$ é um termo de correção de borda, conforme definido anteriormente.

Sob hipótese de completa aleatoriedade espacial (CAE) temos que $K(\mathrm{h}) = K_f(\mathrm{h})$. A escolha de f se dá de acordo com o tipo das marcas. Para marcas quantitativas utilizaremos a função f como sendo

$$f(m(s_k),m(s_l)) = \frac{m(sk),(sl)}{m^2} \tag{2.20}$$

em que $\widehat{m} = \frac{1}{n}\sum m(s_i)$ é a média das marcas.

Neste caso esta função é denominada de função de correlação de marcas ($\kappa_{mm}(\mathrm{h})$) e a sua interpretação é:

a) $\kappa_{mm}(\mathrm{h}) < 1$, sugere-se inibição mútua entre as marcas à uma distância de tamanho h;

b) $\kappa_{mm}(\mathrm{h}) > 1$ estimulação mútua para uma certa distância h;

c) No caso das marcas não serem correlacionadas (rotulagem aleatória) $\kappa_{mm}(\mathrm{h}) \equiv 1$.

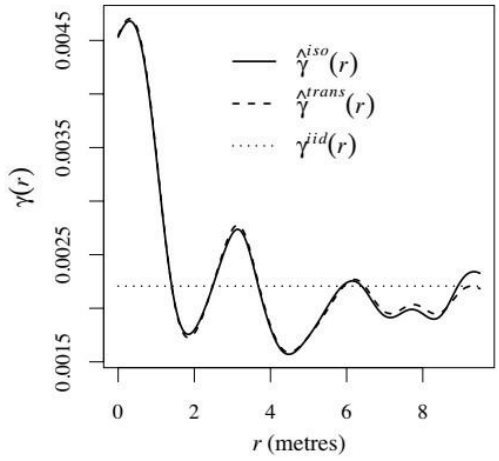
2.4.4 Variograma de marcas

Stoyan e Wälder (2000) mostram que o variograma de marcas pode ser construído de forma análoga ao tradicional variograma da geoestatística, como segue:

$$\gamma(h) = \frac{1}{2}\mathbb{E}[(m(u) - m(v))^2 | u, v \in X] \tag{2.21}$$

De maneira que u e v são posições arbitrarias com $\|u-v\| = h$. Sendo assim, $2\gamma(r)$ é o quadrado da esperança entra os valores das marcas de dois pontos separados por uma distância h. A Figura 2.3 apresenta uma ilustração do variograma gerado por 2.21.

Figura 3 - Variograma empírico para distribuição de variabilidade Abetos Noruegueses



Legenda: Variograma estimado com base no banco de dados levantado pelo estudo da distribuição dos Abetos Noruegueses.
Fonte: Baddeley et al. (2015)

A Figura 2.3 fornece uma pista que parece indicar uma variância maior quando a distância é menor entre os pontos, o que corrobora a natural dinâmica de dissipação da variabilidade conforme o aumento da distância entre os pontos.

2.5 Independência entre marcas e pontos no plano

Diversos autores tais como Diggle et al. (2000), Diggle (2003), Schlather (2002) e Guan (2006) têm mostrado a importância de analisar a independência entre pontos e marcas antes de

conduzir análises de processos pontuais marcados. Inclusive, Guan (2006) apresenta um método para testar a hipótese de independência entre pontos e marcas.

Se as marcas forem independentes, as análises estatísticas baseadas na teoria de processos pontuais não fazem sentido. Então, basta analisar a distribuição espacial da marca usando a teoria de campos aleatórios (geoestatística).

A partir do trabalho desenvolvido por Schlather (2002), Schlather et al. (2004) propuzeram utilizar as funcões E e V para analisar a hipótese de independência entre marcas e pontos em processos pontuais marcados. Essas funções também podem ser utilizadas para caracterizar as propriedades de segunda ordem de um processo pontual marcado no plano e representam, respectivamente e de maneira bem simples, a força e a interação entre as marcas e as suas localizações no plano contínuo. Essas funções são apresentadas a seguir.

2.5.1 As funções descritoras E e V

Considerando que o processo pontual marcado é estacionário e isotrópico e que as marcas seguem uma distribuição normal, é possível substituir o momento de segunda ordem pela variância condicional. Schlather (2002) definem as funções E(h) e V(h) que denotam, respectivamente, a esperança e a variância condicional de uma marca dado que existe um outro ponto do processo pontual em rede a uma distância h.

$$E(h) = E\{m(\mathrm{o})|\mathrm{o},\mathrm{h} \in \psi, ||\mathrm{h}|| = \mathrm{h}\} \tag{2.22}$$

e

$$V(h) = E[\{m(\mathrm{o}) - \mathrm{E}(\mathrm{h})\}^2|\mathrm{o},\mathrm{h} \in \psi, ||\mathrm{h}|| = \mathrm{h}] \tag{2.23}$$

Segundo Schlather (2002) e Schlather, Ribeiro e Diggle (2004) estas duas funções medem a força e a amplitude de interação entre as marcas e as suas localizações, respectivamente, e sob a suposição de que as marcas são independentes e identicamente distribuidas seguindo campo aleatório gaussiano, então E(h) e V(h) seriam constantes.

2.5.2 Estimadores para as funções descritoras E e V

Existem vários estimadores propostos para as funções E(h) e V(h) conforme pode ser visto em Cressie (1993), Baddeley, Møller e Waagepetersen (2000) e Schlather, Ribeiro e

Diggle (2004). Nesta tese vamos utilizar estimadores propostos por Schlather, Ribeiro e Diggle (2004) que são do tipo:

$$\left[\frac{1}{N_h}\sum_{|\|x_i-x_j\|-h|\leq\frac{\varepsilon}{2}} f\{m(x_i,x_j\}.\right] \tag{2.24}$$

A forma como é definida a função $f\{m(x_i,x_j\}$é que determina o estimador das funções variograma $\hat{\gamma}(h)$, esperança condicional $\hat{E}(h)$ e variância condicional $\hat{V}(h)$. Assim temos que:

a) $\hat{\gamma}(h)$ é obtido $f\left\{m(x_i,x_j\right\} = \frac{m_1-m_2)^2}{2}$;

b) $\hat{\gamma}(h)$ é obtido quando $f\{m(x_i,x_j\} = m_1$;

c) $\hat{\gamma}(h)$ é obtido quando $f\{m(x_i,x_j\} = m_1 - \hat{E}(h)$.

Nessas equações, m_1 e m_2 são os valores das marcas em duas posições genéricas, h é a distância que deve ser sempre maior ou igual a zero, ε é a largura de banda que deve ser sempre maior que zero e N_h é o número de pares de pontos (x_i,x_j) na qual ($\left|||x_i - x_j|| - h\right| \leq \varepsilon/2$).

Para evitar o efeito de bordas, podem ser utilizados nos estimadores apresentados acima os mesmos corretores adotados na estimação da função *K*, ou seja, do tipo isotrópico proposto por Ripley (1976) e outros descritos em Baddeley, Rubak e Turner (2015).

2.5.3 Independência entre marcas e pontos no plano

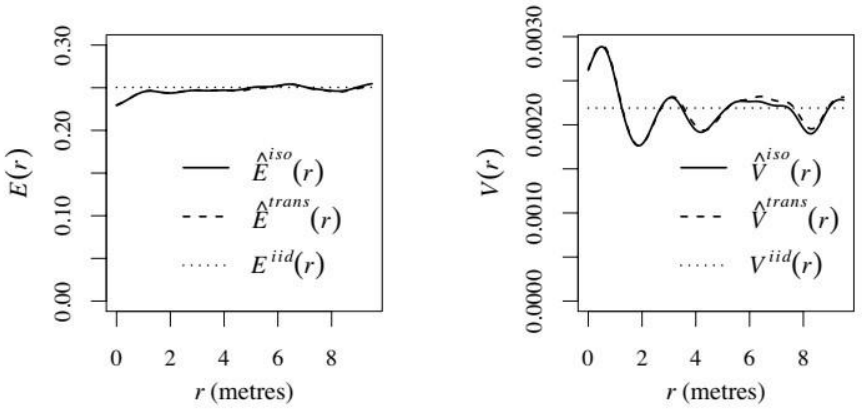
Uma vez obtidas as estimativas das funções E(h) e V(h) para a configuração pontual marcada observada, essas podem ser usadas para avaliar a hipótese nula de independência entre pontos e marcas. Para tal, basta verificar se essas estimativas são aproximadamente constantes para cada distância *h*.

Assim, uma maneira prática de verificar se a hipótese nula de independência entre pontos e marcas é verdadeira é fazer a inspeção visual do gráfico das estimativas das funções E(h) e V(h) contra as distâncias *h*. Quanto maior a discrepância entre as estimativas e uma reta constante, para qualquer distância *h*, maior é a dependência entre marcas e pontos.

A Figura 2.4 reproduz os resultados obtidos por Schlather, Ribeiro e Diggle (2004) e Badddeley et al. (2015) quando o método foi aplicado em um pequeno banco de dados de arvores de pinheiros noruegueses para verificar se o diâmetro a altura do peito dos pinheiros independe das localizações dos mesmos. Na Figura 2.4 a linha reta pontilhada representa independência entre marcas e pontos e as linhas tracejada e contínua representam as funções

estimadas obtidas utilizando dois corretores de bordas. Também adota-se a notação *r* para designar a distância entre os pinheiros apenas para manter a notação do trabalho original de Schlather, Ribeiro e Diggle (2004).

Figura 4 - Gráficos de E(r) e V(r) contra a distância r



Legenda: Teste de independência entre pontos e marcas na configuração das árvores de abetos.
Fonte: Adaptado de Sclather et al. (2004)

Com base na inspeção visual do gráfico esquerdo da Figura 2.4 é possível sugerir que pinheiros próximos (até um metro de distância) tendem a ter um diâmetro a altura do peito menor do que a média. O gráfico a direita sugere uma flutuação das estimativas da variância condicional em torno da reta constante. A interpretação visual dos dois gráficos sugere que as discrepâncias não são marginalmente discrepantes e, portanto, não se rejeita a hipótese nula. Assim, os diâmetros dos pinheiros são independentes de suas localizações e, consequentemente, as análises posteriores poderiam ser conduzidas utilizando métodos usuais da geoestatística. Entretanto, Schlather, Ribeiro e Diggle (2004) sugerem utilizar os mesmos métodos de Monte Carlo, apresentados na seção anterior, para obter a significância estatística.

2.6 Processos pontuais não marcados em redes

Um processo pontual não marcado em rede é uma situações particular e mais restrita do que o processos pontual não marcado no espaço. Neste caso, ao invés do processo estocástico gerar eventos em qualquer local dentro de uma região de estudo, ele gera os eventos que são localizados, exclusivamente, sobre uma rede, dando origem ao denominado processo pontual em rede *L*.

Seja x = $[x_1,...,x_n]$, $x_i \in L$, $n \geq 1$ uma realização do processo pontual X na rede L.

Seguindo, tem-se, por definição, que se X é um processo pontual em uma rede L e $X \subseteq L$, então n(X) é definido como sendo o número de pontos de X em *r*. Assim tem-se em que • *r*

= é o comprimento da janela tomada para estudo e chamada de *bandwidth* (largura de banda). Uma medida usada para definir o tamanho da janela de onde serão feitos os cálculos, ficando mais claro à medida que avançarmos nos detalhes metodológicos.

• N_X = número de pontos em $X \cap S$, ou seja, ocorrências dos eventos em estudo (BADDELEY; RUBAK; TURNER, 2015).

2.6.1 Definições básicas de uma rede

Para definir uma rede, primeiramente, deve-se definir se os eventos estão ou não posicionadas sobre uma rede. Sendo assim, ao longo do trabalho, nomeiam-se todos estimadores e funções como espacial, quando se tratar de um processo pontual espacial em R^2 e *linear* quando for abordado o processo pontual espacial em rede.

Inicia-se com algumas definições, conforme apresentadas por Fuhrmann e Helmke (2015):

Tomando uma seção linear $l \subset R^2$, está é chamada de segmento reta em um plano com limites u e v inferior e superior, respectivamente, e pode ser expressa formalmente por:

$$l = l_{u,v} = \{tu+(1-t)v : 0 \leq t \leq 1\},$$

para alguns pontos $u,v \in R^2, \forall\ u \neq v$ é tido como segmento de reta entre os pontos u e v, em que o módulo (sua amplitude) é chamado de distância euclidiana em R^2 dada por:

$$| l | = | l_{u,v} | = \| u-v \|$$

Logo, pode-se definir uma rede L em R^2 como uma união finita de segmentos de reta l sendo:

$$L = \cup^{n}_{i=1} li.$$

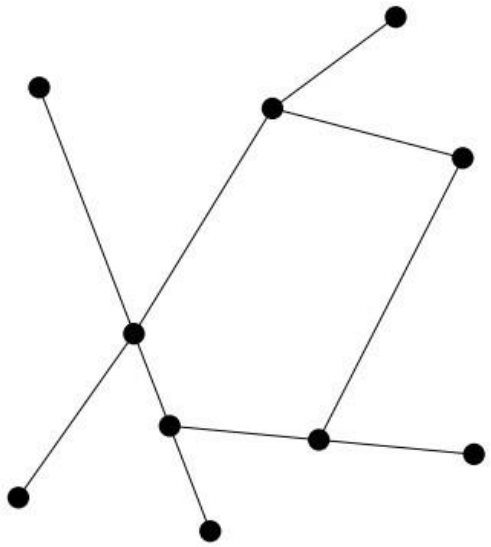
Sabendo que um segmento de reta $l_i = l_{ui,vi}$, tem-se que $l_i \cap l_j = 0, \forall i \neq j$ e o comprimento total de L pode ser escrito como:

$$| L | = \sum_{i=1}^{n} | l_i |$$

Seja L uma rede. Então um ponto $v \in L$ é chamado de vértice de interseção se $v = l_i \cap l_j$ para algum $i, j, i \neq j$ e este será chamado de vértice livre se $l_i = l_{x,v}$ (ou $l_{v,x}$), para algum i,x, e $l_i \cap l_j \neq v, \forall i \neq j$. Será denominado de *interseção* a união dos vértices de interseção e os vértices livres.

Tem-se então, ao fim dos conceitos definidos acima, a seguinte representação geométrica para uma rede:

Figura 5 - Rede formada



Legenda: (•) são as interseções e (-) são segmentos de reta Diagrama ilustrativo de uma rede em um plano.
Fonte: Adaptado de Baddeley et al. (2015)

Uma vez constatado que o processo pontual espacial constitui uma rede, o comportamento do mesmo pode ser caracterizado em termos de efeitos de primeira e segunda ordem.

Para tal, é necessário definir algumas medidas que são apresentadas a seguir.

2.6.2 Métricas de distância em rede

Segundo Borruso (2008) podemos descrever uma distância, a partir do ponto do *grid* escolhido, ou entre pontos, por três maneiras:

- **Distância pareada em uma rede (caminho mais próximo)**

Essa abordagem diz que para um padrão de n eventos em uma região A, exitem $\frac{1}{2}n(n-1)$ distâncias pareadas entre o evento x_i e x_j, logo definimos a função de densidade acumulada por:

$$H_l = \{\mathrm{P}d_{ij} \leq r\}$$

Seguindo podemos escrever a distância do caminho mais próximo como:

$$\widehat{H}\ l(r) = \frac{1}{2} - n(n-1)\}^{-1} = \frac{2}{n(n-1)} \sum_{i \neq j} 1\{|| x_{i-x_j} || l \leq r\}$$

onde: x_i e x_j são eventos observados no processo pontual na rede, $1\{.\}$ uma função indicadora pontual e $\| x_i - x_j \| l$ é distância do caminho mais próximo entre x_i e x_j, que por Baddeley,

Rubak e Turner (2015) é negativamente viesado.

- **Distância do vizinho mais próximo em uma rede**

Definimos a função de densidade acumulada para essa abordagem por:

$$G_l(r) = \mathrm{P}\{d_L(u, \mathrm{X}\{u\}) \leq r \mid u \in \mathrm{X},$$

em que u é um evento qualquer e $d_L(u, \mathrm{X}\{u\})$ é a menor distância de u ao seu vizinho mais próximo, no processo pontual X na rede L.

Um estimador viesado pra $G(r)$, tem função de densidade empírica da forma:

$$\hat{G}_l(r) = n^{-1} \sum_i 1\{d_i \leq r\}.$$

- **Distância do espaço vazio em uma rede**

A função de densidade acumulada é definida nesse caso como:

$$F_l(r) = \mathrm{P}\{d_L(u, \mathrm{X}) \leq r,$$

em que u é uma localização de referência arbitrária.

A distribuição empírica observada, em um *grid* de m pontos amostrais, $u_j = j = 1,...,m$ em A, é definida como:

$$\hat{F}_l(r) = m^{-1} \sum_j 1\{d_L(u, \mathrm{X}) \leq r\},$$

Segundo Diggle (2003), este estimador é positivamente viesado e, portanto, este sugere usar o valor de m como sendo um *grid* igual ao número de eventos observados na configuração pontual. Uma abordagem mais moderna feita por Rakshit, Nair e Baddeley (2017) constrói um estimador para qualquer tipo de métrica de distância em uma rede linear.

2.6.3 Efeitos de primeira ordem em um processo pontual em rede

Os efeitos de primeira ordem, também conhecidos como globais ou de larga escala, descrevem quando o valor esperado (média) varia dentro da rede. O objetivo da análise está em estimar a intensidade do processo, ou seja, o número de eventos por unidade de comprimento.

Os efeitos primeira ordem do processo pontual em rede podem ser definidos pela função intensidade dada por

$$\lambda(u) = \lim_{|du|\to 0} \left\{ \frac{\mathbb{E}[\mathbf{N_X}(du)]}{du} \right\}, \tag{2.25}$$

em que *du* define um pequeno intervalo em torno de *u*, em que o número esperado de pontos por unidade de comprimento é constante e independentemente da localização. Assim, quando a intensidade é igual a algum λ constante diz-se que é a intensidade de primeira ordem é estacionária (homogênea), com função de intensidade $\lambda(u) = \lambda$, em que λ pode ser interpretada como o número esperado de pontos por unidade de comprimento.

Com isso pode-se definir que sendo *L* uma rede e X um processo pontual em *L*, então pode-se dizer que um X é estacionário de primeira ordem se para qualquer $S \subseteq L$, ocorre que

$$E[\mathbf{N_X}(S)] = \lambda|S| \tag{2.26}$$

em que λ é a intensidade de x.

Um estimador para intensidade de um processo pontual espacial em rede estacionário

(intensidade constante em toda a rede) pode ser obtido, segundo Baddeley, Rubak e Turner (2015), por:

$$\hat{\lambda} = \frac{\mathbf{N_X}(S)}{|L|} \tag{2.27}$$

em que N_X é o número de eventos localizados na rede de comprimento total |L|.

Caso o processo pontual espacial em rede não seja estacionário (a intensidade não é constante em toda a rede), pode-se usar um estimador não paramétrico da intensidade do tipo kernel. Esse estimador realiza uma contagem de todos os eventos dentro de uma distância de influência, ponderando-os pela distância de cada evento à localização arbitrária dentro da rede.

Timothée et al. (2010) propõe uma variação da função kernel quadrática espacial para ser utilizada na estimação da intensidade do processo pontual espacial em rede, definida como kernel de redes (NetKDE), descrita como

$$k_{net}(t_{net,i}) = \begin{cases} \frac{1}{3\pi}\ (1 - t_{net,i}^2), & se\ t_{net,i}^2 \ \leq\ 1; \\ 0 & \text{caso contrário} \end{cases} \tag{2.28}$$

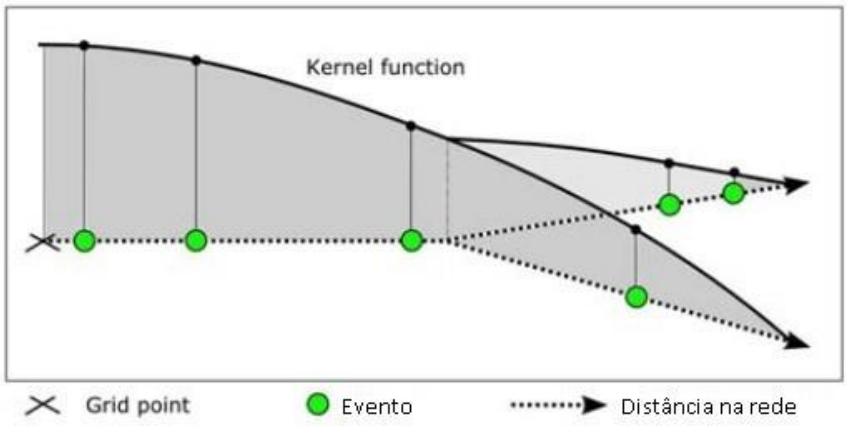
em que $t_{net,i} = d_{net,ij}/h$, h é a largura de banda e $d_{net,ij}$ é a distância do ponto *grid* x_j ao evento observado n_i medido sobre a rede. Pode-se então obter o estimador da intensidade kernel linear na localização x_j dada por

$$Net\ KDE\ (x_j)\ =\ \frac{1}{nh^2}\ \sum_{i\,=\,1}^{n} K_{net}\ (t_{net,i}), \qquad (2.29)$$

em que n é o número de eventos na largura de banda h.

A figura 2.6 ilustra geometricamente o comportamento do estimador da intensidade kernel linear para uma rede em vista em perceptiva lateral, em que os pontos verdes são os eventos observados e a altura até a linha em negrito é a intensidade estimada para cada localização. Nesta figura, pode-se observar de forma mais clara tanto o uso da largura de banda (linha tracejada) quanto a janela de cálculo tomada de um ponto sorteado ao acaso no grid (área sombreada). O parâmetro mais importante é a largura de banda. Apesar de não existirem critérios estatísticos para a escolha dessa largura, Baddeley, Rubak e Turner (2015) sugere uma largura de banda com 10 unidades de comprimento.

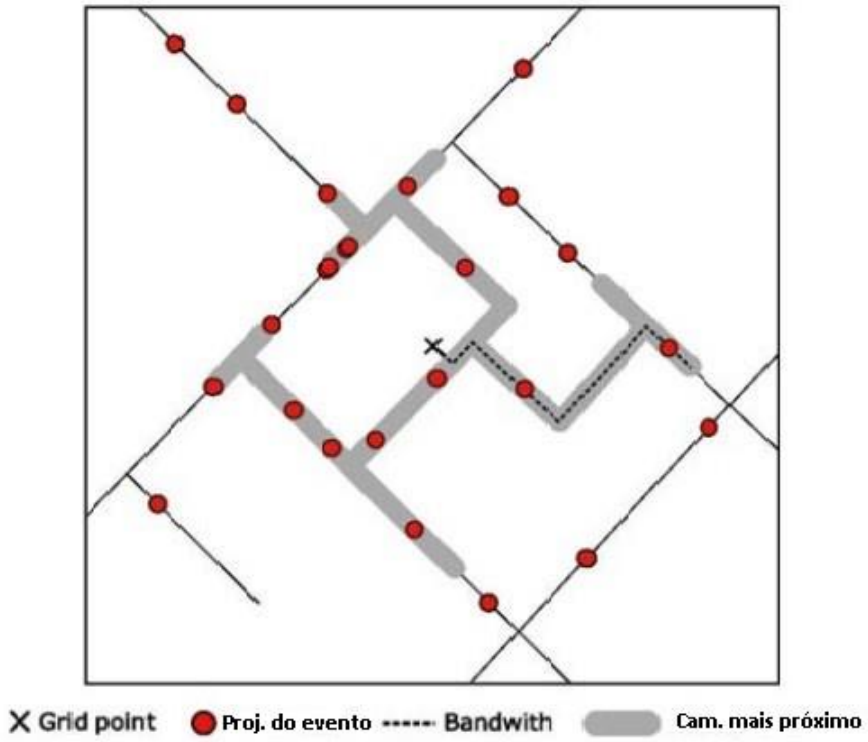
Figura 6 - Estimador kernel linear



Legenda: Intensidade em uma rede onde a altura dos pontos na indica sua intensidade.
Fonte: Adaptado de Timothée (2010)

Um exemplo do uso do estimador kernel em rede linear pode ser visto na figura a seguir:

Figura 7 - Ilustração kernel linear



Legenda: Esquema do ajuste feito nas distância de eventos em uma rede linear.
Fonte: Adaptado de Timothée (2010)

2.6.4 Efeitos de segunda ordem em um processo pontual em rede

Os efeitos de segunda ordem, ou de pequena escala, estão relacionados ao grau de interação entre os eventos localizados na rede linear. Para caracterizar os efeitos de segunda ordem pode-se definir a intensidade do segundo momento (ou densidade do produto) $\lambda_2(u,v)$.

Heuristicamente, segundo Baddeley, Rubak e Turner (2015), dado dois eventos distintos (u,v) na rede, considera-se um segmento muito curto da rede em torno da localização u de comprimento d_1u. De forma semelhante, considera-e um pequeno segmento de rede em torno de v com comprimento d_1v. A probabilidade de $p(u,v)$ que ambos esses pequenos segmentos contenham, pelo menos, um ponto aleatório é dada por $p(u,v) = \lambda_2(u,v)d_1ud_1v$. Então, para quaisquer segmentos de duas linhas $A,B \subset L$ que sejam disjuntos, temos que:

$$E[n(\boldsymbol{X}\cap\boldsymbol{A})n(X\cap B)] = \lambda^{-1} \int A \int B \, \lambda^2 \, (u,v) \, d^1 \, vd^1 \, u \qquad (2.30)$$

Pode-se observar que o processo pontual em rede, com função de intensidade $\lambda(u)$, tem intensidade de segunda ordem como sendo $\lambda_2(u,v) = \lambda(u)\lambda(v)$.

Uma das principais estatísticas descritoras na análise de processos pontuais em rede é a função *K* que é baseada na densidade produto. Sua principal vantagem é permitir a detecção do padrão espacial em diferentes escalas de distâncias simultaneamente. Okabe e Yamada (2001) utilizaram a função *K* de Ripley (1976) como base para propor uma nova função para analisar os padrões espaciais e interações quando o processo ocorre em uma configuração de redes lineares. Um estimador da função *K*-linear (K_{OY}) é obtido diretamente do estimador desta função para o caso planar sendo dado por:

$$\widehat{K}_0 Y\ (\mathrm{r}) = \frac{1}{\lambda^2 \mid L \mid} \sum_{\mathrm{i}} \sum_{\mathrm{j} \neq 1} 1\ (\{d_L\ (x_1, x_j)\} \leq\ r \qquad (2.31)$$

em que *I*(.) representa a função indicadora.

Intuitivamente, sob a suposição de CAE, a função K-linear teórica é dada por $K_{OY}(r)= r$ e $d_L(x_i,x_j)$ valores da distância na rede dos pontos x_i e x_j Assim, a função K-linear pode ser utilizada para testar a hipótese nula CAE utilizando os critérios apresentados na Tabela ??, a seguir.

Quadro 1- Critérios

Tipo de Processo	$\hat{K}_{OY}(r)$
Aleatório	$= r$
Regularidade	$< r$
Agrupamento	$> r$

Fonte: Elaborado pelo autor (2023)

2.7 Processo pontual marcado em redes lineares

Um processo pontual marcado em rede linear pode ser definido da mesma forma que um processo pontual marcado no espaço conforme apresentado anteriormente, ou seja, um processo estocástico $\Psi = [s_i;m(s_i)]$ que gera os eventos (s_i) e as marcas associadas aos eventos $m(s_i)$. Uma realização de Ψ é um conjunto de posições junto às marcas associadas dispostas exclusivamente sob uma rede, e essas marcas podem ser quantitativas ou qualitativas.

Um processo pontual marcado em rede também pode ser descrito em termos de propriedades de primeira e segunda ordem.

2.7.1 Propriedades de primeira e segunda ordem

Assim como ocorre no processo pontual não marcado em rede, no caso estacionário, são utilizados dois tipos diferentes de características de primeira ordem: a que diz respeito aos eventos, isto é, a intensidade λ, e a outra que descreve as marcas (ILLIAN et al., 2008). Os estimadores globais e locais para as propriedades de primeira ordem para processos pontuais marcados são obtidos a partir de extensões diretas dos estimadores de processos pontuais não marcados em rede.

As características de segunda ordem são úteis para caracterizar não somente a variabilidade da distribuição dos eventos, mas também a variabilidade das marcas e, ainda, descrever correlações entre marcas e pontos (WÄLDER; STOYAN, 1996). Estimadores das propriedades de segunda ordem, em geral, também dependem do tipo de marca.

A principal função utilizada para analisar as propriedades de segunda ordem em processos pontuais marcados em redes é a funções *K*.

2.7.2 A função *K* ponderada por marcas em rede

A função *K* ponderada por marcas em rede é uma ferramenta útil para caracterizar as propriedades de segunda ordem. No caso de uma processo pontual marcado de acordo com Penttinen, Stoyan e Henttonen (1992), esta função é uma extensão da função *K* de Ripley, na qual para cada par de eventos existe uma ponderação por uma função teste $f(m_1,m_2)$ de suas respectivas marcas.

Um estimador não-viesado, corrigido e proposto por Ang, Baddeley e Nair (2012) baseado em uma maneira análoga a K_f é dado por:

$$\widehat{K}_f(\mathbf{r}) = \frac{1}{\lambda E\,[\,f\,(M,M')]}\,\mathrm{E}\left[\sum_{x_j \in X}^{n} f\,(m(u)\,), m(xj))1(0 \;<\; ||u - xi|| \;\leq\; \mathrm{r})|u \;\in\; X\right]$$

(2.32)

em que λ é a intensidade de primeira ordem dos pontos marcados na rede de estudo L; f é uma função que atribui um peso a um par dos pontos u e x_j dependendo do tipo de suas marcas $m(u)$ e $m(x_j)$; 1(.) é a função que é igual a 1 se o argumento for verdadeiro e 0 caso contrário.

Pode ser visto em Wälder e Stoyan (1996) que sob hipótese de complete aleatoriedade espacial (CAE) temos que $K(r) = K_f(r)$.

A escolha de f se dá de acordo com o tipo das marcas. Para marcas quantitativas e não negativas utilizaremos a função f como sendo:

$$f(m(u),m(x_j)) = \frac{m(u)m(x_j)}{\hat{m}^2}, \tag{2.33}$$

em que $\hat{m} = \frac{1}{n}\sum m(x_j)$ é a média das marcas, com $f(m, m') = \frac{1}{2}(m - m')^2$.

Nesse caso, esta função é denominada de função de correlação de marcas ($\kappa_{mm}(r)$).

No caso de marcas quantitativas que seguem uma distribuição gaussiana, conforme vamos mostrar no próximo capítulo desta tese, também é possível caracterizar as propriedades de segunda ordem do processo pontual marcado em rede através das funções $E(h)$ e $V(h)$ que denotam, respectivamente, a esperança e a variância condicional de uma marca dado que existe um outro ponto do processo pontual em rede a uma distância h.

3 INDEPENDÊNCIA ENTRE MARCAS E PONTOS EM PROCESSOS PONTUAIS MARCADOS EM REDES LINEARES

Neste capítulo apresenta-se o desenvolvimento teórico do método para a análise da independência entre marcas e pontos em processos pontuais marcados em redes lineares.

Conforme visto anteriormente, um processo pontual marcado em redes lineares pode ser definido da mesma forma que um processo pontual marcado no espaço bidimensional, ou seja, um processo estocástico $\Psi = [s_i;m(s_i)]$ que gera os eventos (s_i) e as marcas associadas aos eventos $m(s_i)$ em reses lineares. Assim, uma realização de Ψ é um conjunto de posições que apresentam às marcas associadas à essas posições que estão dispostas exclusivamente sob uma rede. As marcas podem ser quantitativas ou qualitativas. Nesta tese utiliza-se apenas marcas quantitativas marginalmente gaussianas.

Um processo pontual marcado em rede linear também pode ser descrito em termos das propriedades de primeira e segunda ordem. No caso de marcas quantitativas que seguem uma distribuição normal, também é possível fazer a caracterização das propriedades de segunda ordem através das funções E(r) e V(r) que denotam, respectivamente, a esperança e a variância condicional de uma marca, dado que existe um outro ponto do processo pontual marcado na rede linear a uma distância menor ou igual a *r*.

3.1 Estimadores de E(r) e V(r) para redes lineares

Seguindo a mesma definição utilizada por Schlather (2001b) e Schlather, Ribeiro e Diggle (2004) para processos pontuais no plano, defini-se a esperança condicional de uma marca, dado que existe um outro ponto do processo pontual na rede a uma distância menor ou igual a *r* como sendo

$$E(r) = E\{m(0)|0,\mathrm{x} \in \mathrm{N},||\mathrm{x}|| = \mathrm{r}\} \tag{3.1}$$

Sob a suposição de que as marcas são independentes e identicamente distribuídas seguindo campo aleatório gaussiano, então E(r) seria constante.

Estimadores para $E(r)$ podem ser obtidos a partir dos estimadores propostos por Schlather, Ribeiro e Diggle (2004) apresentados no capítulo anterior. Também podem ser obtidos a partir da função *K* para processos pontuais com marcas qualitativas (bidimensionais) no espaço conforme pode ser visto nos trabalhos de Diggle (2003) e Illian et al. (2008). Gelfand et al. (2010) e Okabe, Okunuki e Shiode (2006) fazem propostas de estimadores para marcas

quantitativas usando a função *K* marcada. Neste trabalho, utiliza-se uma adaptação do estimador proposto por Ang, Baddeley e Nair (2012) para redes lineares, ou seja,

$$\hat{E}_{net}(r) = \hat{K}_{mm}(\mathbf{r}) = \frac{|L|}{n}\sum_{x_k \in X_i}^{n} \sum_{x_l \in X_j}^{n} \frac{1\{d_L(x_k,x_1) \leq r\}}{m\{x_k, d_L(x_k,x_l)\}}. \quad (3.2)$$

O estimador dado pela equação em 3.2 não é fácil de ser implementado computacionalmente. Assim, neste trabalho utiliza-se uma adaptação do estimador proposto por Guan e Afshartous (2007), resultando no estimador da seguinte forma:

$$\hat{E}_{net}(r) = \hat{K}_{mm}(\mathbf{r}) = \hat{\mu} = \frac{\sum \sum w[(r - \|x_2 - x_1\|)/h] \times m(x_1)}{\sum \sum w[(r - \|x_2 - x_1\|)/h]}. \quad (3.3)$$

em que *r* é um vetor de distâncias positivas que a função será obtida (estimada), $m(x_1)$ é um valor real de uma marca de um ponto (evento) localizado no ponto genérico x_1, ($\|x_1 - x_2\|$) é a distância entre dois pontos quaisquer, $w()$ é uma função de densidade simétrica do tipo kernel e *h* é a largura de banda que controla a quantidade de suavização do estimador.

O estimador 3.3 é não viesado para resultados de $r < R$, ou seja, quando analisado na rede de comprimento *R*, a sua interpretação é idêntica à aquela utilizada em configurações pontuais no plano (ANG; BADDELEY; NAIR, 2012; JAMMALAMADAKA et al., 2013), ou seja:

a) $\hat{K}_{mm}(\mathbf{r}) < 1$, sugere-se inibição mútua entre as marcas de dois eventos (pontos) x_1 e x_2 localizados a uma distância de tamanho r;

b) $\hat{K}_{mm}(\mathbf{r}) > 1$, sugere uma atração mútua entre as marcas de dois eventos (pontos) x_1 e x_2 localizados a uma certa distância r;

c) $\hat{K}_{mm}(\mathbf{r}) \equiv 1$ sugere independência entre as marcas e os dois pontos x_1 e x_2 separados por uma certa distância r.

Seguindo a mesma definição utilizada por Schlather (2002) para processos pontuais marcados no espaço contínuo, defini-se a variância condicional de uma marca, dado que existe um outro ponto do processo pontual marcado em rede linear a uma distância menor ou igual a *r* como sendo

$$V(r) = E[\{m(\mathbf{o}) - \mathbf{E}(\mathbf{r})\}^2 \mid \mathbf{o}, \mathbf{r} \in \mathbf{N}, \| \mathbf{x} \| = \mathbf{r}] \quad (3.4)$$

Sob a suposição de que as marcas são independentes e identicamente distribuidas seguindo um campo aleatório gaussiano, então V(r) seria constante.

Pode-se pensar em vários estimadores para 3.4. Para o caso de processos pontuais marcados no espaço, Schlather, Ribeiro e Diggle (2004) sugerem usar o estimador do semivariograma marcado que é definido como

$$\gamma(r) = \frac{1}{2}\mathbb{E}\left[\left(m(x_1) - m(x_2)\right)^2 \mid x_1, x_2 \in X\right] \tag{3.5}$$

em que $m(x_1)$ e $m(x_2)$ são as marcas de dois eventos x_1 e x_2 separados por uma distância r.

Cressie (1993) e Stoyan e Wälder (2000) propuzeram estimadores não paramétricos para 3.5. Entretanto, como observam Stoyan e Wälder (2000), esses estimadores não podem ser interpretados como os estimadores de variogramas usuais da geoestátistica, pois podem exibir propriedades que são impossíveis ou implausíveis no contexto de processos pontuais marcados.

De qualquer maneira, neste trabalho utiliza-se uma adaptação para redes lineares do estimador proposto por Schlather, Ribeiro e Diggle (2004) que denominaremos de $V_{net}(r)$ (veja capítulo anterior). De maneira análoga ao funcionamento do estimadores para os processos pontuais no R^2, proposto por Schlather, Ribeiro e Diggle (2004), o estimador $V_{net}(r)$ para redes lineares continua sintetizar a esperança quadrática para todas as observações x_1,x_2 observadas até uma distância r dentro de uma rede linear de comprimento R.

Assim, um resultado típico de estimativas obtidas pelo estimador $V_{net}(r)$, quando existe independência entre as marcas, é uma curva que flutua em torno de uma reta constante.

3.2 Método de Monte Carlo para testar a independência entre pontos e marcas em uma rede linear

Para a análise da independência entre pontos e marcas na rede linear, pode-se comparar as funções estimadas $E_{net}(r)$ e $V_{net}(r)$ com as funções teóricas de $E(r)$ e $V(r)$ sob a suposição da hipótese nula de independência entre pontos e marcas para um conjunto de distâncias r.

Para avaliar se as diferenças entre as funções estimadas e teóricas (retas constantes) são estatisticamente significantes, pode-se construir envelopes de simulação Monte Carlo sob a hipótese nula de independência entre marcas e pontos. O método é análogo ao caso de processos pontuais no plano, conforme pode ser visto em Schlather, Ribeiro e Diggle (2004) e Baddeley, Rubak e Turner (2015). Observe que sempre será necessário construir dois gráficos: um para

esperança condicional e outro para a variância condicional. A análise é conduzida utilizando os seguintes passos:

a) calcula-se os valores estimados de E_{net} (e V_{net}) para a rede linear observada para um conjunto de distâncias r. Denomina-se esses valores de $\hat{J}$;

b) calcula-se $\hat{J}_i$: $i = 2,3,...,s$ para cada $s-1$ simulações independentes e identicamente distribuídas de configurações de redes lineares, com o mesmo número de pontos da rede observada, sob a hipótese de independência entre marcas e pontos para o mesmo conjunto de distâncias r. Isso é feito embaralhando as marcas, mantendo-se inalteradas as posições dos pontos (eventos);

c) São construídos os envelopes de simulações superior e inferior para um conjunto de distâncias r a partir dos valores máximos e mínimos de $\hat{J}_i$, definidos como:

$$U(r) = \max\{\hat{J}_i\}, i = 2,...,s. \; L(r) = \min\{\hat{J}_i\}, i = 2,...,s; \tag{3.6}$$

d) Constrói-se um gráfico, em que os valores de $\hat{J}$, $U(r)$ e $L(r)$ são colocados no eixo das coordenadas e as distâncias r no eixo das abcissas.

Se para qualquer distância r, a curva de $\hat{J}$ estiver dentro dos envelopes de simulação, aceita-se a hipótese nula de independência entre marcas e pontos no processo pontual marcado na rede linear ao nível de significância $2/(s+1)$. Observe que se a linha de $\hat{J}$ estiver fora dos envelopes para, pelo menos, uma das distâncias r para qualquer uma das medidas (E_{net} ou V_{net}) será suficiente para rejeitar a hipótese nula.

4 MATERIAIS E MÉTODOS

Neste capítulo é descrito todo o ferramental teórico, levantamento das observações e ajuste do banco de dados disponibilizado para a aplicação prática da metodologia.

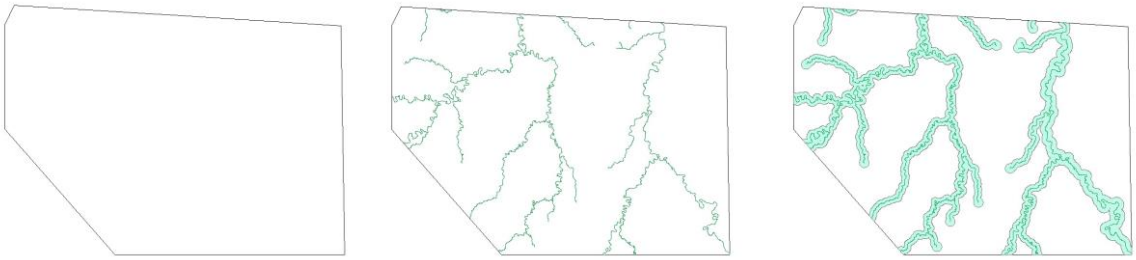
4.1 Materiais

Os dados do presente estudo foram obtidos a partir do censo florestal realizado na Fazenda Canary, localizada no município de Bujari, no estado do Acre. A propriedade detém uma área total de 15.810,14 ha^{-1}, sendo 8.446,80 ha^{-1} de reserva legal, deixando uma área efetiva de manejo florestal de 4465,89 ha^{-1}. Entretanto, neste trabalho será utilizada uma área menor de proteção permanente (APP) pois nesta área que ocorre uma maior concentração de redes fluviais. A vegetação existente é constituída de Floresta Amazônica nativa com bambu, palmeira e outras espécies de árvores como a *Hura crepitans* (ACRE, 2006).

Utiliza-se para o presente estudo uma área de 785 ha^{-1} com uma APP de 160,40 ha^{-1}. Os rios dentro da APP que têm 27.236,12 metros de cumprimento e serão tratados como os segmentos de retas da rede. Como descrito na Figura 4.1 a seguir:

Figura 8 - (a) Área (b) Rios (c) APP

(a) (b) (c)



Legenda: Composição do *shape* para estudo com a janela, rede de rios e área da mata ciliar.
Fonte: Elaborado pelo próprio autor (2023)

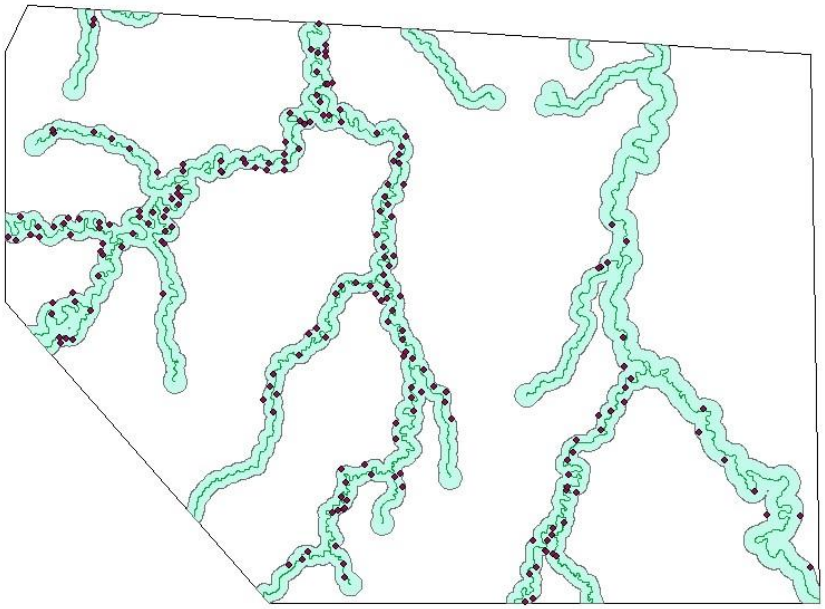
A coleta dos dados iniciou-se com um inventário florestal realizado na área efetiva de manejo, onde foram mensuradas as 56 espécies de árvores comerciais que apresentavam um

diâmetro mínimo de corte (DMC) > 50cm. Foram registradas as informações de diâmetro a altura do peito (DAP), que é tomado a 1,30m de altura verticalmente medidos, utilizando uma fita métrica para tirar a medida da seção circular. Também foi utilizada uma trena a laser para medir a altura comercial. Foram realizados ainda cálculos do volume de madeira, identificação botânica e classes de qualidade do fuste. Na sequência, esta foi marcada com uma placa de alumínio contendo informações relacionadas a seu número de registro. As árvores foram georreferenciadas em coordenadas UTM, utilizando um aparelho GPS. Posteriormente, as informações geradas foram convertidas em um banco de dados espaciais. O inventário florestal que deu origem aos dados que serão utilizados neste trabalho está descrito em detalhes no trabalho de Junior et al. (2014).

Neste trabalho foram utilizadas apenas árvores da espécie *Hura crepitans*, também conhecida como árvore dinamite, açacu ou árvore-do-diabo. Esta espécie é nativa da América do Sul e Central, podendo atingir alturas de 60 metros. No Brasil, sua maior incidência está na Floresta Amazônica. Os nomes populares de açacu e árvore-do-diabo é devido ao fato que seu caule é todo coberto por espinhos de 1 a 2 cm de comprimento, que secretam uma seiva venenosa, capaz de provocar queimaduras graves em contato com a pele. Já o apelido de árvore dinamite surgiu porque seus frutos, que lembram pequenas abóboras, explodem com um grande estrondo, espalhando as sementes em um raio de até 100 m. O som pode ser comparado ao de uma pequena explosão e a força é capaz de ferir pessoas ou animais próximos. A espécie desperta interesse pois a infusão das flores masculinas (espigas) ou as brácteas frescas, aplica-se nos furúnculos. As folhas trituradas com água aplicam-se nos reumatismos. O suco leitoso (látex) é um poderoso antihelmíntico e é usado por pescadores locais para envenenar peixes e para ajudar na caça. A sua madeira macia e sem cheiro é muito apreciada comercialmente para a fabricação de mobiliário e peças artesanais (JUNIOR et al., 2014)).

Neste trabalho foram utilizadas 201 árvores da espécie *Hura crepitans* localizadas dentro da APP como ilustrado na Figura 4.2, a seguir:

Figura 9 - Área de estudo



Shape completo para estudo com as observações(árvores) em suas coordenadas ajustadas.
Fonte: Elaborado pelo próprio autor (2023)

Como a área de estudo não apresenta uma rede completamente conectada, esta área foi dividida em duas partes (oeste e leste). Para uma maior rapidez no trabalho computacional, as análises foram realizadas exclusivamente na parte leste. Essa separação das áreas acarretou a perda de duas observações.

4.2 Métodos

Primeiramente, é importante observar que a malha de rios da região de estudo não pode ser considerada uma rede linear pois as árvores não estão posicionadas dentro do rio mas nas margens. Assim, a rede foi linearizada em toda sua extensão através de uma interpolação de forma ortogonal para que todos os pontos (árvores) ficassem localizados exatamente sobre a rede (rios).

A Figura 4.2 mostra que a APP apresenta duas redes (Leste e Oeste). Nesta tese foi utilizada apenas a rede linear Leste com comprimento de 740,8 m e contendo 40 pontos (árvores). É nos dados desta rede linear que serão conduzidas as análises de primeira e segunda

ordem da configuração pontual marcada, assim como a aplicação do Método de Monte Carlo para testar a hipótese de independência entre pontos e marcas.

Para a análise das propriedades de primeira ordem local do processo pontual não marcado foi utilizado o estimador não paramétrico com função kernel quadrática e uma largura de banda igual a 10 metros. A visualização das intensidades locais estimadas foi feita através de mapas de calor e de círculos.

Para análise das propriedades de segunda ordem do processo pontual não marcado foi calculada a função K até uma distância máxima de 350 metros. Os envelopes de confiança foram obtidos a partir de 99 simulações de Monte-Carlo sob a hipótese nula de completa aleatoriedade espacial.

O método proposto de Monte Carlo para testar a independência entre pontos e marcas em redes lineares foi avaliado através do cálculo da taxa empírica de erro do tipo I (rejeitar a hipótese nula (H_0) de independência entre marcas e pontos quando ela é verdadeira) utilizando simulações de Monte Carlo. Foram avaliados dois cenários: uma rede com 50 eventos (pontos) e outra rede com 100 eventos localizados em redes lineares, cuja configuração é a mesma da rede de rios descrita no capítulo anterior. Para cada um desses cenários foram simuladas 500 configurações (realizações) sob a hipótese nula (H_0) de independência entre marcas e pontos. Para cada uma dessas configurações foi aplicado o método proposto no capítulo anterior, ou seja, foram obtidas as funções V_{net} e E_{net} estimadas e verificado o quanto elas se afastavam das hipóteses $E(r)$ = constante e $V(r)$ = constante para algum valor de r. Envelopes de confiança obtidos a partir de 99 simulações, sob a hipótese nula, foram aplicados a cada realização. Finalmente, foi feita a contagem do número de vezes que as funções estimadas $E(r)$ e $V(r)$ ficaram fora dos envelopes de simulação. As taxas de erro obtidas foram comparadas entre si e como valor nominal do nível de significância adotado $\alpha = 0{,}05$).

O método proposto de Monte Carlo para testar a independência entre pontos e marcas em redes lineares foi aplicado na rede linear Leste de rios da Fazenda Canary com comprimento de $740{,}8m$ e contendo 40 pontos (árvores). As funções E(r) e V(r) foram estimadas até uma distância máxima de 25 metros. Os envelopes de confiança foram obtidos a partir de 99 simulações de Monte-Carlo sob a hipótese nula de independência entre pontos e marcas.

4.3 Softwares

No presente estudo, o *software* QGIS (QGIS DEVELOPMENT TEAM, 2021) foi utilizado para tratamento e filtragem da base de dados, ou seja, georreferenciamento dos pontos, atribuição das marcas aos pontos, construção do *shape* da rede de rios e as delimitações da área de estudo (propriedade). O *software* SANET (OKABE et al., 2012) foi utilizada para os cálculos das intensidades de primeira ordem (kernel) e segunda ordem (função k ajustada para redes conforme proposta apresentada por Okabe e Yamada (2001) e Ang et al. (2012). A análise de independência entre marcas e pontos em redes lineares foi conduzida utlizando funções desenvolvidas especificamente para este fim no R (R CORE TEAM, 2022). Para tal, foi necessário adaptar funções utilizadas na análise de configurações pontuais no plano e disponibilizadas no Spatstat (BADDELEY; TURNER, 2005; BADDELEY et al. 2015; BIVAND et al. 2022).

Essas funções estão disponíveis no Apêndice A.

5 RESULTADOS E DISCUSSÃO

A caracterização de configurações pontuais no plano pode ser conduzida a partir da análise das propriedades de primeira e segunda ordem do processo estocástico que gerou a configuração pontual. A teoria para esse tipo de análise é bem desenvolvida conforme pode ser visto em Diggle (2003) e Illian et al. (2008). Além disso, bibliotecas como o Spatstat (BADDELEY et al., 2005, 2015) permitem conduzir diversas análises (ex. exploratórias e modelagem) de configurações pontuais no plano de forma rápida e segura.

Deve-se observar que existem outras situações de configurações pontuais mais específicas, como é o caso de árvores que são localizadas nas margens de rios ou o caso de acidentes de trânsitos que ocorrem em ruas e estradas. Nesses casos, os dados somente podem estar localizados em locais específicos dentro de uma região contínua de estudo. Esses dados são denominados de configurações pontuais em redes lineares.

A caracterização de configurações pontuais em redes lineares (marcadas ou não) também pode ser conduzida a partir da análise das propriedades de primeira e segunda ordem do processo estocástico que gerou a configuração pontual na rede linear (BADDELEY et al., 2015). A teoria para esse tipo de análise está em desenvolvimento, mas já existem funções disponíveis para esse tipo de análise na biblioteca (Spatstat) do *R* (BADDELEY et al., 2005, 2015). São essas funções que serão utilizadas para as análises das propriedades de primeira e segunda ordem das localizações de árvores da espécie *Hura crepitans* sobre uma rede de rios localizada na Fazenda Canário.

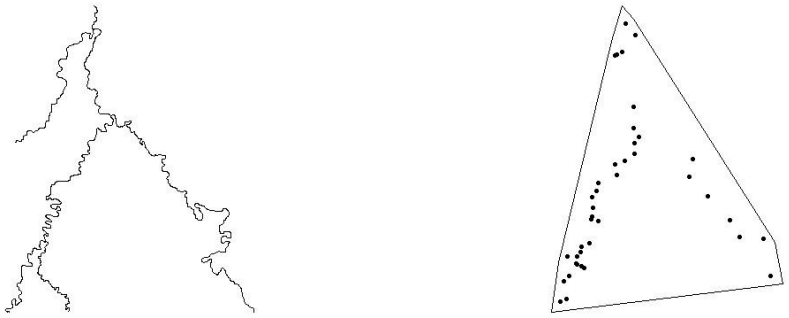
5.1 Análise de primeira ordem

A Figura 5.1 apresenta a rede linear de rios localizada na parte Leste da Fazenda Canário (a), assim como a posição das árvores (pontos) da espécie *Hura crepitans* sobre a rede (b).

• Rede Leste:

Figura 10 – Rede leste

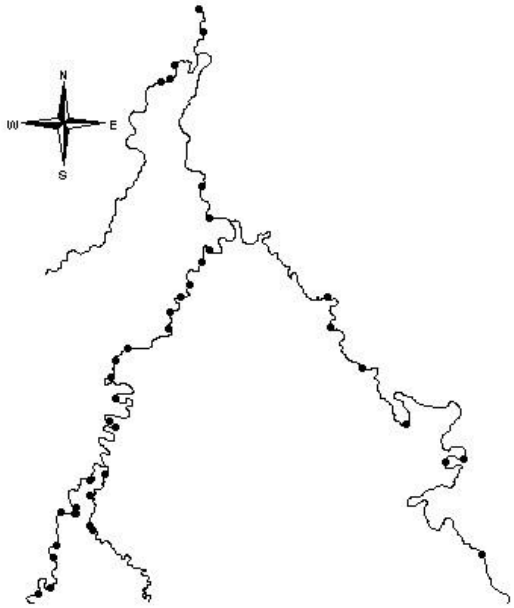
a) Rede b) Pontos



Legenda: Rede definida para o início do estudo e os pontos coleados e ajustados.
Fonte: Elaborado pelo próprio autor (2023)

É importante observar na Figura 5.1 que a configuração da rede não é linear pois os pontos não estão necessariamente posicionados sobre a rede. Nesse caso tem-se que linearizar a rede em toda sua extensão e interpolar os seus pontos de forma ortogonal para que todos os pontos (árvores) estejam localizados exatamente sobre a rede de rios. A Figura 5.2 mostra a rede linear (ajustada) de $740{,}8m$ de comprimento contendo 40 pontos (árvores). Os métodos apresentados nos capítulos anteriores serão aplicados nesta rede linear.

Figura 11 - Rede linear Oeste com as ocorrências



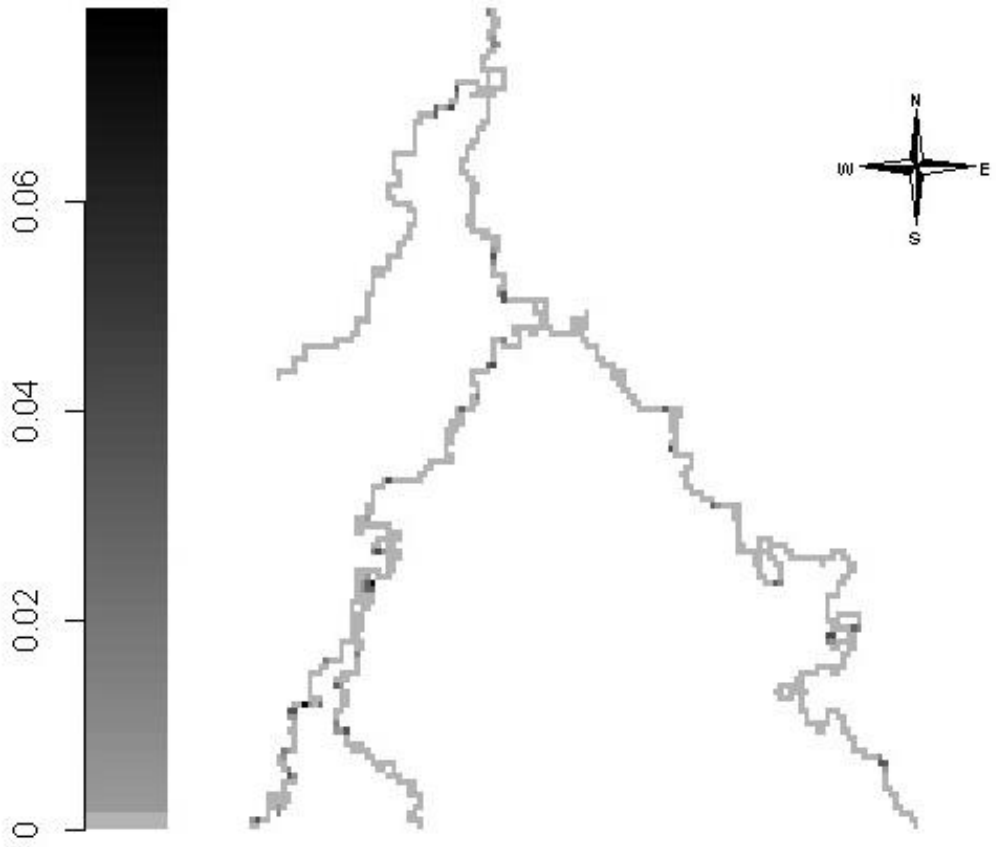
Legenda: Rede ajustada onde serão conduzidas as análises.
Fonte: Próprio autor (2023)

A intensidade global estimada da rede pontual não marcada é de $\hat{\lambda} = 0{,}054$ árvores/metro. Este valor indica que a rede de rios apresenta, em média, apenas uma árvore a cada 18,5 m, ou seja, a espécie *Hura crepitans* não é muito abundante nas margens dos rios da região estudada.

Caso existam suspeitas que a configuração pontual não marcada na rede linear não seja estacionária, ou seja, caso a intensidade não seja constante em toda a rede, pode ser interessante utilizar estimativas locais de intensidade. Timothée et al. (2010) propõe utilizar um estimador não paramétrico do tipo kernel conforme visto no Capítulo 2.

As estimativas de intensidade local podem ser apresentadas de diversas formas gráficas. Uma apresentação popular para configurações pontuais no plano é o "mapa de calor"(DIGGLE, 1995). Entretanto, essa imagem não é muito informativa no caso de redes lineares, conforme pode ser visto na Figura 5.3.

Figura 12 - Alisamento Kernel Leste em tons de cinza

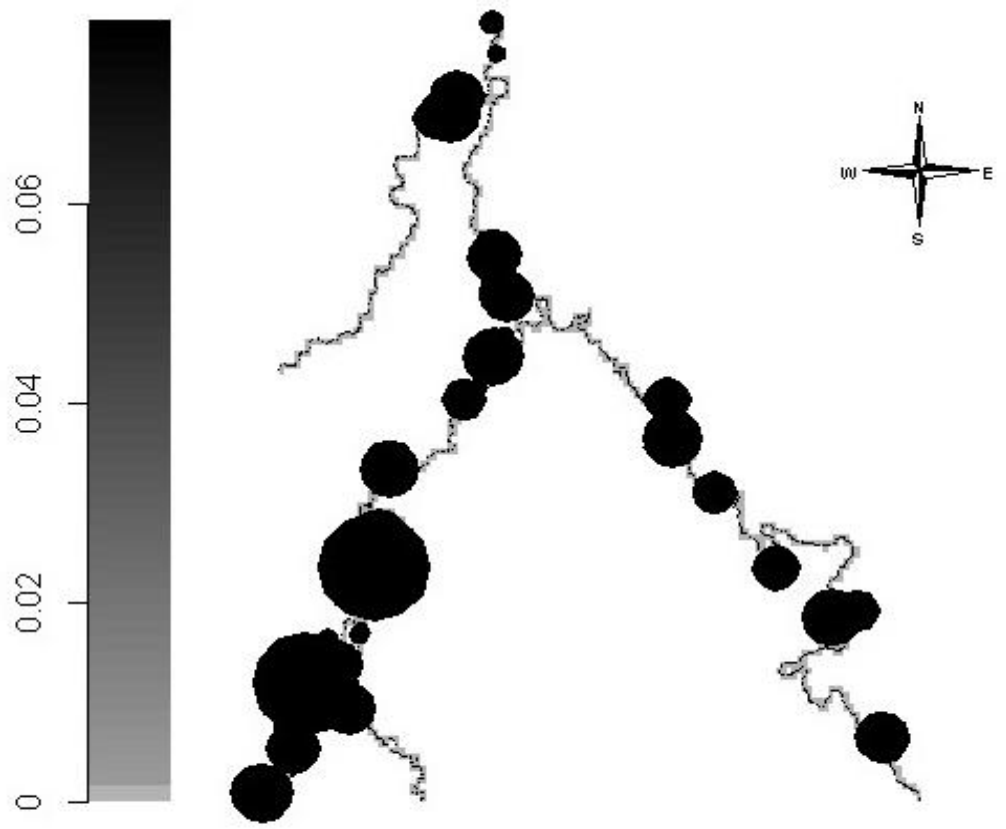


Legenda: Alisamento Kernel em rede mostrando as faixas mais escuras indicam uma maior intensidade.
Fonte: Próprio autor (2023)

Uma alternativa ao "mapa de calor"que realça a presença de altas intensidades locais na rede linear é proposta por Okabe, Okunuki et al. (2012) e recomendada por Baddeley, Jammalamadaka e Nair (2014). Nesta representação gráfica, são gerados círculos com

circunferências proporcionais as intensidades locais, conforme pode-se observar na Figura 13:

Figura 13 - Alisamento Kernel Leste

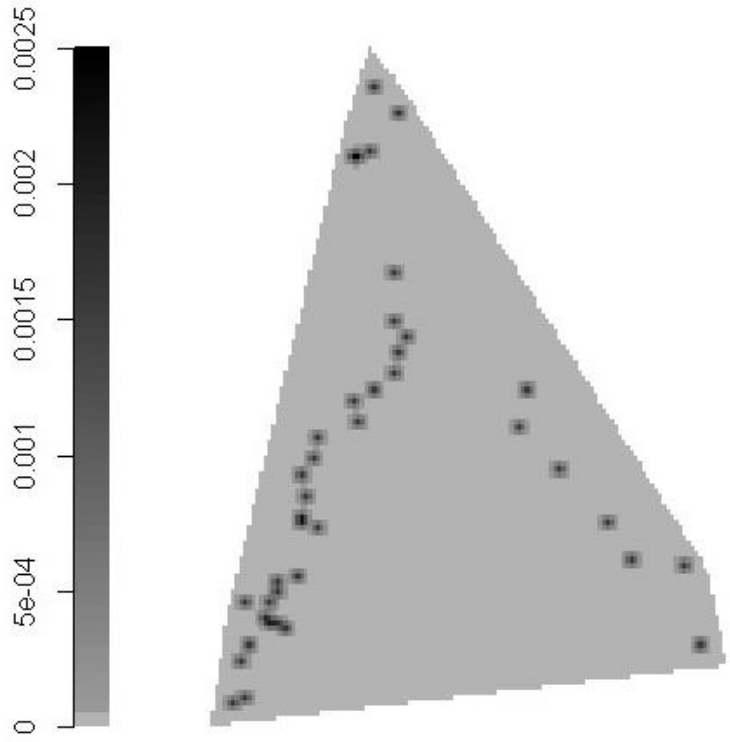


Legenda:Alisamento kernel em rede com um *zoom* na intensidade.
Fonte: Próprio autor (2023)

A Figura 5.4 mostra que existem indícios que o processo estocástico que gerou a configuração pontual na rede linear é não homogêneo, ou seja, a intensidade local de árvores varia ao longo da rede de rios.

A Figura 5.5 mostra como seria o "mapa de calor" das intensidades locais considerando o estimador kernel para configurações pontuais no plano, ou seja, ignorando que os pontos estão dispostos em uma rede linear. Pode-se observar que este procedimento produz um mapa pouco realista das propriedades de primeira ordem do processo estocástico que gerou a configuração pontual não marcada na rede linear.

Figura 14 - Alisamento kernel Leste



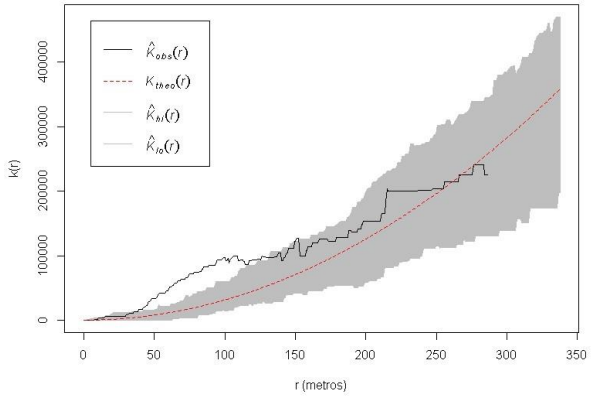
Legenda: Alisamento kernel tradicional apenas para a janela no plano.
Fonte: Próprio autor (2023)

5.2 Análise de segunda ordem

Os efeitos de segunda ordem, ou de pequena escala, estão relacionados ao grau de interação (dependência ou correlação) entre os eventos localizados na rede linear. Existem diversas funções descritoras para caracterizar esses efeitos em configurações pontuais no plano, conforme pode ser visto em Diggle (2003), Illian et al. (2008) e Baddeley et al. (2015). O mais importante desses descritores é a função *K* de Ripley (RIPLEY, 1976). No caso de redes lineares, a função *K* também desempenha um papel importante nessa caracterização, sendo o estimador proposto por Okabe e Yamada (2001), apresentado no Capítulo 2, um dos mais utilizados.

A Figura 5.6 mostra a função *K* para a rede linear de rios da região Leste, com os respectivos envelopes de confiança obtidos a partir de 99 simulações de Monte-Carlo sob a hipótese nula de completa aleatoriedade espacial.

Figura 15 - Função K de Ripley na rede leste



Legenda: Aplicação que visa verificar o agrupamento das árvores
Fonte: Próprio autor (2023)

O 5.6 mostra que a curva da função *K* observada está acima do limite superior do envelope de simulação, indicando a presença de agrupamentos de árvores da espécie *Hura crepitans* nas margens dos rios até uma distância aproximada de até de 130 m.

Deve-se observar que o estimador da função *K* proposto por Okabe e Yamada (2001) parte da suposição que o processo estocástico que gerou a configuração pontual na rede linear é homogêneo (intensidade constante). Entretanto, a análise de intensidade local apresentada anteriormente sugere que o processo é não homogêneo. Com isso, a dependência espacial (presença de agrupamentos de árvores da espécie *Hura crepitans*) detectada utilizando na Figura 5.6 deve ser feita com cautela, uma vez que os agrupamentos podem ser resultados de maiores intensidades locais de árvores e não exclusivamente da interação existente entre as árvores. De qualquer maneira, ambas as análises levam a rejeição da hipótese nula que as ávores da espécie *Hura crepitans* estão localizadas completamente ao acaso na rede de rios da região de estudo.

5.3 Independência entre pontos e marcas em redes lineares

Até o momento analisamos as propriedades de primeira e segunda ordem de um processo pontual não marcado em rede linear, ou seja, não foram consideradas nas análises as marcas (diâmetros a altura do peito) associadas cada ponto (localização das árvores).

Uma abordagem para analisar essas configurações pontuais marcadas em redes lineares é utilizar métodos desenvolvidos tendo como base a teoria de processos pontuais marcados. Entretanto, para utilizar esses métodos é absolutamente necessário que os pontos e as marcas

sejam dependentes. Essa suposição pode ser incorretamente ignorada para que seja possível utilizar métodos geoestatísticos (ex. variograma e krigagem) que são muito mais desenvolvidos e difundidos que os métodos baseados na teoria de processos pontuais marcados. Embora seja uma suposição conveniente, a independência entre marcas e pontos pode não valer na prática em muitas situações. Nesse sentido, métodos estatísticos para detectar a dependência entre marcas e pontos é de fundamental importância para garantir análises apropriadas dos dados.

Existem diversos métodos para detectar a independência entre marcas e pontos em dados referenciados no plano tais como a função de correlação de marca/variograma (WÄLDER; STOYAN, 1996), a função J (LIESHOUT, 2006), as funções E e V (SCHLATHER; RIBEIRO;

DIGGLE, 2004), a função K (GUAN, 2006), a estatística do Qui-quadrado (GUAN; AFSHARTOUS, 2007) e a estatística de Kolmogorov-Smirnov (ZHANG, 2014). Entretanto, não existe disponível na literatura um método para testar a hipótese nula de independência entre marcas e pontos em configurações pontuais marcadas em redes lineares.

Nesse sentido, esta tese propôs um método, apresentado no Capítulo anterior, que faz uma adaptação das funções E e V, propostas por (SCHLATHER; RIBEIRO; DIGGLE, 2004), para o caso de configurações pontuais marcados em redes lineares.

5.3.1 Resultados da simulação

Os resultados do método de Monte Carlo para testar a independência entre pontos e marcas em uma rede linear marcada são apresentados na Tabela 5.1 que mostra a taxa empírica de erro do tipo I (rejeitar a hipótese nula (H_0) de independência entre marcas e pontos quando ela é verdadeira).

Tabela 1 - Desempenho do teste em 500 realizações

Número de vezes que H_0 foi rejeitada (%)		
n	50 pontos	100 pontos
E_{net}	63(13)	75(15)
V_{net}	38(8)	42(8)

Fonte: Elaborado pelo próprio autor (2023)

A Tabela 5.1 apresenta os valores em que os testes propostos rejeitam a hipótese nula de independência entre pontos e marcas nas 500 realizações para os dois cenários (n = 50 *pontos* e n = 100 *pontos*). Pode-se observar que em todos os cenários a taxa empírica de erro do tipo I supera o valor nominal do nível de significância adotado ($\alpha = 0{,}05$), ou seja, o método proposto parece não apresentar um bom controle do erro do tipo I. Isso pode ser devido ao tipo de configuração de rede linear que foi utilizado na simulação. O ideal seria utilizar diversas configurações de rede para uma análise mais abrangente.

O tamanho da amostra parece não afetar expressivamente as taxas de erro do tipo I nos dois testes. Talvez porque 50 e 100 pontos já sejam números suficientemente grandes. Mais uma vez, o ideal seria testar o método em configurações com menor número de pontos.

A função E(r) parece levar a taxas de rejeição maiores que a função V(r). Esse resultado já era esperado pois a obtenção de E_{net} depende da função kernel que por sua vez é muito influenciada pela largura de banda que controla o grau de suavização da curva. Este tem sido um dos maiores problemas ao usar esses métodos em configurações pontuais no espaço, conforme apontam os trabalhos de Schlather, Ribeiro e Diggle (2004) e Guan (2006). Tudo indica que este continua a ser um problema na análise de dados de processos pontuais marcados em redes lineares.

Finalmente é importante destacar que o método deveria ser testado em realizações simuladas que exibem dependência entre pontos e marcas para avaliar o poder dos testes para rejeitar a hipótese nula que, por construção, é falsa.

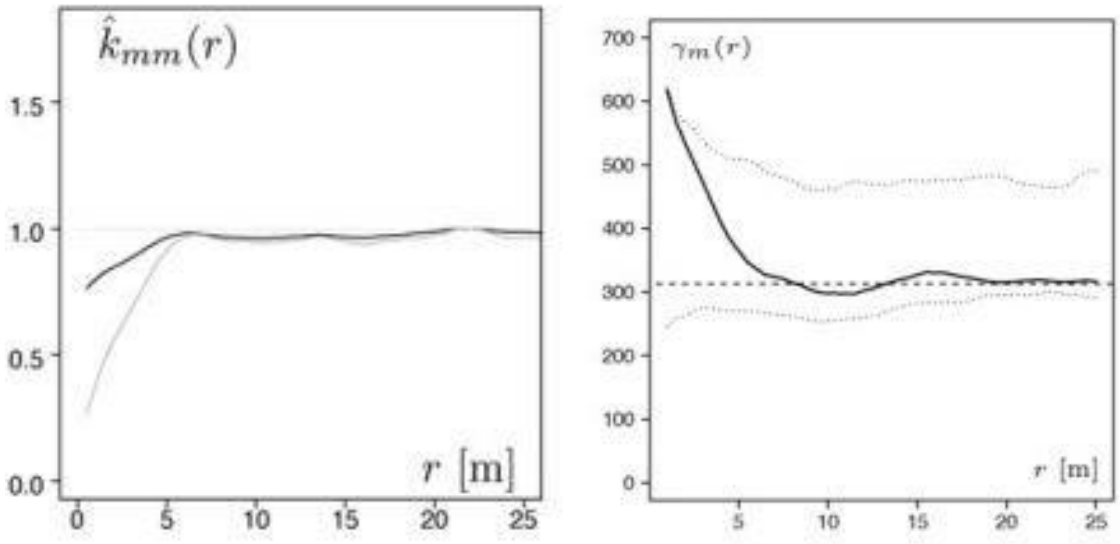
5.3.2 Aplicação em dados reais

Apesar dos resultados do trabalho de simulação mostrar que os testes não fazem um bom controle da taxa de erro do tipo I, nesta seção apresenta-se uma aplicação do método proposto ao conjunto de dados florestais apresentado no capítulo anterior. Neste contexto, os pontos representam localizações individuais de árvores da espécie *Hura crepitans* e as marcas são medidas do diâmetro à altura do peito (DAP) das árvores.

Os resultados da aplicação do método na rede linear de rios da Fazenda Canário são apresentados na Figura 5.7, em que as linhas cheias representam estimativas das funções $V_{net}(r)$ e $E_{net}(r)$, enquanto as linhas pontilhadas representam os envelopes de simulação $V(r)$ e $E(r)$ obtidos a partir de 99 realizações (simulações Monte Carlo) sob a hipótese nula.

Figura 16 - Teste de independência para rede Leste

a) $E_{net}(r)$ b)$V_{net}(r)$



Legenda: Linhas cheias representam estimativas das funções, linha pontilhadas os envelopes de simulação obtidos a partir de 99 realizações sob a hipótese nula.
Fonte: Próprio autor (2023)

Uma análise rápida da Figura 5.7 poderia sugerir que as estimativas de $V_{net}(r)$ e $E_{net}(r)$ se afastavam das hipóteses $E(r)$ = constante e $V(r)$ = constante para distâncias até 5 metros. Entretanto, uma observação mais atenta mostra que as estimativas de $V_{net}(r)$ e $E_{net}(r)$ então dentro dos envelopes de simulação ao nível de α = 0,02 de significância e, portanto, não apresentando evidências estatísticas para rejeitar a hipótese nula de independência entre as localizações individuais de árvores da espécie *Hura crepitans* e os diâmetro à altura do peito (DAP) dessas árvores.

Numa floresta suficientemente esparsa, poderíamos esperar que a competição entre as árvores estivesse ausente, inclusive, nas margens dos rios. Consequentemente, uma variação no DAP da árvore seria causada principalmente pela variação no microambiente subjacente e a independência entre pontos e marcas é uma hipótese plausível.

Em florestas mais densas, espera-se que os efeitos competitivos diretos entre árvores vizinhas violariam esta hipótese. Os resultados obtidos nesta tese mostram que árvores da espécie *Hura crepitans* poderiam ser incluídas na primeira situação e a análise da distribuição do DAP das árvores dessa espécie nas margens dos rios da Fazenda Canário poderia ser realizada utilizando métodos da geoestatística para redes lineares.

6 CONSIDERAÇÕES FINAIS

Na análise de dados de processos pontuais marcados em redes lineares é muito importante ter um teste para a hipótese nula de independência entre as marcas e as localizações (pontos). Ao assumir falsamente a independência quando na verdade ocorre uma relação de dependência entre os valores das marcas e as localizações, essencialmente, assumi-se que o processo pontual marcado na rede linear se ajusta ao modelo de campo aleatório (ex. gaussiano), ou seja, as marcas representam os valores do campo aleatório subjacente nos locais pré-definidos, o que, usualmente, é denominado de variável regionalizada.

Embora os dados para uma variável regionalizada sejam adequados para serem analisados por métodos geoestatísticos clássicos (ex. variograma, krigagem, etc.) a aplicação de métodos geoestatísticos para processos pontuais marcados em rede que violam a suposição de independência é questionável, levando a variogramas inesperados. Este tipo de problema é bem conhecido e estudado no contexto de processos pontuais marcados no plano, conforme pode ser visto em Wälder e Stoyan (1996) e não há razões para supor que seja diferente no contexto de processos pontuais marcados em redes lineares.

Assim, nesta tese é apresentado um novo método para testar a hipótese nula de independência entre as marcas e as localizações de um processo pontual marcado em rede linear. Como em Schlather et al. (2004), nosso teste também assume que as marcas são identicamente distribuidas seguindo campo aleatório gaussiano e o processo pontual marcado na rede linear é estacionário, Com isso, é possível determinar a esperança E(r) e variância condicional V(r) estimadas de uma marca, dado que existe um outro ponto do processo localizado a uma distância *r* na rede linear.

O método é fortemente dependente da suposição de estacionariedade do processo pontual. No futuro é necessário ampliar o método para que seja possível analisar dados de processos pontuais não estacionários como ocorre no plano (veja Baddeley et al., 2015)

O método proposto para análise de independendência entre marcas e pontos em processos pontuais marcados em redes lineares somente é válido para marcas que seguem, aproximadamente, a distribuição normal. Se as marcas forem discretas, então não podemos transformálas em variáveis marginalmente gaussianas e utilizar o método proposto. Assim, para acomodar marcas gaussianas e não gaussianas, em um futuro trabalho poderia adaptar o teste proposto por

Guan (2006) para redes lineares.

Nesta tese, assim como em Schlater et al (2004), foi considerado testes separados para E(r) = constante e V(r) = constante, isso porque o interesse foi preservar suas interpretações separadamente. Neste caso, deve-se observar que, duas vezes o menor dos dois valores-*p* correspondentes fornece um teste combinado conservador. No entanto, a taxa de erro do tipo I para um teste de V(r) = constante tende a ser menor do que a de um teste correspondente de E(r) = constante. Isso pode ser devido ao fato que as estimativas empíricas de E(r) estão intimamente ligadas ao método de alisamento utilizado.

O trabalho de simulação foi conduzido em uma única estrutura de rede linear que poderia comprometer as análises dos erros do tipo I e também poderia interferir no poder do teste. Assim, é necessário que o método proposto para análise de independência entre marcas e pontos em processos pontuais marcados em redes lineares seja testado em diferentes estruturas morfológicas de redes lineares e diferentes tipos de modelos para a hipótese alternativa para uma melhor avaliação do método proposto.

Os resultados obtidos nos dados da espécie *Hura crepitans* mostram que o raio de influência das árvores poderia ser de cerca de 5 m. Entretanto, uma inspeção visual dos envelopes de simulação de Monte Carlo não mostraram que essa influência é estatisticamente significante. Sabe-se que toda interpretação baseada em inspeção visual deve ser tratada com cuidado, principalmente, quando é difícil fazer a discriminação para algumas distâncias, como é o presente caso. Assim, poderia se pensar em conduzir testes formais de Monte Carlo, baseados nas distâncias de Kolmogorov-Smirnov ou Cramer-von Mises, como já é usualmente feito no plano conforme pode ser visto em Diggle (2013).

O método proposto para análise de independendência entre marcas e pontos em processos pontuais marcados em redes lineares ignora a possível disponibilidade de covariáveis como tipo de solo, idade das árvores, etc. A incorporação de tais covariáveis tornaria o método muito mais complexo, entretanto permitiria inferências mais específicas sobre os dados. Poderiam ainda ser incorporados nas análises de processos pontuais em redes lineares modelos híbridos como visto em Garreta (2010) e a análise de fluxo proposta por Sharma et al. (2021)

Uma vez rejeitada a hipótese nula de independência, um interesse fundamental é modelar a estrutura de dependência entre marcas e pontos. Wälder e Stoyan (1996) e Schlather et al. (2004) propuseram modelos úteis para o plano. No entanto, a literatura sobre modelagem de processos pontuais marcados no plano ainda é limitado e merece uma investigação mais aprofundada. No contexto de redes lineares, esses modelos são inexistentes, mas tudo indica que os modelos utilizados no plano podem ser adaptados para redes lineares.

Finalmente, os *scripts* (funções) desenvolvidos para este trabalho precisam de um polimento para serem submetidos ao CRAN do R Project (2022) ou até mesmo ao corpo técnico do SNANET (OKABE; OKUNUKI et al., 2012) para serem incorporados em uma rotina protocolar para a análise de processo pontuais em redes.

REFERÊNCIAS

ACRE. **Zoneamento Ecologico-Economico do Acre**. Fase II: documento sintese-escala 1: 250.000. [S.l.]: Sema Rio Branco, 2006.

ANG, Q. W.; BADDELEY, A.; NAIR, G. Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. **Scandinavian Journal of Statistics**, [s.l.], Wiley Online Library, v. 39, n. 4, p. 591–617, 2012.

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, Springer, p. 1–75, 2007.

BADDELEY, A.; JAMMALAMADAKA, A.; NAIR, G. Multitype point process analysis of spines on the dendrite network of a neuron. Journal of the Royal Statistical Society: Series C (Applied Statistics), **Wiley Online Library,** [s.l], v. 63, n. 5, p. 673–694, 2014.

BADDELEY, A.; RUBAK, E.; TURNER, R. **Spatial point patterns:** methodology and applications with R. [S.l.]: CRC Press, 2015.

BADDELEY, A. J.; MØLLER, J.; WAAGEPETERSEN, R. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. **Statistica Neerlandica,** Wiley Online Library, New Jersey, v. 54, n. 3, p. 329–350, 2000.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. [S.l.]: Longman Scientific & Technical Essex, 1995. v. 413.

BORRUSO, G. Network density estimation: a gis approach for analysing point patterns in a network space. Transactions in GIS, **Wiley Online Library**, [s.l.], v. 12, n. 3, p. 377–402, 2008.

CÂMARA, G.; CARVALHO, M. S. **Análise espacial de eventos.** Análise espacial de dados geográficos. Brasília: Embrapa, 2004.

CÂMARA, G.; MONTEIRO, A. M. V.; MEDEIROS, J. Representações computacionais do espaço: fundamentos epistemológicos da ciência da geoinformação. **Geografia**, Rio Claro, v. 28, n. 1, p. 83–96, 2003.

CRESSIE, N. A. **Statistics for spatial data**. [s.l.]: Wiley Online Library, 1993. 900 p.

DIGGLE, P. **A kernel method for smoothing point process data**. Applied statistics, JSTOR, p. 138–147, 1985.

DIGGLE, P. **Statistical analysis of spatial point patterns**. 2nd ed. [s.l.]: London: Arnold, 2003. 160 p.

DIGGLE, P. J. **Statistical analysis of spatial and spatio-temporal point patterns.** [s.l.]: CRC Press, 2013. 268 p.

DIGGLE, P. J.; MILNE, R. K. Bivariate cox processes: some models for bivariate spatial point patterns. Journal of the Royal Statistical Society: Series B (Methodological), **Wiley Online Library**, v. 45, n. 1, p. 11–21, 1983.

FUHRMANN, P. A.; HELMKE, U. **The mathematics of networks of linear systems**. New York: Springer, 2015. 680 p.

GELFAND, A. E. et al. **Handbook of spatial statistics**. Boca Raton: CRC press, 2010. 619 p.

GUAN, Y. Tests for independence between marks and points of a marked point process. Biometrics, **Wiley Online Library**, [s.l.], v. 62, n. 1, p. 126–134, 2006.

GUAN, Y.; AFSHARTOUS, D. R. Test for independence between marks and points of marked point processes: a subsampling approach. **Environmental and Ecological Statistics**, Springer, v. 14, n. 2, p. 101–111, 2007.

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

JAMMALAMADAKA, A. et al. Statistical analysis of dendritic spine distributions in rat hippocampal cultures. BMC bioinformatics, **BioMed Central**, [s.l.], v. 14, n. 1, p. 287, 2013.

JUNIOR, M. et al. Alocação de pátios de armazenamento de madeira em um plano de manejo florestal na Amazônia Ocidental. **Anais do XLVI SBPO-Simpósio Bras. Pesqui. Oper**, v. 46, p. 704–713, 2014.

LIESHOUT, M. V. A j-function for marked point patterns. **Annals of the Institute of Statistical Mathematics**, Springer, v. 58, n. 2, p. 235–259, 2006.

OKABE, A.; OKUNUKI, K.-I. et al. **Sanet. a spatial analysis along networks** (ver. 4.1). Tokyo, v. 14, 2012.

OKABE, A.; OKUNUKI, K.-I.; SHIODE, S. Sanet: a toolbox for spatial analysis on a network. Geographical analysis, **Wiley Online Library**, v. 38, n. 1, p. 57–66, 2006.

OKABE, A.; YAMADA, I. The k-function method on a network and its computational implementation. Geographical Analysis, Wiley Online Library, v. 33, n. 3, p. 271–290, 2001.

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.

RAKSHIT, S.; NAIR, G.; BADDELEY, A. Second-order analysis of point patterns on a network using any distance metric. **Spatial Statistics**, Elsevier, v. 22, p. 129–154, 2017.

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

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

SCALON, J. D.; SILVA, F. M. Power of tests for spatial randomness in patterns with small number of events. **Revista de Ciência e Tecnologia**, Rio de Janeiro, v. 12, n. 2, p. 7–14, 2006.

SCHABENBERGER, O.; GOTWAY, C. **Statistical methods for spatial data analysis**. [s.l.]: London: Chapman & Hall/CRC, 2005. 488 p.

SCHLATHER, M. Simulation and analysis of random fields. **R news**, Citeseer, v. 1, n. 2, p. 18–20, 2001b.

SCHLATHER, M. Characterisation of point processes with gaussian marks independent of locations. **Mathematische Nachrichten**, [s.l.], v. 239, n. 1, p. 204–214, 2002.

SCHLATHER, M.; RIBEIRO, P. J.; DIGGLE, P. J. Detecting dependence between marks and locations of marked point processes. **Journal of the Royal Statistical Society:** Series B (Statistical Methodology), Wiley Online Library, [s.l.], v. 66, n. 1, p. 79–93, 2004.

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, [s.l.], v. 42, n. 2, p. 171–187, 2000.

TIMOTHÉE, P. et al. **A network based kernel density estimator applied to Barcelona economic activities**. p. 32–45, 2010.

WÄLDER, O.; STOYAN, D. On variograms in point process statistics. **Biometrical Journal**, Wiley Online Library, [s.l.], v. 38, n. 8, p. 895–905, 1996.

ZHANG, T. A kolmogorov-smirnov type test for independence between marks and points of marked point processes. **Electronic Journal of Statistics, Institute of Mathematical Statistics and Bernoulli Society**, [s.l.], v. 8, n. 2, p. 2557–2584, 2014.

APÊNDICE A – SCRIPT

O script comentado usado para a análise segue:

```
# Carregando os pacotes necessários

library("spatstat")
library("maptools") library("sp")



# Selecionando o diretório onde estão os dados e a rede setwd("C:/Users/seu
diretório")
################################################
#o mapa já tem que ser preparado antes no QGIS
################################################

# Entrando com os dados, ou seja os pontos dados01<-read.table("teste.txt",sep =
"", dec = ".", header=TRUE)

#janela automática, sendo as coordenadas em UMT w2
= ripras(dados01$EASTING,dados01$NORTHING)
head(dados01)

# Convertendo para o formato ppp os seus pontos dados1ppp =
ppp(x = dados01$EASTING, y = dados01$NORTHING, marks =
NULL, window=w2)

#Plotando os pontos observados plot(dados1ppp, main =
"Configuração pontual \n Homogênea")

# Função que constroi a rede linear a partir de um shape if(require(spatstat,
quietly=TRUE)) {
dname <- system.file("shapes", package="maptools")
fname <- file.path(dname, "New_Shapefile.shp") fylk
```

```
<- readShapeSpatial(fname) L <- as(fylk, "linnet")
print(max(vertexdegree(L)))
L0 <- as.linnet.SpatialLines(fylk, fuse=FALSE)
print(max(vertexdegree(L0))) }

print(L)#informações da rede

#Plotando a rede sem os pontos plot(L)

# Fazendo a interpolação dos pontos para a rede lienar
X=lpp(dados1ppp,L)
#Plotando a rede com os pontos(objeto que será feito as análises) plot(X)

#Estimador global de intensidade intensity.lpp(X)

# Alisamento Kernel linear DSD <-
density(X, sigma=10) plot(DSD,
main="Alisamento Kernel")

#Função K-linear klin=envelope(X, linearK,
correction="ang", nsim=39)
#plot(klin)



################################################################
Contrução dos estimadores
################################################################

#Verificando se o processe é marcado
#Pois as marcas tem necessaraiamente que ser quantitativas e positivas

function (X, f = function(m1, m2) { m1
* m2
```

```
}, r = NULL, correction = c("isotropic", "Ripley", "translate"), method =
"density", ..., weights = NULL, f1 = NULL, normalise = TRUE, fargs = NULL,
internal = NULL)
{ stopifnot(is.ppp(X) &&
is.marked(X)) nX <- npoints(X) if
(missing(f)) f <- NULL
if (missing(correction)) correction <- NULL marx <- marks(X, dfok =
TRUE) if (is.data.frame(marx)) { nc <- ncol(marx) result <- list() for (j in
1:nc) { Xj <- X %mark% marx[, j] result[[j]] <- markcorr(Xj, f = f, r = r,
correction = correction, method = method, ..., weights = weights, f1 = f1,
normalise = normalise, fargs = fargs)
} result <- as.anylist(result)
names(result) <-
colnames(marx) return(result)
}
if (unweighted <- is.null(weights)) { weights <-
rep(1, nX)
} else { weights <- pointweights(X, weights = weights, parent = parent.frame())
stopifnot(all(weights > 0))
}

##############################################
A contruiindo o estimado E_net(v)
##############################################

h <- check.testfun(f, f1, X) f <- h$f f1 <- h$f1 ftype <- h$ftype npts <- npoints(X)
W <- X$window rmaxdefault <- rmax.rule("K", W, npts/area(W)) breaks <-
handle.r.b.args(r, NULL, W, rmaxdefault = rmaxdefault) r <- breaks$r rmax <-
breaks$max if (length(method) > 1) stop("Select only one method, please") if
(method == "density" && !breaks$even) stop(paste("Evenly spaced r values are
required if method=", sQuote("density"), sep = "")) correction.given <-
!missing(correction) && !is.null(correction) if (is.null(correction)) correction <-
c("isotropic", "Ripley", "translate") correction <- pickoption("correction",
```

```
correction, c(none = "none", border = "border", bord.modif = "bord.modif",
isotropic = "isotropic", Ripley = "isotropic", translate = "translate", translation =
"translate", best = "best"), multi = TRUE)

### correção para as k das redes por okabe e guan correction <-
implemented.for.K(correction, W$type, correction.given) Ef <- internal$Ef if
(is.null(Ef)) { Ef <- switch(ftype, mul = { mean(marx * weights)^2
}, equ = {
if (unweighted) { mtable <-
table(marx) } else { mtable <-
tapply(weights, marx, sum)
mtable[is.na(mtable)] <- 0
}
sum(mtable^2)/n
X^2
}, product = { f1m <- do.call(f1,
append(list(marx), fargs)) mean(f1m *
weights)^2 }, general = { mcross <- if
(is.null(fargs)) { outer(marx, marx, f) } else {
do.call(outer, append(list(marx, marx, f), fargs))
} if
(unweighted)
{
mean(mcross)
} else { wcross <- outer(weights,
weights, "*") mean(mcross * wcross)
}
}, stop("Internal error: invalid ftype"))
} if
(normalise)
{ theory <- 1
Efdenom <- Ef
```

```
} else {
theory
<- Ef
Efdeno
m <- 1
}

### é visto algumas mensagens de aviso em inglês pois ### toido o
codigo é adaptado do já implementado no spatstat

if (normalise) { if (Efdenom == 0) stop("Cannot normalise the mark correlation;
the denominator is zero") else if (Efdenom < 0) warning(paste("Problem when
normalising the mark correlation:",
"the denominator is negative"))
} result <- data.frame(r = r, theo = rep.int(theory, length(r))) desc <- c("distance argument r",
"theoretical value (independent marks) for %s") alim <- c(0, min(rmax, rmaxdefault)) if (ftype
%in% c("mul", "equ")) { if (normalise) { ylab <- quote(k[mm](r)) fnam <- c("k", "mm")
} else { ylab <-
quote(c[mm](r))
fnam <- c("c", "mm")
}
} else { if
(normalise) { ylab
<- quote(k[f](r))
fnam <- c("k", "f")
} else { ylab <-
quote(c[f](r)) fnam
<- c("c", "f")
} } result <- fv(result, "r", ylab, "theo", , alim, c("r", "{%s[%s]^{iid}}(r)"), desc, fname
= fnam) close <- closepairs(X, rmax) dIJ <- close$d I <- close$i
J <- close$j
XI <- ppp(close$xi, close$yi, window = W, check = FALSE) mI <-
marx[I] mJ <- marx[J] ff <- switch(ftype, mul = mI * mJ, equ = (mI ==
```

```
mJ), product = { if (is.null(fargs)) { fI <- f1(mI) fJ <- f1(mJ) } else { fI <-
do.call(f1, append(list(mI), fargs)) fJ <- do.call(f1, append(list(mJ),
fargs))
}
fI * fJ }, general = { if (is.null(fargs)) f(marx[I], marx[J]) else
do.call(f, append(list(marx[I], marx[J]), fargs))

}) if (is.logical(ff)) ff <- as.numeric(ff) else if (!is.numeric(ff)) stop("function f did not
return numeric values") if (anyNA(ff)) switch(ftype, mul = , equ = stop("some marks were
NA"), product = , general = stop("function f returned some NA values")) if (any(ff < 0))
switch(ftype, mul = , equ = stop("negative marks are not permitted"), product = , general =
stop("negative values of function f are not permitted")) if (!unweighted) ff <- ff *
weights[I] * weights[J] if (any(correction == "none")) { edgewt <- rep.int(1, length(dIJ))
Mnone <- sewsmod(dIJ, ff, edgewt, Efdenom, r, method,
...) result <- bind.fv(result, data.frame(un = Mnone), "{hat(%s)[%s]^{un}}(r)",
"uncorrected estimate of %s", "un")
} if (any(correction == "translate")) {
XJ <- ppp(close$xj, close$yj, window = W, check = FALSE) edgewt <-
edge.Trans(XI, XJ, paired = TRUE)
Mtrans <- sewsmod(dIJ, ff, edgewt, Efdenom, r, method,
...) result <- bind.fv(result, data.frame(trans = Mtrans),
"{hat(%s)[%s]^{trans}}(r)", "translation-corrected estimate of %s",
"trans") } if (any(correction == "isotropic")) { edgewt
<- edge.Ripley(XI, matrix(dIJ, ncol = 1)) Miso <-
sewsmod(dIJ, ff, edgewt, Efdenom, r, method,

...) result <- bind.fv(result, data.frame(iso = Miso), "{hat(%s)[%s]^{iso}}(r)",
"Ripley isotropic correction estimate of %s", "iso")
} nama2 <- names(result) corrxns <-
rev(nama2[nama2 != "r"])
formula(result) <- (. ~ r)
fvnames(result, ".") <- corrxns
unitname(result) <- unitname(X)
return(result) }
```

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