



ALEX MONITO NHANCOLOLO

**PROCESSOS PONTUAIS ESPACIAIS UNIVARIADOS
APLICADOS À DISTRIBUIÇÃO DE ESPÉCIES ARBÓREAS
EM FLORESTAS NATURAIS**

**LAVRAS – MG
2024**

ALEX MONITO NHANCOLOLO

**PROCESSOS PONTUAIS ESPACIAIS UNIVARIADOS APLICADOS À
DISTRIBUIÇÃO DE ESPÉCIES ARBÓREAS EM FLORESTAS NATURAIS**

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

Prof. Dr. João Domingos Scalon
Orientador

**LAVRAS – MG
2024**

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

Nhancololo, Alex Monito

Processos Pontuais Espaciais Univariados Aplicados à
Distribuição de Espécies Arbóreas em Florestas Naturais / . –
2024.

116 p. :

Dissertação–Universidade Federal de Lavras, 2024.

Orientador: Prof. Dr. João Domingos Scalon.

Coorientador: .

Bibliografia.

1. Estatística espacial. 2. Configurações pontuais. 3.
Modelos de Poisson. 4. Modelos Cox e Gibbs (Markov). 5.
Espécies florestais. I. Scalon, João Domingos. II. Título.

ALEX MONITO NHANCOLOLO

**PROCESSOS PONTUAIS ESPACIAIS UNIVARIADOS APLICADOS À
DISTRIBUIÇÃO DE ESPÉCIES ARBÓREAS EM FLORESTAS NATURAIS**

**UNIVARIATE SPATIAL POINT PROCESSES APPLIED TO THE DISTRIBUTION
OF TREE SPECIES IN NATURAL FORESTS**

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

APROVADA em 16 de julho de 2024.

Prof. Dr. João Domingos Scalon	UFLA
Prof. Dr. Douglas Mateus da Silva	UFLA
Prof. Dr. Fernando Luiz Pereira de Oliveira	UFOP

Prof. Dr. João Domingos Scalon
Orientador

**LAVRAS – MG
2024**

Dedico este trabalho a Fabião Pedro Nhancololo, Vicente Manuel Marime, André Silvestre Cuinica, José Portez Sibine, Agostinho Julião Chambe, Minelda Arão Chone, Osvaldo Dos Santos Gouveia e Eduardo Generoso Tiago Macie.

AGRADECIMENTOS

A conclusão desta Dissertação só foi possível graças ao apoio e colaboração de muitas pessoas, às quais sou profundamente grato.

Primeiramente, agradeço a Deus pela força, saúde e resiliência para enfrentar os desafios e concluir mais esta etapa da minha vida.

Ao meu orientador, Prof. Dr. João Domingos Scalon, por sua orientação incansável, paciência e sábios conselhos ao longo de todo o processo.

À Fernanda Moreira Gianasi, Giovani Festa Paludo, Prof. Dr. Rubens Manoel dos Santos pelos ensinamentos ecológicos, e Wélson António de Oliveira, meu sincero agradecimento pela contribuição valiosa neste trabalho.

À minha mãe, Hortência Alfredo Ngazane, aos meus tios, André Alfredo Ngazane, Dorófia Caetano Macie e Rogério Pedro Nhancololo, às minhas avós, Alzira Norberto Nhanombe e Vitória Manuel Marrime e aos meus irmãos, Efigênia, Júlio, Aguinalda, Rosa e Inelta, pela educação e pelos valores que me transmitiram. Seu amor e suporte incondicional foram fundamentais para meu crescimento pessoal e acadêmico.

Aos meus amigos e colegas, pelo incentivo constante e pelas horas de discussões construtivas que tornaram essa jornada mais leve e prazerosa.

Agradeço também à Universidade Federal de Lavras e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro e recursos disponibilizados para a realização desta pesquisa.

A todos os professores do Departamento de Estatística e do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, do Instituto de Ciências Exatas e Tecnológicas, da Universidade Federal de Lavras que, de alguma forma, contribuíram para minha formação acadêmica e pessoal, meus sinceros agradecimentos. Cada ensinamento e orientação foram fundamentais para meu desenvolvimento.

Finalmente, expresso meu sincero agradecimento a todos que, direta ou indiretamente, contribuíram para a realização desta pesquisa. Cada um de vocês teve um papel fundamental neste trabalho.

“Life is a journey that must be traveled no matter how bad the roads and accommodations.”

(Oliver Goldsmith)

RESUMO

A maior parte das áreas florestais no mundo (3,75 bilhões de hectares) consiste em florestas de regeneração natural e aproximadamente 1,11 bilhões de hectares são de florestas nativas. As florestas naturais, que incluem tanto florestas de regeneração natural quanto florestas nativas, devido à pouca intervenção da ação humana para seu reflorestamento, estabelecem internamente a que distância os indivíduos da mesma espécie ou de espécies diferentes devem estar para sua sobrevivência. Essas distâncias internas de separação definem distintos padrões de distribuição espacial das plantas (aleatório, agrupado e regular). Compreender os fatores responsáveis por esses padrões é crucial para um bom manejo florestal. Este estudo teve como objetivo investigar o padrão de distribuição espacial das espécies *Siparuna guianensis*, *Xylopia brasiliensis*, *Copaifera langsdorffii* e *Myrcia splendens* e modelá-lo para identificar os fatores diretamente ligados a ele. Estas espécies fazem parte de um fragmento florestal de Mata Atlântica de 6,2 hectares, existente no *Campus* da Universidade Federal de Lavras (UFLA), no estado de Minas Gerais, Brasil. Para isso, foram utilizados inventários florestais de 1987 a 2017, juntamente com métodos não paramétricos de kernel para caracterizar tendências de configuração espacial e as funções $L_{inhom}(r)$ e $J_{inhom}(r)$ para caracterizar padrões de distribuição espacial. Modelos Poisson e Cox foram ajustados usando método de máxima pseudo-verossimilhança e máxima verossimilhança Palm e validados por meio de análise de resíduos, Critério de Informação de Akaike e simulações de Monte Carlo (1000). A dinâmica espaço-temporal (1987-2017) revelou baixa mortalidade entre as espécies, com uma leve mortalidade durante 2015-2017 devido ao fenômeno El Niño registrado entre 2015 e 2016. Apesar disso, *Copaifera langsdorffii*, *Myrcia splendens* e *Siparuna guianensis* apresentaram crescimento limitado. A análise do padrão de distribuição espacial mostrou que *Siparuna guianensis* e *Copaifera langsdorffii* exibem tendência espacial mas não dependência espacial. As distribuições espaciais dessas espécies foram modeladas por meio do modelo de Poisson não homogêneo. Esse modelo mostra que a baixa abundância de *Copaifera langsdorffii* foi influenciada por fatores como acidez do solo, capacidade de troca catiônica efetiva, conteúdo de matéria orgânica e composição de argila, enquanto para a espécie *Siparuna guianensis* os fatores que influenciaram negativamente a intensidade são a baixa capacidade de germinação (<30%) e a presença de alumínio. Em contraste, *Xylopia brasiliensis* e *Myrcia splendens* demonstraram dependência espacial intraespecífica. Modelos Cox revelaram que a acidez potencial e a matéria orgânica influenciam positivamente a intensidade da espécie *Xylopia brasiliensis*, enquanto a umidade, cálcio, magnésio e alumínio

influenciam negativamente a intensidade restrita de *Myrcia splendens* ao sudeste. Este estudo contribuiu para a compreensão da dinâmica espaço-temporal e dos padrões de distribuição espacial das espécies estudadas, fornecendo modelos explicativos para sua distribuição.

Palavras-chave: estatística espacial; configurações pontuais; modelos de Poisson; modelos Cox e Gibbs (Markov); espécies florestais.

ABSTRACT

The majority of forested areas worldwide (3.75 billion hectares) consist of naturally regenerating forests, and approximately 1.11 billion hectares are native forests. Natural forests, encompassing both naturally regenerating and native forests, establish internal distances at which individuals of the same or different species must be for their survival, due to minimal human intervention in reforestation efforts. These internal separation distances define distinct spatial distribution patterns of plants (random, clustered, and regular). Understanding the factors responsible for these patterns is crucial for effective forest management. This study aimed to investigate the spatial distribution pattern of the species *Siparuna guianensis*, *Xylopia brasiliensis*, *Copaifera langsdorffii*, and *Myrcia splendens*, and to model these patterns to identify directly linked factors. These species inhabit a 6.2-hectare fragment of Atlantic Forest located at the Campus of the Federal University of Lavras (UFLA), Minas Gerais, Brazil. Forest inventories from 1987 to 2017 were utilized alongside non-parametric kernel methods to characterize spatial configuration trends, and $L_{inhom}(r)$ and $J_{inhom}(r)$ functions to characterize spatial distribution patterns. Poisson and Cox models were fitted using maximum pseudo-likelihood and maximum likelihood Palm methods, validated through residual analysis, Akaike Information Criterion, and Monte Carlo simulations (1000). Spatio-temporal dynamics (1987-2017) revealed low mortality among species, with slight mortality observed during 2015-2017 due to the El Niño phenomenon recorded between 2015 and 2016. Despite this, *Copaifera langsdorffii*, *Myrcia splendens*, and *Siparuna guianensis* showed limited growth. Spatial distribution pattern analysis indicated that *Siparuna guianensis* and *Copaifera langsdorffii* exhibited spatial trends rather than spatial dependence, modeled through inhomogeneous Poisson models. The low abundance of *Copaifera langsdorffii* was influenced by factors such as soil acidity, effective cation exchange capacity, organic matter content, and clay composition, while for *Siparuna guianensis*, negative influences included low germination capacity (<30%) and presence of aluminum. In contrast, *Xylopia brasiliensis* and *Myrcia splendens* demonstrated intra-specific spatial dependence. Cox models revealed that potential acidity and organic matter positively influenced *Xylopia brasiliensis*, whereas moisture, calcium, magnesium, and aluminum negatively influenced the restricted abundance of *Myrcia splendens* in the southeast. This study contributes

to understanding the spatio-temporal dynamics and spatial distribution patterns of the studied species, providing explanatory models for their distribution.

Keywords: spatial statistics; point patterns; Poisson models; Cox and Gibbs (Markov) models; forest species.

INDICADORES DE IMPACTO

A pesquisa sobre Processos Pontuais Espaciais Univariados Aplicados à Distribuição de Espécies Arbóreas em Florestas Naturais oferece uma compreensão detalhada da dinâmica espaço-temporal das quatro espécies florestais *Siparuna guianensis*, *Xylopia brasiliensis*, *Copaifera langsdorffii* e *Myrcia splendens*. Isso permite entender a evolução dessas espécies ao longo do tempo, considerando que elas estão localizadas em uma área que foi impactada por ações humanas, como o corte de árvores pela comunidade local e durante aulas práticas e experimentos com alunos da Universidade Federal de Lavras, antes de a área ser declarada protegida. O estudo dos padrões de distribuição espacial dessas espécies ajuda a identificar suas preferências por determinados locais e a determinar a distância mínima de separação entre plantas da mesma espécie. Isso fornece aos ecologistas dados valiosos para o manejo florestal, especialmente em projetos de reflorestamento envolvendo essas espécies. A modelagem dessas distribuições contribui cientificamente para a ecologia e outras áreas, por ser uma abordagem ainda pouco explorada. Além disso, permite identificar fatores associados aos padrões espaciais observados. Para as quatro espécies em estudo, a pesquisa gerou avanços significativos que beneficiam pesquisadores e gestores florestais na implementação de políticas de conservação e práticas de manejo sustentável. Isso não só ajuda a evitar a extinção das espécies estudadas, mas também aumenta sua presença na floresta por meio de intervenções corretivas baseadas nos fatores identificados como negativamente associados. A pesquisa foi aplicada no setor florestal de Lavras, no Brasil, mas pode ser aplicada a outras regiões do mundo. Além disso, promove a conservação da biodiversidade e a saúde dos ecossistemas florestais, alinhando-se a quatro áreas temáticas da Política Nacional de Extensão: cultura, educação, meio ambiente e saúde, bem como a dois Objetivos de Desenvolvimento Sustentável da ONU: ação contra a mudança global do clima e vida terrestre.

IMPACT INDICATORS

The research on Univariate Point Process Models Applied to the Distribution of Tree Species in Natural Forests provides a detailed understanding of the spatiotemporal dynamics of four forest species: *Siparuna guianensis*, *Xylopia brasiliensis*, *Copaifera langsdorffii*, and *Myrcia splendens*. This approach facilitates an understanding of how these species have evolved over time, considering that they are located in an area previously affected by human activities, such as tree cutting by the local community and during practical classes and experiments involving students from the Federal University of Lavras, before the area was designated as protected. The study of the spatial distribution patterns helps to identify the preferences for specific locations and determine the minimum distance of separation between individuals of the same species. This provides ecologists with valuable data for forest management, especially in reforestation projects. Modeling these distributions makes significant scientific contributions to ecology and related fields, as it is an approach that remains underexplored. Moreover, it enables the identification of factors associated with the observed spatial patterns. For the four species under study, the research has generated significant advancements that benefit researchers and forest managers in the implementation of conservation policies and sustainable management practices. This not only aids in preventing the extinction of the studied species but also enhances their presence in the forest through corrective interventions based on the identified negatively associated factors. The research has been applied to the forestry sector in Lavras, Brazil, but it is also applicable to other regions worldwide. Furthermore, it promotes biodiversity conservation and the health of forest ecosystems, aligning with four thematic areas of the National Extension Policy: culture, education, environment, and health, as well as two United Nations Sustainable Development Goals: action against climate change and life on land.

LISTA DE FIGURAS

Figura 2.1 – Intensidade: (a) não suavizada, (b) suavizada e (c) estimativa de Kernel. . .	32
Figura 2.2 – Funções kernel	33
Figura 2.3 – Padrões de Distribuição Espacial: (a) aleatório, (b) regular e (c) agrupado. .	35
Figura 2.4 – Padrões de distribuição espacial na função $F(r)$: (a) regular, (b) aleatório e (c) agrupado.	36
Figura 2.5 – Padrões de distribuição espacial na função $G(r)$: (a) regular, (b) aleatório e (c) agrupado.	38
Figura 2.6 – (a) distância de um evento (s_i) ao evento vizinho (s_j) mais próximo ($G(r)$), sem efeito da borda; (b) distância de um espaço vazio (z) ao evento s_i mais próximo ($F(r)$), sem efeito da borda; (c) demonstração da situação em que o efeito da borda está presente.	38
Figura 2.7 – Padrões de distribuição espacial na função $J(r)$: (a) regular, (b) aleatório e (c) agrupado.	39
Figura 2.8 – Ideia intuitiva por trás da estimação da função $K(r)$	40
Figura 2.9 – Padrões de distribuição espacial na função $K(r)$: (a) regular, (b) aleatório e (c) agrupado.	40
Figura 2.10 – Padrões de distribuição espacial na função $L(r)$: (a) regular, (b) aleatório e (c) agrupado.	41
Figura 2.11 – Padrões de distribuição espacial na função $g(r)$: (a) regular, (b) aleatório e (c) agrupado.	42
Figura 2.12 – Estrutura de dependência espacial: (a) sem dependência espacial (Presença de tendência), (b) existe dependência espacial (padrão agrupado)	44
Figura 2.13 – Ilustração do campo aleatório (superfície) $\Psi(s)$, que controla a abundância e distribuição dos eventos em um modelo Cox	47
Figura 3.1 – Mapa ilustrativo da localização da área de estudo	60
Figura 3.2 – Espécies florestais: (a) <i>xylopia brasiliensis</i> (pindaíba) e, (b) <i>myrcia splen-</i> <i>dens</i> (guamirim)	62
Figura 3.3 – Espécies florestais: (a) <i>siparuna guianensis</i> (negamina) e, (b) <i>copaifera</i> <i>langsdorffii</i> (copaíba)	63
Figura 4.1 – Dinâmica espaço-temporal das espécies arbóreas de 1987 a 2017.	65
Figura 4.2 – Distribuição espacial das espécies florestais nos anos 1987 (a) e 2017 (b). .	66

Figura 4.3 – Densidade das espécies <i>Xylopia brasiliensis</i> , <i>Copaifera langsdorffii</i> , <i>Myrcia splendens</i> e <i>Siparuna guianensis</i>	68
Figura 4.4 – Distribuição das espécies <i>Xylopia brasiliensis</i> e <i>Copaifera langsdorffii</i> por DAP (o) no ano 2017.	69
Figura 4.5 – Padrão de distribuição espacial das espécies <i>Copaifera langsdorffii</i> e <i>Siparuna guianensis</i>	70
Figura 4.6 – Padrão de distribuição espacial das espécies <i>Xylopia brasiliensis</i> e <i>Myrcia splendens</i>	71
Figura 4.7 – Modelo ajustado para a espécie <i>Copaifera langsdorffii</i>	73
Figura 4.8 – Resíduos do modelo ajustado à espécie <i>Copaifera langsdorffii</i>	74
Figura 4.9 – <i>Siparuna guianensis</i> : (a) modelo ajustado e (b) q-q plot	75
Figura 4.10 – Resíduos do modelo ajustado à espécie <i>Siparuna guianensis</i>	76
Figura 4.11 – Modelos ajustados para: (a) <i>xylopia brasiliensis</i> e (b) <i>myrcia splendens</i> . . .	78

LISTA DE TABELAS

Tabela 4.1 – Estatísticas descritivas das espécies arbóreas estudadas, registradas no inventário florestal de 2017.	67
Tabela 4.2 – Covariáveis selecionadas pelo método <i>stepwise</i> para a espécie <i>Copaifera langsdorffii</i>	72
Tabela 4.3 – Covariáveis selecionadas pelo método <i>stepwise</i> para a espécie <i>Siparuna guianensis</i>	75
Tabela 4.4 – Covariáveis selecionadas pelo método <i>stepwise</i> para a espécie <i>Xylopia brasiliensis</i>	77
Tabela 4.5 – Covariáveis selecionadas pelo método <i>stepwise</i> para a espécie <i>myrcia splendens</i>	78

LISTA DE ABREVIATURAS

AEC	Aleatoriedade espacial completa
DAP	Diâmetro à altura do peito
$F(r)$	Função $F(r)$ homogênea
$G(r)$	Função $G(r)$ homogênea
$J(r)$	Função $J(r)$ homogênea
$K(r)$	Função $k(r)$ ou κ de Ripley homogênea
$L(r)$	Função $L(r)$ homogênea
$g(r)$	Função correlação par homogênea
B	Região de estudo
S	Subregião de estudo
$n(S)$	Número de eventos
s_i	Eventos de estudo
(x, y)	Coordenadas geográficas
$m(s_i)$	Marcas associadas aos eventos $m(s_i)$
$N \stackrel{D}{=} N_t$	Processo pontual N e N_t apresentam mesma distribuição
$ B $	Área ou volume da região B
$\lambda(s_i)$	Intensidade dos eventos s_i
$e(s_i)$	Efeito da borda para os eventos s_i
$\lambda(s_i s_{i'})$	Intensidade condicional
z	Pontos quaisquer ou locais onde não ocorrem eventos dentro da área de estudo
r	Raio no qual se pretende ver o padrão espacial dos eventos
$\hat{F}(r)$	Estimador da função $F(r)$ homogênea
$\hat{G}(r)$	Estimador da função $G(r)$ homogênea

$\hat{J}(r)$	Estimador da função J(r) homogênea
$\hat{K}(r)$	Estimador da função k(r) ou κ de Ripley homogênea
$\hat{L}(r)$	Estimador da função L(r) homogênea
$\hat{g}(r)$	Estimador da função correlação par homogênea
$F_{pois}(r)$	Função F(r) Poisson homogênea
$G_{pois}(r)$	Função G(r) Poisson homogênea
$J_{pois}(r)$	Função J(r) Poisson homogênea
$K_{pois}(r)$	Função k(r) (κ de Ripley) Poisson homogênea
$L_{pois}(r)$	Função L(r) Poisson homogênea
$g_{pois}(r)$	Função correlação par Poisson homogênea
$\hat{F}_{bord}(r)$	Estimador da função F(r) quando existe efeito da borda
$\hat{G}_{bord}(r)$	Estimador da função G(r) quando existe efeito da borda
$\hat{K}_{bord}(r)$	Estimador da função K(r) quando existe efeito da borda
$\hat{L}_{bord}(r)$	Estimador da função L(r) quando existe efeito da borda
$\hat{g}_{bord}(r)$	Estimador da função g(r) quando existe efeito da borda
$F_{inhom}(r)$	Função F(r) não homogênea
$G_{inhom}(r)$	Função G(r) não homogênea
$J_{inhom}(r)$	Função J(r) não homogênea
$K_{inhom}(r)$	Função k(r) ou κ de Ripley não homogênea
$L_{inhom}(r)$	Função L(r) não homogênea
$g_{inhom}(r)$	Função correlação par não homogênea
$\hat{F}_{inhom}(r)$	Estimador da função F(r) não homogênea
$\hat{G}_{inhom}(r)$	Estimador da função G(r) não homogênea
$\hat{J}_{inhom}(r)$	Estimador da função J(r) não homogênea
$\hat{K}_{inhom}(r)$	Estimador da função k(r) ou κ de Ripley não homogênea

$\hat{L}_{inhom}(r)$	Estimador da função $L(r)$ não homogênea
$\hat{g}_{inhom}(r)$	Estimador da função correlação par não homogênea
$L_{inhom}(r) - r$	Função $L(r)$ não homogênea transformada, considerando $(L(r) = r) - r = 0$ para a tornar constante
β	Constante que representa intensidade
θ^\top	Transposta do vetor de parâmetros θ a serem estimados
$\Lambda(s)$	Função intensidade aleatória
$\Psi(s)$	Campo aleatório
$\gamma(r)$	Função covariância exponencial
AIC	Critério de Informação de Akaike
$R_i^P(B)$	Resíduos de Pearson na subregião i (<i>quadrats</i> i)

LISTA DE SÍMBOLOS

pH	Potencial hidrogeniônico
K	Potássio
P	Fósforo
Ca	Cálcio
Mg	Magnésio
Al	Alumínio
$H + Al$	Acidez potencial
SB	Soma de Bases
t	Capacidade de Troca de Cátions Efetiva
T	Capacidade de Troca de Cátions Total
V	Saturação de Bases
m	Saturação de Alumínio
MO	Matéria Orgânica
μ	média
σ^2	Variância

SUMÁRIO

1	INTRODUÇÃO	22
1.1	Objetivos	24
1.1.1	Geral	24
1.1.2	Específicos	24
2	REFERENCIAL TEÓRICO	25
2.1	Processos pontuais	25
2.2	Elementos de processos pontuais	26
2.2.1	Marcas	26
2.2.2	Covariáveis	26
2.3	Propriedades fundamentais de processos pontuais	27
2.3.1	Estacionaridade e Isotropia	27
2.3.2	Homogeneidade, Independência e Tendência Espacial	28
2.4	Principais distribuições de probabilidade em processos pontuais	28
2.4.1	Distribuição Uniforme	28
2.4.2	Distribuição Binomial	29
2.4.3	Distribuição Poisson	29
2.5	Propriedade de primeira e segunda ordem	30
2.5.1	Propriedade de primeira ordem	30
2.5.2	Propriedade de segunda ordem	34
2.6	Padrões de distribuição espacial e sua identificação	34
2.6.1	Funções $F(r), G(r), J(r), K(r), L(r)$ e $g(r)$	35
2.6.1.1	Função $F(r)$	35
2.6.1.2	Função $G(r)$	37
2.6.1.3	Função $J(r)$	39
2.6.1.4	Função $K(r)$ ou função κ de Ripley	39
2.6.1.5	Função $L(r)$	41
2.6.1.6	Função $g(r)$	41
2.6.2	Funções $F_{inhom}(r), G_{inhom}(r), J_{inhom}(r), K_{inhom}(r), L_{inhom}(r)$ e $g_{inhom}(r)$	42
2.7	Modelos de processos pontuais	44
2.7.1	Modelos Poisson	45
2.7.1.1	Modelo Poisson homogêneo	45

2.7.1.2	Modelo Poisson não homogêneo	45
2.7.2	Modelos Cox	46
2.7.2.1	Modelo Log Cox Gaussiano	47
2.7.2.2	Modelos de Neyman-Scott: Modelo Matérn, Thomas, Variância-Gama e Cauchy	48
2.7.2.2.1	Modelo Matérn	49
2.7.2.2.2	Modelo Thomas	49
2.7.2.2.3	Modelo Cauchy	49
2.7.2.2.4	Modelo Variância Gama	50
2.7.3	Modelos Gibbs (Markov)	50
2.7.3.1	Modelos de interação par a par	50
2.7.3.1.1	Modelo Hard core	52
2.7.3.1.2	Modelo de nascimento e morte	52
2.7.3.1.3	Modelo Strauss	52
2.7.3.1.4	Modelo Strauss Hard Core	53
2.7.3.1.5	Modelo de interação soft-core	53
2.7.3.1.6	Modelo de interação Diggle-Gates-Stibbard	54
2.7.3.1.7	Modelo de interação Diggle-Gratton	54
2.7.3.2	Modelos de interação por área	54
2.8	Diagnóstico e Validação de Modelos	56
2.8.1	Diagnóstico	56
2.8.2	Validação	58
3	MATERIAL E MÉTODOS	60
3.1	Material	60
3.1.1	Descrição das florestais em estudo	61
3.2	Métodos	64
4	RESULTADOS E DISCUSSÕES	65
4.1	Dinâmica espaço-temporal das espécies arbóreas em estudo	65
4.2	Modelagem da distribuição espacial das árvores das espécies florestais .	72
5	CONCLUSÕES E RECOMENDAÇÕES	80
	REFERÊNCIAS	82
A	ANEXO - Códigos em R	88

1 INTRODUÇÃO

As florestas desempenham um papel crucial no alcance dos objetivos de desenvolvimento sustentável, abrangendo desde a promoção da produção e consumo sustentável até a mitigação da pobreza, garantia da segurança alimentar, conservação da biodiversidade e enfrentamento das mudanças climáticas (FRA, 2020). No entanto, no Brasil e não só, esses ecossistemas enfrentam sérias ameaças devido à expansão agrícola, exploração madeireira, mineração e mudanças climáticas, resultando em perdas irreparáveis de habitat e biodiversidade (Vieira, 2011; Belchior, 2022; Fruehauf; Silva; Lombardo, 2022). Em 2023, por exemplo, os incêndios florestais no Brasil atingiram aproximadamente 68.960 km^2 , conforme relatado pelo INPE (2023).

Dada a importância que as florestas apresentam, 31% da área total da Terra (4,06 bilhões de hectares) é composta por florestas, sendo 290 milhões de hectares de florestas plantadas, que correspondem a 7% do total da área florestal, e 3,75 bilhões de hectares de florestas de regeneração natural¹ que correspondem a 93% do total da área florestal. Estando no topo de países com maior extensão florestal (54% do total), a Rússia (815 milhões de hectares), Brasil (497 milhões de hectares), Canadá (347 milhões de hectares), Estados Unidos da América (310 milhões de hectares) e China (220 milhões de hectares) (FAO, 2020; FRA, 2020). No entanto, apenas 1,11 bilhões de hectares são de florestas nativas², com a maior extensão (61%) dessas florestas nativas encontrada na Rússia, Canadá e Brasil, que abriga cerca de 485.396.000 hectares, das quais apenas 8.700 foram registradas e aproximadamente 1.000 são encontradas em Minas Gerais.

As florestas naturais, que incluem tanto florestas de regeneração natural quanto florestas nativas, por não serem diretamente influenciadas pela intervenção humana para seu reflorestamento, estas, estabelecem internamente distâncias a que da mesma espécie (distância intra-específica) e de espécies diferentes (distância interespecífica) devem estar. Essas distâncias de separação resultam em diferentes padrões de distribuição espacial das plantas, que podem ser identificados por meio do georreferenciamento de cada planta por espécie florestal. Esse processo permite a adoção de estratégias eficazes de conservação e manejo sustentável das florestas.

¹ Florestas de regeneração natural são áreas onde a vegetação se recupera naturalmente após ter sido perturbada por atividades humanas, como agricultura, exploração madeireira, incêndios florestais ou outros tipos de degradação (Hecht, 2014; Schelhas; Brandeis; Rudel, 2021; Paiva et al., 2022).

² Florestas nativas ou primárias são ecossistemas florestais que nunca foram expressivamente alterados por atividades humanas (FRA, 2020).

Nesse contexto, a estatística espacial, com ênfase em processos pontuais, desempenha um papel crucial. Ela oferece uma ferramenta eficaz para compreender a dinâmica espaço-temporal das espécies arbóreas em áreas florestais naturais. Isso inclui a determinação das distâncias apropriadas entre no caso de reflorestamento exótico, a investigação do padrão de distribuição e da estrutura de dependência espacial, a análise da segregação espacial, e a modelagem dos processos subjacentes a essa distribuição espacial das espécies arbóreas em florestais naturais (Shimatani, 2001; Ferrari et al., 2011; Wiegand; Moloney, 2014; Jalilian; Safari; Sohrabi, 2020). No entanto, os modelos de processos pontuais ainda não são amplamente utilizados pelos ecologistas, que tendem a se concentrar em características de resumo, como índice de Morisita (1959) e funções $F(r)$, $G(r)$, $J(r)$, $K(r)$, $L(r)$, $g(r)$ ou suas extensões não homogêneas. Essa restrição, conforme descrito por Illian (2019), pode estar associada à complexidade matemática dos modelos e à dificuldade associada ao seu ajuste e interpretação.

O presente estudo visa preencher essa lacuna, aplicando técnicas de processos pontuais para investigar o padrão de distribuição espacial das espécies *Siparuna guianensis*, *Xylopia brasiliensis*, *Copaifera langsdorffii* e *Myrcia splendens* e, modelá-lo para identificar os fatores diretamente ligados a ele. Estas espécies fazem parte do fragmento florestal de Mata Atlântica, conhecido localmente como “matinha”, localizado no *Campus* da Universidade Federal de Lavras (UFLA).

Esta pesquisa está estruturada em cinco capítulos: introdução, que contextualiza o estudo, define os objetivos, apresenta a problematização e justifica a pesquisa; referencial teórico, que descreve os processos pontuais, suas subdivisões e respectivos modelos; materiais e métodos, que detalha sucintamente os materiais e os métodos utilizados; resultados e discussões, onde são analisados os dados e discutidos os resultados obtidos; conclusão e recomendações, onde são apresentadas as conclusões do estudo e recomendações.

1.1 Objetivos

Para a realização da pesquisa foram definidos os seguintes objetivos:

1.1.1 Geral

Caracterizar a distribuição espacial intraespecífica das espécies arbóreas *Myrcia splendens*, *Siparuna guianensis*, *Xylopia brasiliensis* e *Copaifera langsdorffii* localizadas na “matinha” da Universidade Federal de Lavras (UFLA).

1.1.2 Específicos

- a) descrever a dinâmica espaço-temporal das espécies arbóreas *Myrcia splendens*, *Siparuna guianensis*, *Xylopia brasiliensis* e *Copaifera langsdorffii* de 1987 a 2017;
- b) estudar a estrutura de tendência e dependência espacial intraespecífica apresentada pelas espécies arbóreas *Myrcia splendens*, *Siparuna guianensis*, *Xylopia brasiliensis* e *Copaifera langsdorffii*, no ano 2017;
- c) modelar a distribuição espacial das espécies arbóreas *Myrcia splendens*, *Siparuna guianensis*, *Xylopia brasiliensis* e *Copaifera langsdorffii*.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta uma revisão sobre os conceitos fundamentais relacionados aos processos pontuais. Serão exploradas as principais distribuições e padrões (configurações espaciais) que podem ocorrer. Discutiremos a tendência espacial e as técnicas para sua identificação, além de analisar padrões espaciais utilizando as funções F , G , J , K , L , g ou suas extensões não homogêneas. Ademais, serão abordados os modelos de processos pontuais, incluindo os modelos de Poisson, Cox e Gibbs, com ênfase em seus métodos de ajuste e validação.

2.1 Processos pontuais

Constituem conjunto de dados que fornecem as localizações espaciais de fenômenos ou eventos observados (Diggle, 2013; Baddeley; Rubak; Turner, 2015; Scalon, 2024). São modelos matemáticos que descrevem a disposição de eventos que estão irregular ou aleatoriamente distribuídos no plano, ou no espaço (Illian et al., 2008). São conjuntos contáveis de eventos espaciais que surgem como realizações de processos estocásticos (aleatórios ou probabilísticos) de eventos espaciais tomando valores em uma região plana $B \subset \mathbb{R}^2$ (Moller; Waagepetersen, 2003; Mateus, 2013; Leininger, 2014; Moraga, 2023).

Seja $S \subseteq \mathbb{R}^d$ um espaço métrico, onde \mathbb{R}^d denota um espaço euclidiano d -dimensional, onde os eventos ocorrem. Um processo pontual espacial em $B \subseteq \mathbb{R}^d$, é uma coleção de variáveis aleatórias $n(S) : S \subseteq B$, onde $n(S)$ representa o número eventos ocorridos em um conjunto mensurável S de B .

Processos pontuais referem-se a uma área da estatística espacial que estuda a distribuição de eventos em uma área geográfica específica delimitada, onde as coordenadas geográficas são a própria informação. Isto é, os eventos são geralmente representados por pontos e descritos utilizando coordenadas geográficas. São exemplos de processos pontuais, coordenadas geográficas da localização de: árvores, estabelecimentos comerciais, acidentes rodoviários, epicentros de terremotos, crimes/raptos, ninhos de pássaros, Universidades, etc.

Dependendo do número de tipos diferentes de eventos considerados, um processo pontual pode ser classificado como univariado, quando envolve um único tipo de evento ou mais de um tipo de evento sendo estudado cada um separadamente, ou seja, sem investigar a relação entre diferentes eventos; e multivariado, quando envolve mais de um tipo de evento e o objetivo

é investigar a relação entre diferentes eventos (dependência interespecífica). Esta pesquisa se concentra exclusivamente em processos pontuais univariados.

2.2 Elementos de processos pontuais

Nesta seção, apresentamos os principais elementos que compõem os processos pontuais, incluindo as marcas e as covariáveis.

2.2.1 Marcas

Um processo pontual espacial pode envolver não apenas a localização dos eventos no espaço, mas também informações qualitativas ou quantitativas adicionais associadas a cada evento, ou localização $s_i = (x, y)$, em que x e y são as respectivas coordenadas. Essas informações adicionais são denominadas marcas, e são representadas por $m(s_i)$ (Illian, 2019; Baddeley; Rubak; Turner, 2015; Scalco, 2024). Assim, considerando $B \subseteq \mathbb{R}^d$ e $C \subseteq \mathbb{R}$, em que B representa a região em que ocorrem os eventos ou fenômenos $s_i = (x, y)$ de interesse e C as informações adicionais para cada evento ou fenômeno $s_i = (x, y)$, um processo pontual marcado $M(B \times C)$, pode ser representado,

$$M(B \times C) = \{[s_i, m(s_i)]\} = \{[s_1, m(s_1)], [s_2, m(s_2)], \dots, [s_n, m(s_n)]\}, s_i \in B, m(s_i) \in C.$$

Em um processo pontual onde os eventos são localizações de árvores em uma floresta, exemplos de marcas incluem a altura das árvores, tipo de espécie florestal, diâmetro à altura do peito (DAP), entre outros. Em um processo pontual com eventos sendo acidentes de trânsito em uma cidade, as marcas podem representar a gravidade do acidente (ferimentos leves, graves ou fatalidades). Em um processo pontual onde os eventos são estabelecimentos comerciais em cidade, as marcas podem representar o tipo de estabelecimento (restaurantes, lojas de roupas, supermercados).

2.2.2 Covariáveis

Além das marcas, um processo pontual pode incluir variáveis explicativas, também conhecidas como covariáveis. Essas covariáveis são variáveis adicionais que auxiliam na compreensão dos motivos pelos quais os eventos ocorrem em determinados locais ou por que ocorrem de maneira desigual.

Baddeley, Rubak & Turner (2015) definem covariáveis como quaisquer dados tratados como explicativos e não como resposta, tomando como exemplo, uma função espacial $z(s_i)$ que descreve a altitude no evento s_i , em uma floresta. Em um processo pontual onde os eventos são crimes em uma cidade, as covariáveis podem incluir densidade populacional, níveis de desemprego e acesso a serviços policiais.

2.3 Propriedades fundamentais de processos pontuais

Nesta seção, apresentamos as propriedades fundamentais de processos pontuais, incluindo estacionaridade, isotropia, homogeneidade, independência e tendência espacial.

2.3.1 Estacionaridade e Isotropia

Um processo pontual é estacionário se a distribuição espacial dos eventos é invariante sob condição de translação (Illian et al., 2008; Baddeley; Rubak; Turner, 2015). Seja $N = \{s_1, s_2, \dots, s_n\}$ um processo pontual e $N_t = \{s_1 + t, s_2 + t, \dots, s_n + t\}$ a sua translação, este será estacionário se para qualquer $t \in \mathbb{R}^d$, N e N_t apresentarem a mesma distribuição espacial dos eventos, ou seja,

$$\forall t \in \mathbb{R}^d, N \stackrel{D}{=} N_t \Leftrightarrow \{s_1, s_2, \dots, s_n\} \stackrel{D}{=} \{s_1 + t, s_2 + t, \dots, s_n + t\}, \quad (2.1)$$

em que t indica as unidades transladadas. No caso de um processo pontual marcado, onde $m(s_i)$ é a marca associada ao evento, a expressão 2.1 fica representada por:

$$\begin{aligned} \forall t \in \mathbb{R}^d, N_{[s_i, m(s_i)]} &\stackrel{D}{=} N_{[s_i+t, m(s_i)]}, \\ \Leftrightarrow \{[s_1, m(s_1)], \dots, [s_n, m(s_n)]\} &\stackrel{D}{=} \{[s_1 + t, m(s_1)], \dots, [s_n + t, m(s_n)]\}. \end{aligned}$$

Um processo pontual é isotrópico se a distribuição espacial dos eventos é invariante sob rotação em torno da origem (Moller; Waagepetersen, 2003; Illian et al., 2008; Wiegand; Moloney, 2014; Scalon, 2024).

Seja $N = \{s_1, s_2, \dots, s_n\}$ um processo pontual e $R_\alpha N = \{R_\alpha s_1, R_\alpha s_2, \dots, R_\alpha s_n\}$ a sua rotação, onde $\alpha \in [0^\circ, 360^\circ]$. Este processo será isotrópico se a distribuição espacial dos eventos N e $R_\alpha N$ for a mesma, ou seja,

$$\forall \alpha \in [0^\circ, 360^\circ], N \stackrel{D}{=} R_\alpha N \Leftrightarrow \{s_1, s_2, \dots, s_n\} \stackrel{D}{=} \{R_\alpha s_1, R_\alpha s_2, \dots, R_\alpha s_n\}. \quad (2.2)$$

No caso do processo pontual ser marcado, onde $m(s_i)$ é a marca associada a cada evento, a expressão 2.2 fica representada por: $\forall \alpha \in [0^\circ, 360^\circ], N_{[s_i, m(s_i)]} \stackrel{D}{=} R_\alpha N_{\{[s_i, m(s_i)]\}}$,

$$\Leftrightarrow \{[s_1, m(s_1)], \dots, [s_n, m(s_n)]\} \stackrel{D}{=} \{R_\alpha [s_1, m(s_1)], \dots, R_\alpha [s_n, m(s_n)]\}.$$

Um processo pontual pode ser isotrópico sem ser estacionário, mas o contrário não (Baddeley; Rubak; Turner, 2015).

2.3.2 Homogeneidade, Independência e Tendência Espacial

Um processo pontual é considerado homogêneo quando os eventos ocorrem em proporções iguais na área de estudo. Isso significa que a probabilidade de encontrar um evento em qualquer sub-região da área de estudo onde ocorrem os eventos, é constante e igual. Em contraste, se essa probabilidade não for constante e variar, indicando que certas áreas são mais propensas a ter um maior ou menor número de eventos, ou se os eventos ocorrerem apenas em algumas sub-regiões específicas da área de estudo, diz-se que o processo pontual apresenta tendência espacial.

Pela teoria de probabilidade sabe-se que dois eventos S e Z são independentes se e somente se $P(S \cap Z) = P(S)P(Z)$, ou seja, eles serão independentes se e somente se sua probabilidade conjunta (ocorrência simultânea) for igual ao produto de suas probabilidades marginais (ocorrências individuais), como evidenciado por Ferreira (2020).

Da mesma forma, pode-se aplicar essa lógica ao conceito de independência de dois ou mais eventos em processos pontuais. Em um processo pontual, dois ou mais eventos são considerados independentes se a ocorrência de um evento em um determinado local ou sub-região da área de estudo não influencia a ocorrência de outros eventos.

2.4 Principais distribuições de probabilidade em processos pontuais

Nesta seção, apresentamos as principais distribuições de probabilidade, incluindo distribuição uniforme, binomial e Poisson.

2.4.1 Distribuição Uniforme

Seja B a região de estudo e $|B|$ a sua área ou volume correspondente. Diz-se que os eventos de um processo pontual qualquer estão uniformemente distribuídos em B , se a probabilidade

de ocorrência for constante na região B e zero fora dela, ou seja,

$$f(s_i) = \begin{cases} \frac{1}{|B|} & \text{se } (x,y) \in B \\ 0 & \text{se } (x,y) \notin B. \end{cases}$$

Considerando $S \subseteq B$, como uma sub-região de B , onde $|S|$ e $|B|$ são as áreas ou volumes correspondentes, a probabilidade dos eventos $s_i \in S$ ocorrerem na sub-região S é,

$$P(s \in S) = \int f(x,y) dx dy = \frac{1}{|B|} \int_S 1 dx dy = \frac{|S|}{|B|},$$

em que $s_i = (x,y)$.

2.4.2 Distribuição Binomial

Seja $S \subseteq B$ e $n(S \cap B) \in B$ representando os eventos no processo pontual S na região de estudo B , onde $|S|$ e $|B|$ são as áreas ou volumes correspondentes. Diz-se que os eventos aleatórios $n(S \cap B)$ do processo pontual S seguem distribuição binomial, se resultam de ensaios independentes e idênticos, com apenas dois resultados possíveis, sucesso (p) ou fracasso ($1 - p$), constantes em cada ensaio, ou seja,

$$P\{n(S \cap B) = \kappa\} = \binom{n}{\kappa} p^\kappa (1-p)^{n-\kappa}; \kappa = 0, 1, 2, \dots, n; p = \frac{|S|}{|B|}.$$

Sabe-se da teoria de probabilidade que o valor esperado (média) de uma distribuição binomial X é dado por $\mu = \mathbb{E}(X) = n \times p$, onde n é o número de ensaios realizados. Substituindo $p = \frac{|S|}{|B|}$ e X por $n(S \cap B)$, obtém-se o número médio de eventos de um processo pontual dado por,

$$\mu = \mathbb{E}[n(S \cap B)] = n \times \frac{|S|}{|B|} \Leftrightarrow \frac{\mathbb{E}[n(S \cap B)]}{|S|} = \frac{n}{|B|} \Rightarrow \hat{\lambda}(s) = \frac{n}{|B|},$$

onde $\hat{\lambda}(s)$ é denominado estimador da intensidade $\lambda(s)$, e corresponde ao número médio de eventos por unidade de área ou volume sob condição de homogeneidade.

2.4.3 Distribuição Poisson

A distribuição Poisson é comumente empregada para modelar a ocorrência de eventos raros e independentes em uma área de estudo. Assim, um processo pontual que segue a distri-

buição Poisson descreve o número de eventos raros que ocorrem em uma determinada área de estudo, e sua distribuição de probabilidade é dada por,

$$P(S = K) = \frac{\mu^k \times e^{-\mu}}{k!} = \frac{(\lambda(s)|S|)^k \times e^{-\lambda(s)|S|}}{k!}.$$

Um processo pontual que segue uma distribuição Poisson e é homogêneo, é denominado processo pontual Poisson homogêneo. Este processo é caracterizado por ser estacionário e isotrópico, razão pela qual é também denominado processo de aleatoriedade espacial completa (não exibe dependência espacial), e se a região B de estudo é dividida em S sub-regiões diferentes, a intensidade $\lambda(s)$ nessas S sub-regiões, estimada por $\hat{\lambda}(s)$ será constante (não variará), ou seja, $\forall s_i, s_j \in B : \lambda(s_i) \equiv \lambda(s_j)$. Isso decorre do fato de que a distribuição Poisson possui média e variância iguais, e sua relação com a distribuição Binomial é demonstrada em Lima (2005), Mateus (2013). Caso o processo pontual seja Poisson, mas apresente tendência, ele é denominado processo pontual Poisson não homogêneo. Nesse contexto, embora o processo não exiba dependência espacial, a intensidade $\lambda(s)$ nas S sub-regiões não é constante, ou seja, $\exists s_i \in B : \lambda(s_i) \neq \lambda(s_j)$ e, o número médio de eventos nessa situação é dado por, $\mu = \int_B \lambda(s)$.

2.5 Propriedade de primeira e segunda ordem

Nesta seção, apresentamos as propriedades (intensidade) de primeira e segunda ordem.

2.5.1 Propriedade de primeira ordem

Sejam S_i e S_j sub-regiões de B contendo os eventos $s_i = (x, y)$ e $s_j = (x', y')$, onde o número de eventos é dado por $n(s_i)$ e $n(s_j)$ e sua área dada por $|S_i|$ e $|S_j|$ respectivamente. A propriedade de primeira ordem, também designada intensidade $\lambda(s_i)$, corresponde o número de eventos esperados por unidade de área, ou seja, $\lambda(s_i) = \lim_{|S_i| \rightarrow 0} \left\{ \frac{\mathbb{E}[n(s_i)]}{|S_i|} \right\}$, conforme descrito por Mateus (2013), Pebesma & Bivand (2023), Moraga (2023). Se o processo pontual for homogêneo, a intensidade $\lambda(s_i)$ é constante, e dado por, $\lambda(s_i) = \lambda = \frac{\mathbb{E}[n(s_i)]}{|S_i|}$, $\hat{\lambda} = \frac{n(s_i)}{|S_i|}$.

Uma maneira de verificar se a intensidade é constante consiste em dividir a área de estudo em sub-regiões ou *quadrats* do mesmo tamanho (área), contar os eventos em cada sub-região ou *quadrats* e dividir pela respectiva área da sub-região ou *quadrats*. No entanto, este procedimento resulta em uma intensidade não suavizada (Figura 2.1 (a)).

Outra alternativa é usar uma janela móvel de tamanho fixo, centrada em vários locais da área de estudo. Ao mover essa janela pela região, os eventos são contados e divididos pela área da janela, proporcionando uma estimativa mais suave e interpretável da variação da intensidade $\lambda(s_i)$ (Figura 2.1 (b)). No entanto, em cada uma das estimativas de intensidade (com janela móvel ou sub-regiões do mesmo tamanho), não se considera a localização relativa dos eventos dentro janela/sub-região (se os eventos estão próximos ou afastados) e a escolha de um tamanho de janela adequado não é clara (Gatrell et al., 1996).

Vale ressaltar que a variação na intensidade é mais significativa em processos pontuais não homogêneos. Assim, no caso do processo pontual não homogêneo, uma maneira de estimar a função intensidade $\lambda(s_i)$, é usando os estimadores não paramétricos de kernel, onde a janela ou *quadrats* outrora descrita (o), é substituída por uma função (Kernel) que descreve um objeto tridimensional móvel, e que pondera os eventos dentro de sua esfera de influência consoante a distância do ponto onde a intensidade está sendo estimada (Gatrell et al., 1996) (Figura 2.1 (c)). Essencialmente, neste método estima-se a função de densidade de probabilidade $f(z)$ e não função intensidade $\lambda(s_i)$.

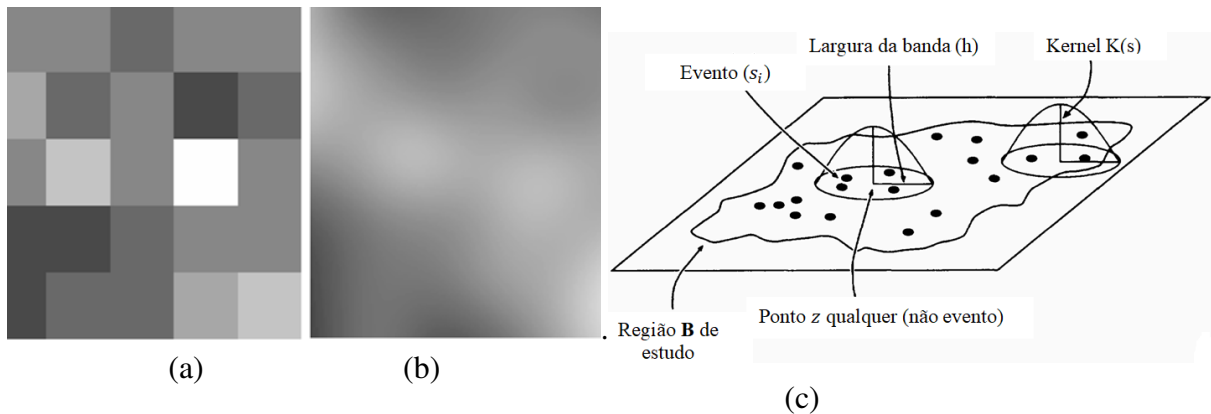
A função de densidade de probabilidade descreve a probabilidade de observar um evento em uma determinada região de estudo e esta função é não-negativa e integra para 1, enquanto a função intensidade $\lambda(s_i)$ não é uma medida de probabilidade, mas sim uma medida da taxa de ocorrência de eventos em uma determinada área, que embora seja não-negativa, não precisa integrar para 1, pois ela não descreve uma distribuição de probabilidade (Moraga, 2023).

Embora a função de densidade de probabilidade $f(z)$ e a função intensidade $\lambda(s_i)$ não sejam iguais, estas são proporcionais, o que significa que locais com maior taxa de ocorrências dos eventos (intensidade), terão também maior densidade de probabilidade $f(z)$. Assim, a função densidade de probabilidade e função intensidade são dadas por,

$$\lambda(z) = f(z) \int_B \lambda(s_i) ds_i, \hat{f}(z) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{z-s_i}{h}\right) \text{ e } \hat{\lambda}(z) = \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{z-s_i}{h}\right),$$

onde $\int_B \lambda(s_i) ds_i$ é o número esperado de eventos na região S_i , $\hat{f}(z)$ e $\hat{\lambda}(z)$ são respectivamente estimadores Kernel não paramétricos da função densidade e intensidade, no local z da região B em estudo, com base nos eventos s_i . $K(s)$ é uma função densidade de probabilidade simétrica tal que $\forall s, K(s) \geq 0$, com $\int_B K(s) = 1$, designada função Kernel e h é um parâmetro de suavização ou largura da banda (Figura 2.1(c)).

Figura 2.1 – Intensidade: (a) não suavizada, (b) suavizada e (c) estimativa de Kernel.

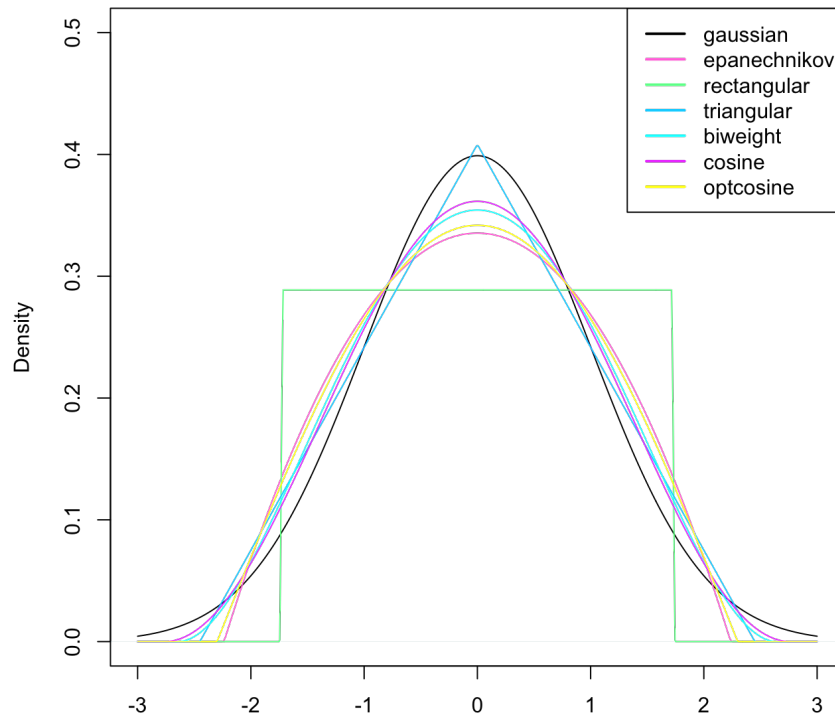


Fonte: Adaptado de Gatrell et al. (1996), Baddeley, Rubak & Turner (2015)

Conforme descrito por Gatrell et al. (1996), Gelfand et al. (2010), o estimador de Kernel $K(s)$, é sensível à escolha da largura de banda h , de tal modo que, um valor maior de h resulta em maior suavização na variação espacial da intensidade (viés aumenta e a variância diminui), enquanto menor valor de h resulta em menor suavização (viés diminui e a variância aumenta). Portanto, testar várias larguras de banda é uma opção viável para selecionar o valor ideal do parâmetro de suavização, mas também existem métodos computacionais que permitem uma escolha mais eficiente, como a validação cruzada por verossimilhança ou outros descritos por Cronie & Lieshout (2016). As escolhas comuns para o Kernel $k(s)$ incluem $K(s) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s^2}{2}\right)$, $K(s) = \frac{3}{4}(1-s^2)\mathbb{I}(|s| < 1)$, $K(s) = \frac{15}{16}(1-s^2)^2\mathbb{I}(|s| < 1)$, $K(s) = \frac{1}{2}\mathbb{I}(|s| < 1)$, que correspondem a kernel Gaussiano, Kernel Epanechnikov, Kernel Quadrático e Kernel Uniforme, respectivamente. Além destas, existem outras na literatura e algumas são apresentadas na Figura 2.2.

A escolha do Kernel dependerá da pesquisa e do pesquisador, tomando em consideração que o Kernel Uniforme atribui peso igual a todos os eventos em uma distância fixa do evento de referência, sendo útil quando a distribuição dos eventos é uniforme, embora seja menos sensível a variações na densidade dos dados. O Kernel Gaussiano atribui pesos baseados na distribuição gaussiana (normal), capturando padrões suaves e contínuos nos dados e sendo adequado para distribuições não uniformes, embora mais sensível a variações locais. O Kernel Epanechnikov, com sua forma parabólica e suporte compacto, é mais eficiente que o Gaussiano, equilibrando características locais e evitando sensibilidade excessiva ao ruído. O Kernel Quadrático, também parabólico, oferece um compromisso entre os Kernels Uniforme e Gaussiano, sendo menos sensível a outliers que o Gaussiano, mas mais sensível que o Uniforme, ideal para equilibrar robustez e suavidade (García-Portugués, 2024; Scalón, 2024).

Figura 2.2 – Funções kernel



Fonte: García-Portugués (2024)

Conforme descrito por Moraga (2023), Scalon (2024), os efeitos de borda tendem a distorcer as estimativas do Kernel, próximo à fronteira (margens) da região de estudo, uma vez que os eventos próximos à fronteira (margens) têm menos vizinhos locais do que os eventos no interior. Uma maneira de lidar com esse problema é modificar a estimativa do Kernel dividindo-a pelo seguinte termo de correção de borda $e_h(s) = \int_B h^{-2} K(\frac{z-s}{h}) dx$, que representa o volume sob o Kernel centrado em z que está na região de estudo B (Gatrell et al., 1996).

Conforme descrito por Baddeley, Rubak & Turner (2015), dentre os estimadores de Kernel usuais da função de intensidade destacam-se, $\hat{\lambda}^0(z) = \sum_{i=1}^n K(z-s_i)$, $\hat{\lambda}^U(z) = \frac{1}{e(z)} \sum_{i=1}^n K(z-s_i)$ e $\hat{\lambda}^D(z) = \sum_{i=1}^n \frac{1}{e(s_i)} K(z-s_i)$, e correspondem a estimador não corrigido, estimador uniformemente corrigido e estimador com correção de Diggle, respectivamente. Nestes estimadores, $K(\cdot)$ é função densidade de probabilidade Kernel e $e(s_i)$ a correção da borda, dada por $e(s_i) = \int_B K(z-s) dx$.

Além do estimador não paramétricos de kernel, outra maneira de estimar a função de intensidade $\lambda(s_i)$, é usar estimadores paramétricos (2.7).

2.5.2 Propriedade de segunda ordem

Segundo Mateus (2013), Pebesma & Bivand (2023), a propriedade de segunda ordem, também conhecida como intensidade de segunda ordem ou densidade, descreve a forma como os eventos estão distribuídos no espaço, podendo indicar se estão distribuídos independentemente uns dos outros (aleatoriedade espacial completa), se tendem a se agrupar (agrupamento) ou se repelem mutuamente (distribuição mais regular do que sob aleatoriedade espacial completa) e, é representada por $\lambda(s_i, s_j) = \lim_{|S_i| \rightarrow 0} \lim_{|S_j| \rightarrow 0} \frac{\mathbb{E}[n(s_i)n(s_j)]}{|S_i||S_j|}$, e estimada usando as funções correlação par $g(r)$, função $L(r)$ e função $K(r)$ ou variantes não homogêneas dessas funções.

Segundo Gatrell et al. (1996), Diggle (2013), se o processo pontual for estacionário, a propriedade de segunda ordem dependerá apenas da diferença vetorial d (direção e distância), entre s_i e s_j , e não de suas localizações absolutas, ou seja, $\lambda(s_i, s_j) \equiv \lambda(\|s_i - s_j\|)$. Todavia, se for estacionário e isotrópico, a intensidade de segunda ordem dependerá apenas da distância $d(s_i, s_j)$ e não da sua orientação ou direção.

Além da intensidade de primeira e segunda ordem, existe a chamada intensidade condicional.

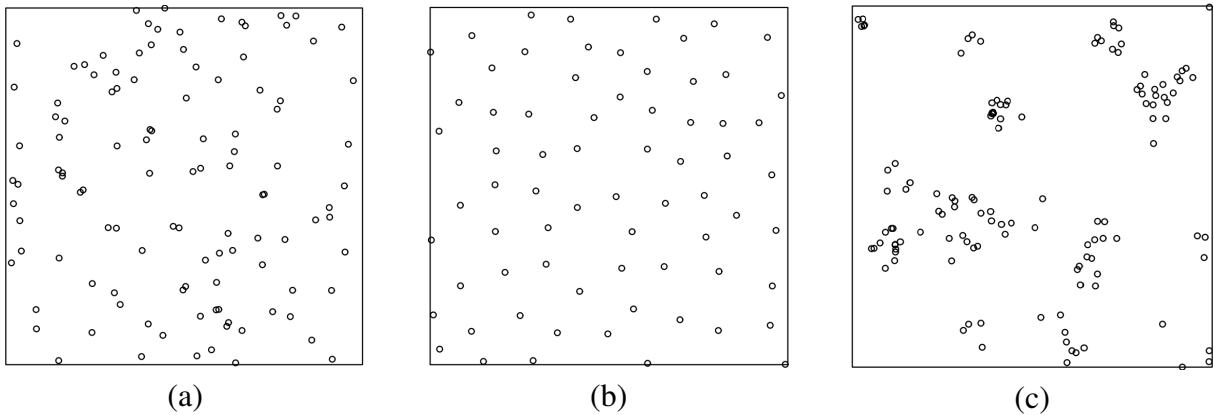
Segundo Diggle (2013), a intensidade condicional corresponde à intensidade do evento s_i condicionada à (dada) informação do evento s_j , ou seja, $\lambda(s_i|s_j) = \frac{\lambda(s_i, s_j)}{\lambda(s_j)}$. Pela propriedade de independência entre os eventos, descrita na seção 2.3.2, se os eventos s_i e s_j forem independentes, a intensidade condicional $\lambda(s_i|s_j)$ será dada por $\lambda(s_i|s_j) = \frac{\lambda(s_i, s_j)}{\lambda(s_j)} = \lambda(s_i)$.

2.6 Padrões de distribuição espacial e sua identificação

Um processo pontual pode exibir três padrões espaciais distintos (Figura 2.3), que são, o aleatório, onde os eventos (representados por "o") estão distribuídos de forma aleatória, com média e variância iguais; regular, quando os eventos estão espaçados de maneira uniforme em comparação com um padrão aleatório, resultando em uma variância menor que a média (subdispersão); e agrupado, quando os eventos estão mais próximos uns dos outros do que seria esperado em um padrão aleatório, resultando em uma variância maior que a média (superdispersão) (Lima, 2005; Scalon, 2024).

Para identificar os padrões apresentados na Figura 2.3, há várias abordagens, que vão desde estatísticas simples como o índice de Morisita (1959), o índice de Clark & Evans (1954), o índice de Hopkins & Skellam (1954), até o uso de funções (Scalon, 2024).

Figura 2.3 – Padrões de Distribuição Espacial: (a) aleatório, (b) regular e (c) agrupado.



Fonte: Oliveira (2022)

Neste estudo, prioriza-se o uso de funções, pois estas permitem analisar os padrões em diferentes distâncias (r) e podem ser visualizadas graficamente, o que não é possível com simples estatísticas. Entre as funções, destacam-se as funções sumárias homogêneas de distância, $F(r)$, $G(r)$, $J(r)$, bem como as funções sumárias homogêneas de segunda ordem $K(r)$, $L(r)$ e $g(r)$, ou suas extensões não homogêneas. Essas funções são comparadas com um processo pontual Poisson homogêneo $F_{pois}(r) = G_{pois}(r) = 1 - e^{-\pi\lambda r^2}$ e $J_{pois}(r) = 1$ ou $K_{pois}(r) = \pi r^2$, $L_{pois}(r) = r$ e $g_{pois}(r) = 1$. Se forem equivalentes, diz-se que o padrão de distribuição espacial é de aleatoriedade espacial completa (AEC), caso contrário, pode ser regular (Inibição) ou agregado (Atração) conforme descrito no Quadro 2.1.

2.6.1 Funções $F(r)$, $G(r)$, $J(r)$, $K(r)$, $L(r)$ e $g(r)$

Nesta seção, apresentamos as funções $F(r)$, $G(r)$, $J(r)$, $K(r)$, $L(r)$ e $g(r)$ homogêneas, úteis para identificação do padrão de distribuição (regular, aleatório ou agrupado) dos processos pontuais.

2.6.1.1 Função $F(r)$

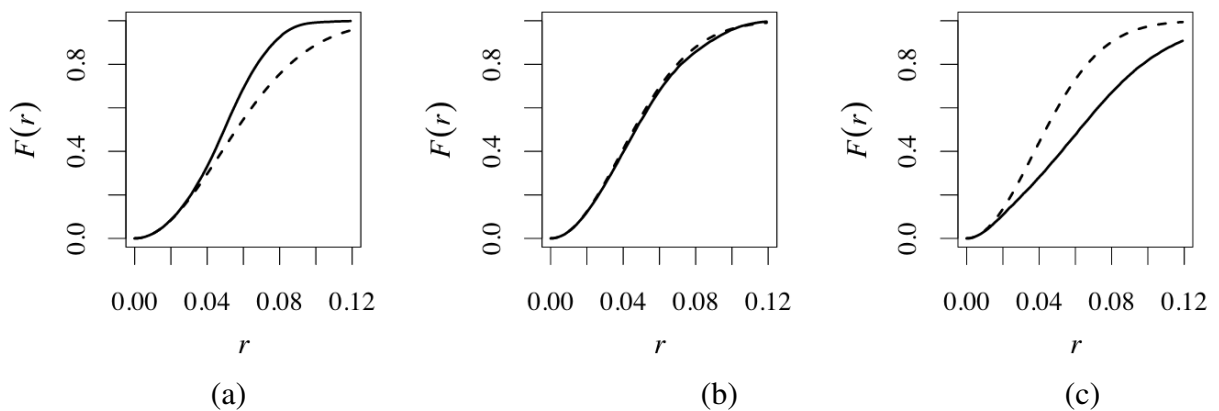
Em um processo pontual S , a distância $d(z, s_i) = \min\{\|z - s_i\| : s_i \in S\}$ representa a menor distância de quaisquer pontos fixos imaginários $z \in \mathbb{R}^2$ para os i -ésimos eventos s_i de estudo mais próximos, conforme ilustrado na Figura 2.6 (b). Esta distância é comumente referida como a distância do espaço vazio e a função que representa todas as possíveis distâncias (função cumulativa) é denominada função $F(r)$ e é definida como:

$$F(r) = P\{d(z, s_i) \leq r\} \text{ e } \hat{F}(r) = \frac{1}{n(z)} \sum_{k=1}^K \mathbb{I}\{d(z_k, s_i) \leq r\}, \forall r \geq 0, \quad (2.3)$$

onde $\hat{F}(r)$ é seu estimador, \mathbb{I} é função indicadora e $d(z_k, s_i)$ a distância real entre k -ésimos pontos z_k (não eventos) e i -ésimos eventos s_i mais próximos (Gelfand et al., 2010; Diggle, 2003; Baddeley; Rubak; Turner, 2015; Scalon, 2024).

Se a curva $\hat{F}(r)$ (representada por “—”) estiver acima da curva $F_{pois}(r)$ (representada por “- -”), ou seja, $\hat{F}(r) > F_{pois}(r)$, indica que a distância entre quaisquer pontos z até o evento mais próximo s_i é menor do que seria sob aleatoriedade espacial completa. Isto sugere que existe menor espaço vazio (não ocupado pelos eventos) e, conseqüentemente, o padrão espacial dos eventos é regular (Figura 2.4 (a)). Se $\hat{F}(r)$ é equivalente a $F_{pois}(r)$ ($\hat{F}(r) \equiv F_{pois}(r)$), o padrão é considerado aleatório, também designado aleatoriedade espacial completa (AEC) (Figura 2.4 (b)). Se a curva $\hat{F}(r)$ estiver abaixo da curva $F_{pois}(r)$ ($\hat{F}(r) < F_{pois}(r)$), significa que a distância entre quaisquer pontos z até o evento mais próximo s_i é maior do que seria sob aleatoriedade espacial completa. Nesse caso, há mais espaço vazio e, conseqüentemente, o padrão espacial dos eventos é agrupado (Figura 2.4 (c)).

Figura 2.4 – Padrões de distribuição espacial na função $F(r)$: (a) regular, (b) aleatório e (c) agrupado.



Fonte: Baddeley, Rubak & Turner (2015)

Considerando a região de estudo S , uma sub-região de uma área infinita, o estimador $\hat{F}(r)$, representado pela expressão 2.3, apresentará viés. Esse viés ocorre porque a distância $d(z_k, s_i)$ de um ponto z_k localizado na borda (limite da área de estudo) até o evento mais próximo s_i pode parecer maior do que a distância r , porque na sua determinação, não são considerados os eventos fora da sub-região em estudo (Scalon, 2024).

No entanto, se todos os eventos existentes na sub-região fora da área de estudo fossem levados em consideração, essa distância seria menor ou igual a r (Figura 2.6 (c)) (Diggle, 2003; Leininger, 2014). Esse viés é conhecido como o “efeito da borda”. A correção para esse viés é denominada “correção da borda” (e_k), que corresponde à distância de um ponto z_k até o limite

da área de estudo B (borda). Portanto, um estimador que não apresenta viés para essa situação é, $\hat{F}_{bord}(r) = \frac{\sum_k \mathbb{I}\{d(z_k, s_i) \leq r\} \mathbb{I}\{e_k > r\}}{\sum_k \mathbb{I}\{e_k > r\}}$, onde $e_k = d(z_k, S \cap B)$, $r \geq 0$, onde $d(z_k, S \cap B)$ representa a distância entre quaisquer pontos amostrais (não eventos) z_k para o evento $s_i \in S$ mais próximo, considerando a existência dos efeitos borda na região B de estudo.

2.6.1.2 Função $G(r)$

Diferentemente da função $F(r)$, que mede a distância entre um ponto e um evento, a função $G(r)$ mede a distância entre eventos (Figura 2.6 (a)). Seja s_i um evento de um processo pontual S , a distância até o vizinho mais próximo $d_{\min}(s_i, s_{i'}) = \min_{i' \neq i} \|s_{i'} - s_i\|$ pode ser expressa por $d_{\min}(s_i, s_{i'}) = d(s_i, s_{i'} \setminus s_i)$, que corresponde à distância mínima de um evento s_i até os outros eventos $s_{i'}$ diferentes de s_i (Baddeley; Rubak; Turner, 2015; Scalon, 2024).

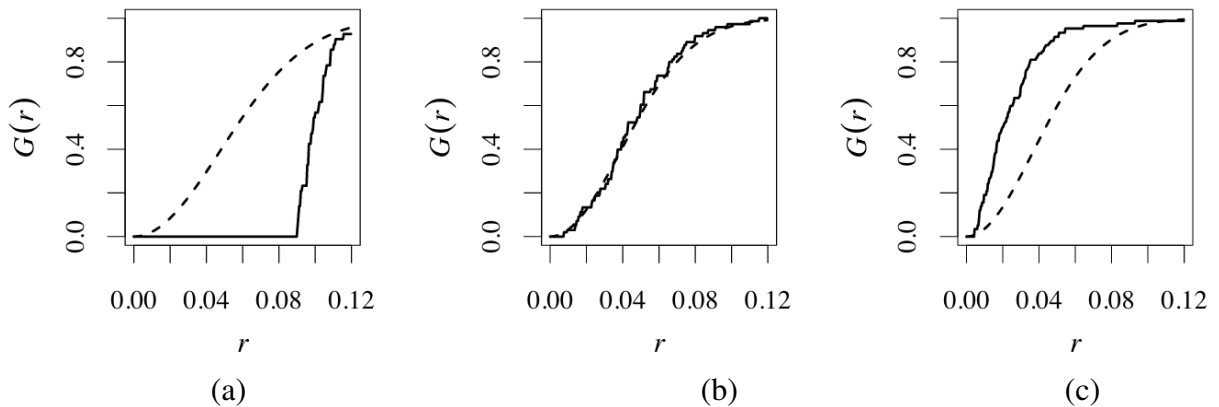
Conforme mencionado por Diggle (2003), Leininger (2014), Scalon (2024), considerando $n(s)$ como o número de eventos S em uma região de estudo B e $d_{\min}(s_i, s_{i'}) = d(s_i, s_{i'} \setminus s_i)$ como a distância do i -ésimo evento s_i até o evento mais próximo $s_i \setminus s_{i'}$, $d_{\min}(s_i, s_{i'})$ é designada como a distância do vizinho mais próximo. Esta medida inclui distâncias repetidas em pares de vizinhos mais próximos recíprocos $\{d(s_i, s_{i'} \setminus s_i) \text{ e } d(s_{i'} \setminus s_i, s_i)\}$, ou seja,

$$G(r) = P\{d(s_i, s_{i'} \setminus s_i) \leq r\} \quad e \quad \hat{G}(r) = \frac{1}{n(s)} \sum_i \mathbb{I}\{d(s_i, s_{i'} \setminus s_i) \leq r\}, \quad r \geq 0, \quad (2.4)$$

onde $\hat{G}(r)$ é respectivo estimador.

Se a curva $\hat{G}(r)$ (representada por “—”) estiver abaixo da curva $G_{pois}(r)$ (representada por “- - -”), ou seja, $\hat{G}(r) < G_{pois}(r)$, indica que a distância entre quaisquer eventos s_i até os eventos $s_{i'}$ mais próximo é maior do que seria sob aleatoriedade espacial completa. Nesse caso, há maior distância de separação entre os eventos e, conseqüentemente, o padrão espacial dos eventos é regular (Figura 2.5 (a)). Se $\hat{G}(r)$ é equivalente a $G_{pois}(r)$ ($\hat{G}(r) \equiv G_{pois}(r)$), o padrão é considerado aleatório, também designado aleatoriedade espacial completa (AEC) (Figura 2.5 (b)). Se a curva $\hat{G}(r)$ estiver acima da curva $G_{pois}(r)$ ($\hat{G}(r) > G_{pois}(r)$), significa que a distância entre quaisquer eventos s_i até os eventos $s_{i'}$ mais próximo é menor do que seria sob aleatoriedade espacial completa. Isto sugere que existe menor distância de separação entre os eventos e, conseqüentemente, o padrão espacial dos eventos é agrupado (Figura 2.5 (c)).

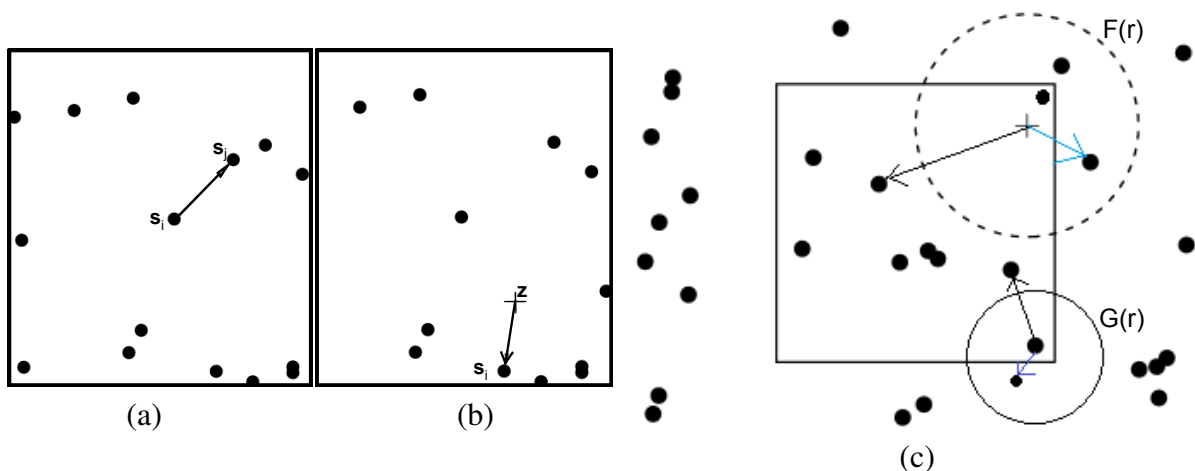
Figura 2.5 – Padrões de distribuição espacial na função $G(r)$: (a) regular, (b) aleatório e (c) agrupado.



Fonte: Baddeley, Rubak & Turner (2015)

Considerando a região de estudo S , uma sub-área de uma infinita área B , o estimador $\hat{G}(r)$ representado pela expressão 2.4, apresentará viés, pois a distância $d(s_i, s_j \setminus s_i)$ de um evento s_i que esteja na borda para o evento s_j mais próximo pode aparentemente ser maior que a distância r , enquanto na realidade, se fossem considerados os eventos que estão fora da área de estudo, esta distância seria menor ou igual a r (Figura 2.6 (c)). Este viés ocorre porque a distância $d(s_i, s_j \setminus s_i)$ de qualquer evento em B , para o evento mais próximo que esteja fora da área de estudo não é considerada, resultando em um viés. Portanto, um estimador que não apresentara viés para esta situação em concreto é, $\hat{G}_{bord}(r) = \frac{\sum_i \mathbb{I}\{e_i \geq r \cap d(s_i, s_j \setminus s_i) \leq r\}}{\sum_i \mathbb{I}\{e_i \geq r\}}$, onde $e_i = d(s_i, S \cap B) \quad r \geq 0$.

Figura 2.6 – (a) distância de um evento (s_i) ao evento vizinho (s_j) mais próximo ($G(r)$), sem efeito da borda; (b) distância de um espaço vazio (z) ao evento s_i mais próximo ($F(r)$), sem efeito da borda; (c) demonstração da situação em que o efeito da borda está presente.



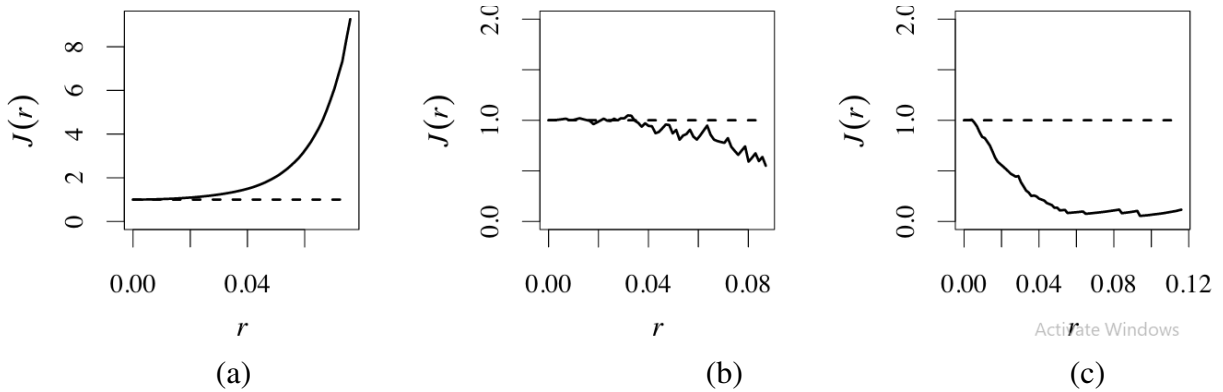
Fonte: Adaptado de Baddeley, Rubak & Turner (2015)

2.6.1.3 Função $J(r)$

A função $J(r)$ é uma função que não apresenta viés e resulta da razão entre as funções $G(r)$ e $F(r)$ acima descritas, ou seja, $J(r) = \frac{1-G(r)}{1-F(r)}$ e $\hat{J}(r) = \frac{1-\hat{G}(r)}{1-\hat{F}(r)} = \frac{1-\frac{1}{n(s)} \sum_i \mathbb{I}\{d(s_i, s'_i \setminus s_i) \leq r\}}{1-\frac{1}{j} \sum_{j=1}^j \mathbb{I}\{d(z_j, s'_j) \leq r\}}$, $F(r), \hat{F}(r) < 1$, onde $\hat{J}(r)$ é o respectivo estimador (Lieshout; Baddeley, 1996; Scalon, 2024).

Como $F_{pois}(r) = G_{pois}(r) = 1 - e^{-\pi\lambda r^2}$, $J_{pois}(r) = \frac{1-G_{pois}(r)}{1-F_{pois}(r)} = \frac{1-(1-e^{-\pi\lambda r^2})}{1-(1-e^{-\pi\lambda r^2})} = 1$, consequentemente, se $\hat{J}(r) > 1$, o processo pontual apresenta um padrão regular (Figura 2.7 (a)). Se $\hat{J}(r) \equiv 1$ o processo pontual apresenta um padrão aleatório (Figura 2.7 (b)). Se $\hat{J}(r) < 1$ o processo pontual apresenta um padrão agrupado (Figura 2.7 (c)).

Figura 2.7 – Padrões de distribuição espacial na função $J(r)$: (a) regular, (b) aleatório e (c) agrupado.



Fonte: Baddeley, Rubak & Turner (2015)

2.6.1.4 Função $K(r)$ ou função κ de Ripley

A função de $K(r)$ ou função κ de Ripley estima o número esperado de r -ésimos vizinhos de um evento S , dividido pela intensidade $\lambda(s)$. Em outras palavras, esta função estima o número médio de eventos ($\mathbb{E}[d(s'_i, r, s_i) | s'_i \in S]$) que estão contidos em um círculo de raio r , centrado em um evento s_i de referência, sem contar o próprio evento s_i (r é a distância entre o evento de referência s_i até outro evento s'_i). Em seguida, número médio de eventos estimados é dividido pela intensidade $\lambda(s)$ (Figura 2.8), ou seja,

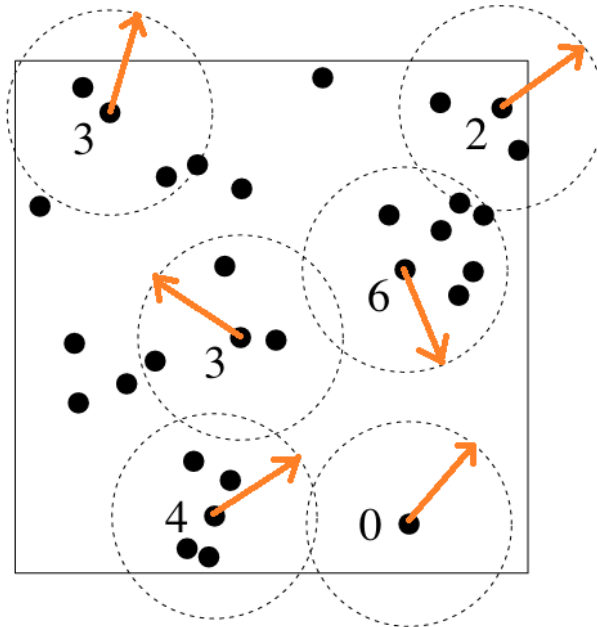
$$K(r) = \frac{\mathbb{E}[d(s_i, r, s'_i) \leq r | s'_i \in S]}{\lambda(s_i)\lambda(s'_i)} = \frac{\mathbb{E}[\sum_{i' \neq i} \mathbb{I}\{0 < \|s_i - s'_i\| \leq r\} | s'_i \in S]}{\lambda(s_i)\lambda(s'_i)}, \lambda(s) > 0, r \geq 0, \quad (2.5)$$

$$\hat{K}(r) = \frac{1}{|B|} \sum_{i=1} \sum_{i' \neq i} \frac{\mathbb{I}[\|s_i - s'_i\| \leq r]}{\hat{\lambda}(s_i)\hat{\lambda}(s'_i)} \text{ ou } \hat{K}_{bord}(r) = \frac{1}{|B|} \sum_{i=1} \sum_{i' \neq i} \frac{\mathbb{I}[\|s_i - s'_i\| \leq r]}{\hat{\lambda}(s_i)\hat{\lambda}(s'_i)e_{ii'}}$$

onde $\hat{K}(r)$ e $\hat{K}_{bord}(r)$ são respectivos estimadores na ausência e na presença dos efeitos da borda respectivamente. $e_{ii'}$ é correção da borda, $\mathbb{I}[\cdot]$ função indicadora, $|B|$ área de região em estudo e

$\hat{\lambda}(s) = n/|B|$ é intensidade (Cressie, 1993; Gatrell et al., 1996; Diggle, 2013; Leininger, 2014; González; Moraga, 2023; Scalón, 2024).

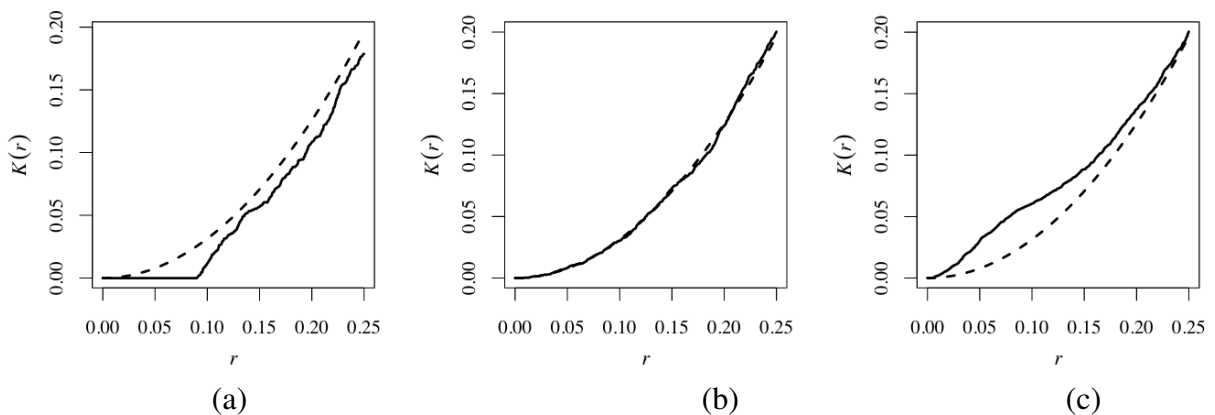
Figura 2.8 – Ideia intuitiva por trás da estimação da função $K(r)$.



Fonte: Adaptado de Baddeley, Rubak & Turner (2015)

A interpretação da função $K(r)$ é idêntica à função $G(r)$. Se $K(r) < K_{pois}(r)$ (abaixo de), significa que na distância considerada, há poucos eventos contidos no círculo do que seria esperado se o padrão fosse aleatório. Consequentemente, o padrão é regular (Figura 2.9 (a)). Se $K(r) \equiv K_{pois}$, o padrão é aleatório (Figura 2.9 (b)). Se $K(r) > K_{pois}$ (acima de), significa que na distância (r) considerada, existem muitos eventos contidos no círculo do que seria esperado se o padrão fosse aleatório. Consequentemente, o padrão é agrupado (Figura 2.9 (c)).

Figura 2.9 – Padrões de distribuição espacial na função $K(r)$: (a) regular, (b) aleatório e (c) agrupado.

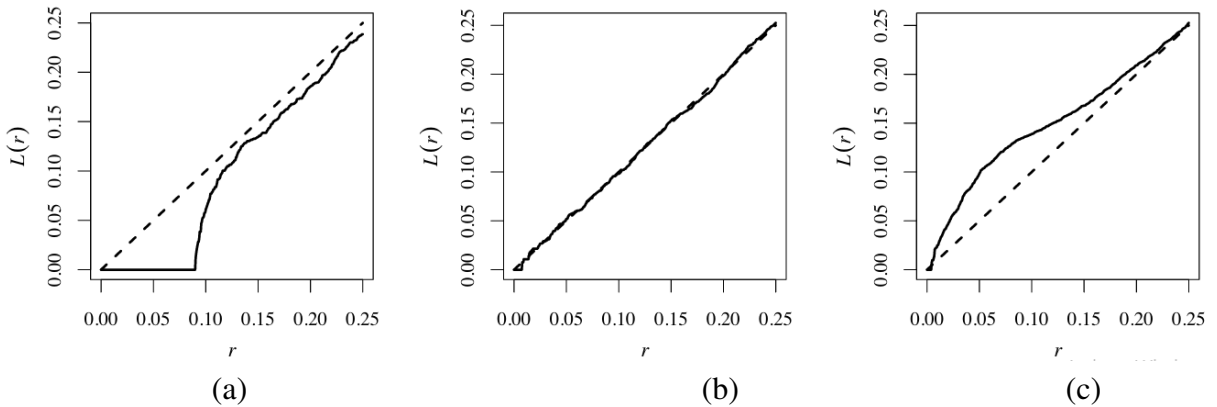


Fonte: Baddeley, Rubak & Turner (2015)

2.6.1.5 Função $L(r)$

Conforme descrito em Scalon (2024), a função $L(r)$ é uma transformação da função $K(r)$ sugerida por Besag (1977), visando transformar o modelo Poisson teórico (representada por “- - -”) em uma linha reta, bem como estabilizar a variância nos dados sob completa aleatoriedade espacial (Figura 2.10), ou seja, $L(r) = \sqrt{\frac{K(r)}{\pi}}$ e $\hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}}$.

Figura 2.10 – Padrões de distribuição espacial na função $L(r)$: (a) regular, (b) aleatório e (c) agrupado.



Fonte: Baddeley, Rubak & Turner (2015)

Como $L_{pois}(r) = r$, que corresponde a uma linha reta de inclinação 1, é comum subtrair r de $L(r)$, resultando em $L_{pois}(r) - r = 0$, o que permite estudar o padrão espacial em torno de zero, conforme será detalhado na seção dos resultados.

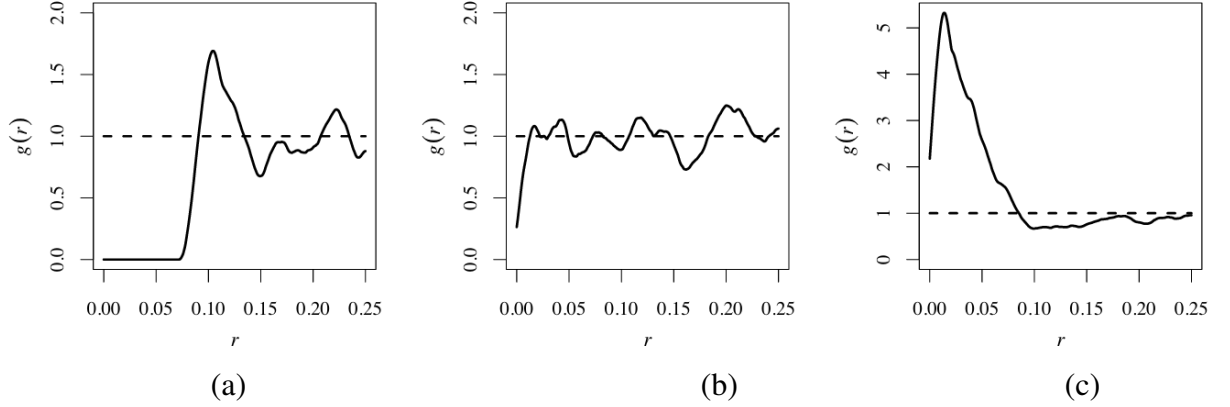
Em qualquer uma das possibilidades aqui apresentadas, se $\hat{L}(r) < L_{pois}(r)$, o processo pontual apresenta um padrão regular (Figura 2.9 (a)). Se $\hat{L}(r) \equiv L_{pois}(r)$ o processo pontual apresenta um padrão aleatório (Figura 2.9 (b)). Se $\hat{L}(r) > L_{pois}(r)$ o processo pontual apresenta um padrão agrupado (Figura 2.9 (c)).

2.6.1.6 Função $g(r)$

A função $g(r)$, também conhecida como função correlação par, é uma função de densidade de probabilidade padronizada que descreve a ocorrência conjunta dos eventos s_i e $s_{i'}$ em uma determinada região de estudo e, diferentemente de todas as funções já apresentadas, não é cumulativa (Figura 2.11), ou seja, $g(r) = \frac{\lambda(s_i, s_{i'})}{\lambda(s_i)\lambda(s_{i'})}$ e $\hat{g}_{bord}(r) = \frac{1}{2\pi r|B|} \sum_{i=1}^n \sum_j \frac{K_h(\|s_i - s_{i'}\| - r)}{\hat{\lambda}(s_i)\hat{\lambda}(s_{i'})e_{ii'}}$, onde $g(r) \neq G(r)$, $\hat{g}_{bord}(r)$ é seu estimador na presença dos efeitos da borda, K_h é função Kernel, h é largura da banda e $e_{ii'}$ correção da borda. Na ausência destes, $\hat{g}(r)$ é obtido omitindo $e_{ii'}$ em $\hat{g}_{bord}(r)$ (Diggle, 2010; González; Moraga, 2023; Scalon, 2024).

Segundo Waagepetersen & Guan (2009), Scalon (2024), se o processo pontual for isotrópico, a função $g(r)$ é dada pela razão entre a derivada da função $K(r)$ e $2\pi r$, ou seja, $g(r) = \frac{K'(r)}{2\pi r}$ e $\hat{g}(r) = \frac{\hat{K}'(r)}{2\pi r}$.

Figura 2.11 – Padrões de distribuição espacial na função $g(r)$: (a) regular, (b) aleatório e (c) agrupado.



Fonte: Baddeley, Rubak & Turner (2015)

Se $\hat{g}(r) < 1$, o processo pontual apresenta um padrão regular (Figura 2.11 (a)). Se $\hat{g}(r) \equiv 1$ o processo pontual apresenta um padrão aleatório (Figura 2.11 (b)). Se $\hat{g}(r) > 1$ o processo pontual apresenta um padrão agrupado (Figura 2.11 (c)).

2.6.2 Funções $F_{inhom}(r)$, $G_{inhom}(r)$, $J_{inhom}(r)$, $K_{inhom}(r)$, $L_{inhom}(r)$ e $g_{inhom}(r)$

As funções $F(r)$, $G(r)$, $J(r)$, $K(r)$, $L(r)$ e $g(r)$ apresentadas na seção 2.6.1 são úteis quando a condição de estacionariedade ou homogeneidade é satisfeita. Caso contrário, elas têm sua extensão para processos pontuais não homogêneos, sendo, $F_{inhom}(r) = 1 - \mathbb{E} \left[\prod_{s_i} \left(1 - \frac{\lambda_{\min}}{\lambda(s_i)} \right) \right]$,

$$\hat{F}_{inhom}(r) = 1 - \frac{\sum_i \left(\mathbb{I}(w_i > r) \left[\prod_i \mathbb{I}(\|s_i - z_i\| \leq r) \left(1 - \frac{\lambda_{\min}}{\lambda(s_i)} \right) \right] \right)}{\sum_i \mathbb{I}(w_i > r)}, \text{ para a função } F(r) \text{ não homogênea.}$$

$$G_{inhom}(r) = 1 - \mathbb{E} \left[\prod_{s_i} \left(1 - \frac{\lambda_{\min}}{\lambda(s_i)} \right) \middle| z \in S \right], \hat{G}_{inhom}(r) = 1 - \frac{\sum_i \left(\mathbb{I}(e_i > r) \left[\prod_i \mathbb{I}(\|s_i - s'_i\| \leq r) \left(1 - \frac{\lambda_{\min}}{\lambda(s_i)} \right) \right] \right)}{\sum_i \mathbb{I}(e_i > r)},$$

para a função $G(r)$ não homogênea ($G_{inhom}(r)$).

$$J_{inhom}(r) = \frac{1 - G_{inhom}(r)}{1 - F_{inhom}(r)}, r \geq 0, F_{inhom}(r) > 1, \text{ para a função } J(r) \text{ não homogênea } (J_{inhom}(r)).$$

$$K_{inhom}(r) = \frac{1}{|B|} \mathbb{E} \left[\sum_{s_i} \sum_{s'_j \setminus \{s_i\}} \frac{\mathbb{I}(\|s_i - s'_j\| \leq r)}{\lambda(s_i)\lambda(s'_j)} \right] \text{ e } \hat{K}_{inhom}(r) = \frac{1}{|B|} \sum_i \sum_{i'} \frac{\mathbb{I}(\|s_i - s'_i\| \leq r)}{\hat{\lambda}(s_i)\hat{\lambda}(s'_i)e_{ii}}, \text{ para a função}$$

$K(r)$ não homogênea ($K_{inhom}(r)$).

$$L_{inhom}(r) = \sqrt{\frac{K_{inhom}(r)}{\pi}}, \hat{L}_{inhom}(r) = \sqrt{\frac{\hat{K}_{inhom}(r)}{\pi}}, \text{ para a função } L(r) \text{ não homogênea } (L_{inhom}(r)) \text{ e,}$$

$$g_{inhom}(r) = \frac{K'_{inhom}(r)}{2\pi r} \text{ e } \hat{g}_{inhom}(r) = \frac{\hat{K}'_{inhom}(r)}{2\pi r}, \text{ para a função } g(r) \text{ não homogênea } (g_{inhom}(r)).$$

É comum encontrar na literatura, como em Baddeley, Møller & Waagepetersen (2000), Baddeley, Rubak & Turner (2015), a correção de borda $e_{ii'}$ no numerador, pois apenas eventos dentro de uma certa distância da borda são considerados. Por outro lado, em algumas literaturas, como as de Diggle (2010), Leininger (2014), González & Moraga (2023), a correção de borda é colocada no denominador, pois o ajuste é feito normalizando a estatística de teste, ou seja, $e_{ii'}$ é igual a 1 quando o círculo centrado em s_i ou z_k , passando pelo evento s_i , está completamente contido na área de estudo. Se parte do círculo estiver fora da área de estudo, então $e_{ii'}$ é calculada como a proporção do círculo contida dentro da área de estudo. Assim, a escolha da correção de borda dependerá da pesquisa e da preferência do pesquisador.

Apesar das funções homogêneas e não homogêneas apresentarem expressões matemáticas diferentes, a sua interpretação é igual (Quadro 2.1).

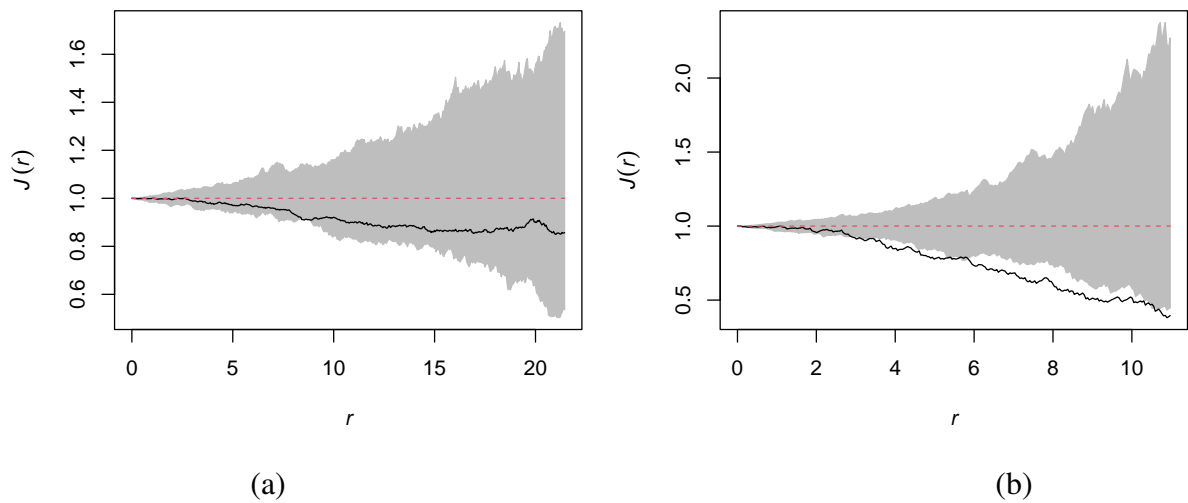
Quadro 2.1 – Resumo da interpretação gráfica das funções que capturam o padrão de distribuição espacial.

↓ Padrão	$F(r)$	$G(r)$	$J(r)$	$K(r)$	$L(r)$	$g(r)$
AEC	$F(r) \equiv F_{pois}(r)$	$G(r) \equiv G_{pois}(r)$	$J(r) \equiv 1$	$K(r) \equiv K_{pois}(r)$	$L(r) \equiv L_{pois}(r)$	$g(r) \equiv 1$
Agrupado	$F(r) < F_{pois}(r)$	$G(r) > G_{pois}(r)$	$J(r) < 1$	$K(r) > K_{pois}(r)$	$L(r) > L_{pois}(r)$	$g(r) > 1$
Regular	$F(r) > F_{pois}(r)$	$G(r) < G_{pois}(r)$	$J(r) > 1$	$K(r) < K_{pois}(r)$	$L(r) < L_{pois}(r)$	$g(r) < 1$
Padrão ↑	$F_{inhom}(r)$	$G_{inhom}(r)$	$J_{inhom}(r)$	$K_{inhom}(r)$	$L_{inhom}(r)$	$g_{inhom}(r)$

Fonte: Do Autor (2024)

Em um processo pontual, a presença de tendência pode levar as funções apresentadas na seção 2.6.1 a identificar dependência espacial entre eventos (padrão regular ou agrupado), mesmo que essa dependência não seja devido à afinidade entre os eventos ou à influência de um evento s_i sobre a presença de outro evento s_j . A dependência espacial identificada pode resultar da competição por recursos (como nutrientes no solo), e na ausência desses recursos, os eventos não apresentariam dependência espacial. Uma forma eficiente de investigar o padrão de distribuição espacial é a inclusão de envelopes de simulação nas funções $F(r)$, $G(r)$, $J(r)$, $K(r)$, $L(r)$ e $g(r)$ ou suas extensões não homogêneas. Os envelopes representam um intervalo mínimo e máximo dentro do qual é aceitável assumir a existência ou não de dependência espacial. Se as funções $F(r)$, $G(r)$, $J(r)$, $K(r)$, $L(r)$ e $g(r)$ ou suas extensões não homogêneas estiverem totalmente dentro do envelope, mas acima ou abaixo das funções $F_{pois}(r)$, $G_{pois}(r)$, $J_{pois}(r)$, $K_{pois}(r)$, $L_{pois}(r)$ e $g_{pois}(r)$, isso indicará a presença de tendência (Figura 2.12 (a)). No entanto, se pelo menos uma parte dessas funções estiver fora do envelope (Figura 2.12 (b)), mantêm-se todas as regras de interpretação apresentadas no Quadro 2.1.

Figura 2.12 – Estrutura de dependência espacial: (a) sem dependência espacial (Presença de tendência), (b) existe dependência espacial (padrão agrupado)



Fonte: Do Autor (2024)

Como pode ser observado na Figura 2.12 (a), a função $\hat{J}(r)$ está abaixo de $J_{pois}(r)$ e, se não houvesse envelope (área cinza), dir-se-ia que existe dependência espacial e o padrão seria agrupado, conforme descrito na seção 2.6.1.3 e no Quadro 2.1. No entanto, apesar de $\hat{J}(r)$ estar abaixo de $J_{pois}(r)$, como está totalmente dentro do envelope, isso significa que os eventos não apresentam dependência espacial (não possuem afinidade natural), mas sim tendência, ou seja, o padrão observado sem considerar o envelope é resultado da tendência, provavelmente porque os eventos (por exemplo, espécies florestais) estão competindo por um recurso que está restrito a uma sub-região.

Na Figura 2.12 (b), se não houvesse envelope (área cinza), haveria dependência espacial e o padrão seria agrupado. Com a incorporação do envelope, a dependência espacial permanece e o padrão continua agrupado, indicando que os eventos possuem uma afinidade natural que não é influenciada por fatores externos (como recursos).

2.7 Modelos de processos pontuais

Identificar o padrão de distribuição (estrutura de dependência) ou tendência espacial dos eventos em um processo pontual é importante, mas por si só não fornece uma análise completa dos possíveis fatores (variáveis) que influenciam essa estrutura observada. É necessário modelar essa distribuição dos eventos, tomando (ou não) como covariáveis os possíveis fatores que podem ajudar a explicar cada padrão observado. Assim, em processos pontuais, existem três classes de modelos mais utilizadas, que são: modelos Poisson, modelos Cox e modelos Gibbs.

2.7.1 Modelos Poisson

Os modelos Poisson são utilizados quando não há dependência espacial entre os eventos, ou seja, o padrão não é regular nem agrupado, como ocorre na presença de tendência ou padrão aleatório. Os modelos Poisson podem ser divididos em duas categorias principais, que são: modelo Poisson homogêneo e modelo Poisson não homogêneo (Isham, 2010; Scalon, 2024).

2.7.1.1 Modelo Poisson homogêneo

O modelo Poisson homogêneo é utilizado quando não existem locais com mais eventos em relação aos outros na região de estudo (intensidade constante), ou seja, é um modelo utilizado quando o padrão é aleatório, sinônimo da não existência da tendência.

No modelo Poisson homogêneo, os eventos são independentes e identicamente distribuídos, com função intensidade denotada por $\lambda(s) = \beta$ e função densidade de probabilidade por

$$f(s) = \beta^{n(s)} \exp\{(1 - \beta) |B|\}, \quad (2.6)$$

onde, β é uma constante que representa a intensidade (λ), $n(s)$ é o número de eventos existentes, B é a região de estudo e $|B|$ é a respectiva área (Baddeley; Rubak; Turner, 2015; Scalon, 2024).

2.7.1.2 Modelo Poisson não homogêneo

O modelo Poisson não homogêneo é uma extensão do modelo Poisson homogêneo e é utilizado quando existe tendência, ou seja, quando existem locais com uma maior concentração de eventos em relação a outros na região de estudo (Baddeley; Rubak; Turner, 2015; Illian, 2019; Scalon, 2024). Neste modelo, a intensidade não é constante, mas é uma função determinística (fixa) e pode ser modelada utilizando diversas covariáveis que podem explicar o padrão observado, ou seja,

$$\lambda(s) = \exp \left[\alpha + \theta^\top Z(s) \right] = \exp \left[\alpha + \theta_1 Z_1(s) + \theta_2 Z_2(s) + \dots + \theta_n Z_n(s) \right],$$

onde α é intercepto, $Z(s)$ é um vetor de covariáveis e θ^\top são parâmetros (coeficientes) $\theta_1, \theta_2, \dots, \theta_n$, associados às covariáveis. Exemplos de covariáveis incluem características do solo, altitude, precipitação, temperatura, diâmetro à altura do peito de árvores, entre outras.

Na ausência de covariáveis, é possível utilizar as coordenadas geográficas (latitude e longitude) ou transformá-las em coordenadas cartesianas (x, y), conforme descrito por (Baddeley;

Rubak; Turner, 2015; Scalon, 2024). Conforme descrito por Scalon (2024), o uso das coordenadas cartesianas faz com que a função intensidade assuma várias formas, que são, Linear: $\lambda_\theta(x, y) = \exp(\theta_0 + \theta_1 x + \theta_2 y)$, onde, $\lambda_\theta(x, y)$ corresponde a função intensidade, variando em função das coordenadas (x, y) (possíveis covariáveis não conhecidas) e $\theta_{i=\{0,1,2\}}$ parâmetros a serem estimados. Quadrática: $\lambda_\theta(x, y) = \exp(\theta_0 + \theta_1 x + \theta_2 y + \theta_3 xy + \theta_4 x^2 + \theta_5 y^2)$. Cúbica: $\lambda_\theta(x, y) = \exp(\theta_0 + \theta_1 x + \theta_2 y + \theta_3 xy + \theta_4 x^2 + \theta_5 y^2 + \theta_6 x^2 y + \theta_7 xy^2 + \theta_8 x^3 + \theta_9 y^3)$, ou n-ésima forma de maneira geral dada por

$$\lambda_\theta(x, y) = \exp\left(\sum_{i=0}^n \sum_{i'=0}^n \theta_{ii'} x^i y^{i'}\right). \quad (2.7)$$

Esta classe de modelo Poisson apresenta a seguinte função densidade de probabilidade,

$$f(s) = \beta(s_1) \dots \beta(s_n) \exp\left[\int_B (1 - \beta(s)) dz\right] = \prod_{i=1}^n \beta(s_i) \exp\left[\int_B (1 - \beta(s)) dz\right], \quad (2.8)$$

onde $\beta = \lambda$, outrora designada intensidade dos eventos s_i e B é a região de estudo.

2.7.2 Modelos Cox

Conforme descrito na seção 2.7.1.2, a disposição dos eventos na área de estudo pode ser influenciada por diversos fatores (covariáveis), como os fatores ambientais. A escassez desses fatores em certas partes da região de estudo pode resultar em uma menor propensão à ocorrência dos eventos, criando uma tendência na distribuição espacial. Como essa tendência é atribuída a fatores ambientais, é possível, por meio do modelo Poisson não homogêneo, modelar a distribuição dos eventos por área (intensidade) levando em consideração essas variáveis ambientais. O modelo resultante é capaz de capturar (explicar) a variabilidade na distribuição dos eventos observados.

No entanto, é comum que as covariáveis disponíveis não consigam explicar completamente o padrão observado, porque, além da influência das covariáveis, os eventos apresentam uma afinidade natural (dependência espacial). Essa afinidade natural faz com que, ao ajustar um modelo Poisson não homogêneo, este, não seja capaz de explicar completamente a variabilidade na distribuição dos eventos observados.

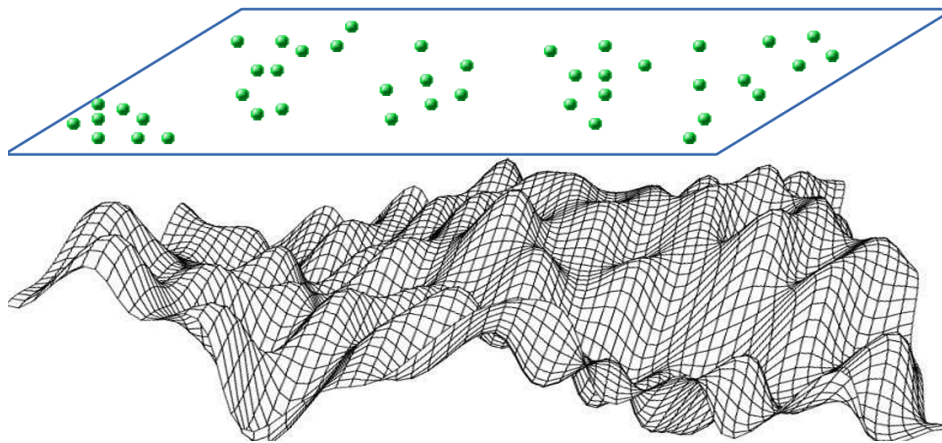
Considerando essa limitação, é possível, em um processo pontual, modelar a função de intensidade como uma função aleatória $\Lambda(s)$, que além de considerar a influência das covariáveis, pressupõe a existência de dependência espacial entre os eventos. Esse tipo de modelagem

resulta em um processo pontual duplamente estocástico (aleatoriedade na intensidade e nos eventos), conhecido como modelo Cox.

Os modelos Cox são uma extensão dos modelos Poisson, nos quais a intensidade originalmente denotada por $\lambda(s)$ é expandida para incorporar um campo aleatório $\Psi(s)$, conforme apresentado na Figura 2.13 (Jalilian; Safari; Sohrabi, 2020). Este campo aleatório torna a função de intensidade aleatória e denotada por $\Lambda(s)$, onde $\{\Lambda = \Lambda(s) : s \in \mathbb{R}^2\}$.

Estes modelos são frequentemente utilizados para modelar padrões pontuais que apresentam dependência espacial do tipo padrão agrupado (Cressie, 1993; Scalon, 2024).

Figura 2.13 – Ilustração do campo aleatório (superfície) $\Psi(s)$, que controla a abundância e distribuição dos eventos em um modelo Cox



Fonte: Jalilian, Safari & Sohrabi (2020)

Na Figura 2.13, o plano representa a região de estudo onde os eventos estão distribuídos. Os pontos no plano indicam a localização dos eventos. Abaixo do plano, há um campo aleatório que influencia o padrão de distribuição dos eventos no plano. As variações neste campo aleatório refletem áreas com maiores ou menores quantidades da variável aleatória. Dentre as subclasses dos modelos Cox, incluem-se o modelo Log-Cox Gaussiano e os modelos de Neyman-Scott, que podem ser Matérn, Thomas, Cauchy e Variância gama.

2.7.2.1 Modelo Log Cox Gaussiano

O papel do campo aleatório apresentado na Figura 2.13 é de explicar a configuração espacial dos eventos, que não é explicada pelas covariáveis (Illian, 2019). Seja, $\Psi(s)$ um campo aleatório Gaussiano de média μ e função covariância $\gamma(r)$, ou seja, uma função cujo valor em qualquer ponto é uma variável aleatória, que explica a variabilidade espacial não explicada pelas covariáveis. Em outras palavras, um campo aleatório $\Psi(s)$ é gaussiano se, para qualquer

coleção finita de eventos s_1, s_2, \dots, s_n , qualquer combinação linear $\beta_1 \Psi_{s_1} + \beta_2 \Psi_{s_2} + \dots + \beta_n \Psi_{s_n}$, com $\beta_i \in \mathbb{R}$, tem em uma distribuição normal unidimensional (Illian et al., 2008).

Assim, sob essa condição, pode-se definir uma função intensidade aleatória $\Lambda(s)$, que além das covariáveis, incorpora um campo aleatório Gaussiano, como,

$$\Lambda(s) = Z(s)\beta^T + \Psi(s). \quad (2.9)$$

No entanto, a expressão 2.9 não pode ser diretamente utilizada como função de intensidade em um modelo de Cox, pois o campo aleatório $\Psi(s)$ sendo Gaussiano pode assumir valores negativos, o que não é admissível em processos pontuais (a intensidade deve ser não-negativa) (Møller; Syversveen; Waagepetersen, 1998; Møller; Waagepetersen, 2007). Para resolver essa questão, uma transformação apropriada é aplicar o logaritmo neperiano (ln), resultando no modelo conhecido como Log-Cox Gaussiano, dado por,

$$\ln[\Lambda(s)] = Z(s)\beta^T + \Psi(s) \Leftrightarrow \Lambda(s) = \exp\{Z(s)\beta^T + \Psi(s)\}.$$

Em síntese, um modelo Cox será Log Cox Gaussiano se o campo aleatório $\Psi(s)$ for Gaussiano (possuir distribuição normal).

2.7.2.2 Modelos de Neyman-Scott: Modelo Matérn, Thomas, Variância-Gama e Cauchy

Desenvolvidos por Neyman e Scott (1958), os modelos Neyman-Scott constituem uma classe de modelos Cox utilizados para descrever padrões de agrupamento quando se presume que os eventos observados são descendentes de um ou mais eventos progenitores (Møller; Waagepetersen, 2003; Baddeley; Rubak; Turner, 2015; Scalon, 2024).

Os modelos de Neyman-Scott são compostos por duas etapas distintas. Na primeira etapa, os eventos s_i (progenitores) são gerados, como por exemplo árvores. Na segunda etapa, cada evento s_i (progenitor) gera eventos s_j descendentes, como por exemplo mudas. O padrão de eventos que consiste exclusivamente nos eventos descendentes (mudas), independentemente de seus progenitores, forma realização do processo de agrupamento Neyman-Scott. Dependendo da distribuição espacial dos eventos descendentes em relação aos seus progenitores, os modelos podem ser do tipo Thomas, Matérn, variância-gama ou Cauchy.

2.7.2.2.1 Modelo Matérn

Um modelo Matérn é caracterizado pela distribuição de probabilidade $h(s_i')$ dos descendentes em relação aos seus progenitores, onde os descendentes, em média por progenitor, estão distribuídos uniformemente em um círculo de raio r centrado no progenitor, ou seja, $h(s_i') = \frac{\mathbb{I}[\|s_i - s_i'\| \leq r]}{\omega^2 r^2}$, onde ω é um parâmetro (escala) de agrupamento dos descendentes em relação aos progenitores s_i (Scalon, 2024).

Para Illian et al. (2008), no modelo Matérn, a função intensidade $\Lambda(s_i')$ pode ser denotada por $\Lambda(s_i') = \lambda(s_i') \sum_{s_j \in S} \mathbb{I}_{b(s_j, r)}(s)$, onde $\lambda(s_i')$ é a intensidade do i -ésimo aglomerado formado pelos descendentes, $\sum_{s_j \in S} \mathbb{I}_{b(s_j, r)}(s)$ é a soma sobre todos os eventos s_j no processo Poisson homogêneo S , $\mathbb{I}_{b(s_j, r)}(s)$ é uma função indicadora que é igual a 1 se o evento s_j estiver no círculo de raio r centrado no progenitor, e 0 caso contrário.

Baddeley, Rubak & Turner (2015) para a mesma função intensidade propuseram a seguinte expressão, $\Lambda(s_i') = \sum_i h(s_i')$, $h(s_i') = \begin{cases} \frac{s_i'}{\pi r^2} & \text{se } \|s_i'\| \leq r \\ 0 & \text{se } \|s_i'\| > r \end{cases}$, onde $\|\cdot\|$ representa a norma ou distância entre os eventos s_i' e r distância de referência.

2.7.2.2.2 Modelo Thomas

Neste modelo, os eventos descendentes s_i' estão distribuídos ao redor dos eventos principais (progenitores) seguindo uma distribuição normal simétrica, ou seja,

$h(s_i') = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|s_i'\|^2}{2\sigma^2}}$, $s_i' \sim N(0, \sigma^2)$, onde σ é um parâmetro mede o padrão de agrupamento dos descendentes em relação aos progenitores e $\|\cdot\|$ representa a norma ou distância entre os eventos s_i' (Baddeley; Rubak; Turner, 2015; Scalon, 2024).

2.7.2.2.3 Modelo Cauchy

Neste modelo, a densidade de probabilidade $h(s_i')$ dos descendentes é uma distribuição bivariada de Cauchy, dada por $h(s_i') = \frac{1}{2\pi\omega^2} \left[1 + \frac{\|s_i'\|^2}{\omega^2} \right]^{-\frac{3}{2}}$, onde ω é um parâmetro que mede o agrupamento (Jalilian; Guan; Waagepetersen, 2013; Baddeley; Rubak; Turner, 2015; Illian, 2019).

2.7.2.2.4 Modelo Variância Gama

Segundo Jalilian, Guan & Waagepetersen (2013), Baddeley, Rubak & Turner (2015), no modelo variância-gama, a densidade de probabilidade dos descendentes é variância gama, dada por $h(s_i') = \frac{1}{2^{v+1}\pi\eta^2\Gamma(v+1)} \frac{\|s_i'\|^v}{\eta^v} Y_v \left(\frac{\|s_i'\|}{\eta} \right)$, onde η é o parâmetro de escala e v é um parâmetro adicional que controla a forma da densidade e deve satisfazer $v > \frac{1}{2}$. Γ é função gama e Y_v é função Bessel modificada de segunda espécie e ordem v que pode ser encontrada no livro de Bell (2004).

Para o modelo variância gama, Illian et al. (2008), propõe uma possível representação da função intensidade $\Lambda(s_i')$, dada por $\Lambda(s_i') = \sum_{s_j' \in S} \omega_i k_z(z - s_j')$, onde S representa o processo Poisson que gera uma coleção de eventos s_j' no espaço d , ω_i representa uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid), com distribuição gama. k_z representa a função de densidade de probabilidade (Kernel) que modela a influência ou interação dos eventos.

2.7.3 Modelos Gibbs (Markov)

Os modelos apresentados nas seções 2.7.1 e 2.7.2 são úteis para eventos que não apresentam dependência espacial e para eventos que apresentam dependência espacial do tipo padrão agrupado, respectivamente. Para eventos que apresentam dependência espacial do tipo regular, é necessária uma classe diferente de modelos denominada Gibbs.

Os modelos Gibbs podem ser subdivididos em modelos de interação par-a-par e modelos de interação por área.

2.7.3.1 Modelos de interação par a par

Os modelos de interação par-a-par são representados pela densidade de probabilidade dada por

$$f(s) = \alpha \left[\prod_{i=1}^{n(s)} b(s_i) \right] \left[\prod_{i < i'} c(s_i, s_{i'}) \right] \quad \text{ou} \quad g(s) \propto \prod_{s_i \in S} \phi(s_i) \prod_{\{s_i, \eta\} \subseteq S} \phi(\{s_i, \eta\}), \quad (2.10)$$

onde \propto indica proporcionalidade e α constante de normalização (Illian et al., 2008; Diggle, 2013; Baddeley; Rubak; Turner, 2015; Scalón et al., 2022; Scalón, 2024). As funções $b(s_i)$ e $\phi(s_i)$ desempenham o mesmo papel, representando estatísticas de primeira ordem (λ) enquanto

$c(s_i, s_{i'})$ e $\phi(s_i, \eta)$, que também poderiam ser expressos como $\phi(\|s_i - s_{i'}\|)$, representam funções de interação par-a-par (Moller; Waagepetersen, 2003; Baddeley; Rubak; Turner, 2015; Scalon, 2024).

Conforme Illian et al. (2008), a função de interação $c(s_i, s_{i'})$, também conhecida como potencial par, um termo originário da física, mede a “energia potencial” gerada pela interação entre pares de eventos $(s_i, s_{i'})$ em função de sua distância $\|s_i - s_{i'}\|$.

Segundo Scalon (2024), os modelos Poisson descritos nas seções 2.7.1 e 2.7.2, apesar de não apresentarem dependência espacial (interação), fazem parte dos modelos Gibbs, pois, se $b(s_i) \equiv \lambda$ e $c(s_i, s_{i'}) = 1$ na equação 2.10, obtém-se a função densidade de probabilidade do modelo Poisson homogêneo, apresentada na equação 2.6, onde $b(s_i) = \beta$ e $\alpha = \exp\{(1 - \beta)|B|\}$. De forma análoga, se $b(s_i) \equiv \lambda(s_i)$ e $c(s_i, s_{i'}) = 1$ na equação 2.10, obtém-se a função densidade de probabilidade do modelo Poisson não homogêneo, apresentada na equação 2.8.

Na prática, os modelos da classe Gibbs são frequentemente apresentados com base na intensidade condicional $\lambda(s_i|s_{i'})$, que pode ser interpretada como o número esperado de eventos s_i por unidade de área (infinitesimalmente pequena) em uma localização, dada a existência do evento $s_{i'}$.

No entanto, Scalon (2024) mostra que as funções densidade de probabilidade $f(s)$ e intensidade condicional $\lambda(s_i|s_{i'})$ são relacionadas,

$$\lambda(s_i|s_{i'}) = \frac{f(s \cup s_{i'})}{f(s_{i'})} = b(s_i) \left[\prod_{i \neq i'}^n c(s_i, s_{i'}) \right],$$

e esta intensidade apresenta relação com a intensidade dos modelos Poisson, se $\prod_i^n c(s_i, s_{i'}) = 1$ e $b(s_i) = \beta$, obtém-se a intensidade do modelo Poisson homogêneo.

Embora modelos Gibbs sejam adequados para padrões regulares, podem identificar padrões de atração (agrupamento) com baixa eficácia e sem aplicação prática significativa (Baddeley; Rubak; Turner, 2015).

Em processos pontuais, ao utilizar modelos Gibbs de interação par-a-par, diferentes modelos Gibbs podem ser obtidos dependendo da forma que a função de interação $c(s_i, s_{i'})$ assumir e alguns desses modelos incluem modelos Hard Core (núcleo duro), modelo Strauss, Strauss Hard Core, etc.

2.7.3.1.1 Modelo Hard core

Neste tipo de modelo da classe Gibbs considera-se que não existem eventos que estejam mais próximos do que a distância mínima h e o modelo só é útil na situação em que os eventos são centros de partículas não elásticas esféricas ou circulares do mesmo tamanho, e h é o diâmetro dessas partículas, que deve ser o mesmo para todos os eventos (Illian et al., 2008; Baddeley; Rubak; Turner, 2015; Scalon, 2024). Este modelo, é obtido considerando

$$b(s) \equiv \beta \text{ e } c(s_i, s_{i'}) = \begin{cases} 1 & \text{se } \|s_i - s_{i'}\| > h \\ 0 & \text{se } \|s_i - s_{i'}\| \leq h \end{cases}, \text{ na equação 2.8, o que resulta na função densidade de probabilidade } f(x) = \begin{cases} \alpha \beta^{n(x)} & \text{se } \|s_i - s_{i'}\| > h \quad \forall i \neq i' \\ 0 & \text{se caso contrário} \end{cases} \text{ e intensidade condicional } \\ \lambda(s_i | s_{i'} \setminus s_i) = \begin{cases} \beta & \text{se } \|s_i - s_{i'}\| > h \quad \forall i \neq i' \\ 0 & \text{se caso contrário} \end{cases}.$$

2.7.3.1.2 Modelo de nascimento e morte

É um modelo mais adequado para a sucessão florestal que envolve a morte de árvores existentes em momentos aleatórios e a germinação de novas mudas em lugares e tempos aleatórios (Cressie, 1993; Baddeley; Rubak; Turner, 2015).

Supondo que em cada intervalo de tempo Δt , cada árvore existente tenha probabilidade $m\Delta t$ de morrer, independentemente das outras árvores, onde m é taxa de mortalidade por árvore e por unidade de tempo. Durante o mesmo intervalo, em uma região Δa , nova árvore germina com probabilidade $g\Delta a\Delta t$, se estiver a mais de d unidades de distância da árvore mais próxima e considerando g a taxa de germinação por unidade de tempo. Independentemente do estado inicial da floresta, ao longo de tempo, este processo de nascimento e morte espacial alcançará um equilíbrio no qual qualquer evento da floresta é uma realização do processo de Hard Core com parâmetro $\beta = \frac{g}{m}$ e diâmetro de Hard Core h .

$$\text{Este modelo, é obtido considerando } b(s) \equiv \beta = \frac{g}{m} \text{ e a função densidade de probabilidade } f(x) = \begin{cases} \alpha \left(\frac{g}{m}\right)^{n(x)} & \text{se } \|s_i - s_{i'}\| > h \\ 0 & \text{se caso contrário} \end{cases} \text{ e } \lambda(s_i | s_{i'}) = \begin{cases} \left(\frac{g}{m}\right) & \text{se } \|s_i - s_{i'}\| > h \\ 0 & \text{se caso contrário} \end{cases}.$$

2.7.3.1.3 Modelo Strauss

Segundo Baddeley, Rubak & Turner (2015), Scalon (2024), o modelo Hard Core é apropriado quando é fisicamente impossível que dois eventos estejam a uma distância menor que d .

Se não é impossível que dois ou mais eventos estejam próximos, mas é improvável, um modelo apropriado é o modelo Strauss, obtido considerando $c(s_i, s_{i'}) = \begin{cases} 1 & \text{se } \|s_i - s_{i'}\| > d \\ \gamma & \text{se } \|s_i - s_{i'}\| \leq d \end{cases}$ e $f(s) = \alpha \beta^{n(s)} \gamma^{n(s,d)}$ e $\lambda(s_i | s_{i'}) = \beta \gamma^{n(s_i, d, s_{i'})}$, onde $n(s, d)$ é o número de pares não ordenados de eventos em s que estão mais próximos do que a distância de interação d ; $n(s_i, d, s_{i'}) = n(s, d) - n(s_{i'}, d)$ é o número de eventos vizinhos de s_i excluindo de $s_{i'}$ e γ um parâmetro de interação.

Se $0 < \gamma < 1$ o modelo caracteriza inibição; se $\gamma = 0$ e não existirem eventos próximos do que a distância d , o modelo passa a ser de Hard Core; se $\gamma = 1$, o modelo é Poisson homogêneo e o modelo não está definido se $\gamma > 1$ pois, a função densidade $f(s)$ não é integrável.

2.7.3.1.4 Modelo Strauss Hard Core

Neste modelo, é impossível que dois ou mais eventos estejam a uma distância menor que a distância de Hard Core (h), mas eles podem apresentar interação a uma distância $d > h$, onde d é a distância de interação de Strauss. A função de interação é expressa por:

$$c(s_i, s_{i'}) = \begin{cases} 0 & \text{se } \|s_i - s_{i'}\| \leq h \\ \gamma & \text{se } h < \|s_i - s_{i'}\| \leq d \\ 1 & \text{se } \|s_i - s_{i'}\| > d \end{cases} .$$

2.7.3.1.5 Modelo de interação soft-core

Segundo Okabe & Tanemura (2006) observamos frequentemente padrões regulares de eventos no mundo natural e nesse padrão, ocorre um espaçamento entre os i -ésimos eventos, que pode ser devido à competição entre os eventos por territórios, nutrientes e assim por diante. Para representar o alcance e a suavidade das interações, os chamados potenciais Soft-Core são os mais convenientes. Neste modelo, a interação entre os eventos é suave e diminui gradualmente à medida que a distância d entre os eventos aumenta. Em contraste com o modelo de Hard core, onde a interação abruptamente se torna zero após uma certa distância, e no modelo de Strauss, onde a interação é constante em uma certa distância e zero fora dela. A função interação é dado por: $c(s_i, s_{i'}) = \left(\frac{\sigma}{\|s_i - s_{i'}\|} \right)^{\frac{2}{\kappa}}$, $\sigma > 0$ e $0 < \kappa < 1$.

2.7.3.1.6 Modelo de interação Diggle-Gates-Stibbard

O modelo proposto por Diggle, Gates & Stibbard (1987) é um modelo de processo pontual com interação par a par, útil para modelar padrões onde a interação entre os pontos é influenciada pela distância entre eles. Este modelo é especialmente adequado quando há evidências de que a intensidade da interação entre os pontos diminui à medida que a distância entre eles aumenta. A função interação é dado por: $c(s_i, s_{i'}) = \begin{cases} \sin\left(\frac{\pi\|s_i - s_{i'}\|}{2d}\right)^2 & \text{se } \|s_i - s_{i'}\| \leq d \\ 1 & \text{se } \|s_i - s_{i'}\| > d \end{cases}$.

2.7.3.1.7 Modelo de interação Diggle-Gratton

Proposto por Diggle & Gratton (1984), o modelo Diggle-Gratton é um tipo de modelo estatístico utilizado para descrever a distribuição espacial de eventos pontuais. A função de potencial de par nesse modelo é definida em termos da distância entre pares de pontos, e quantifica como a presença de um evento afeta a probabilidade de encontrar outro ponto em sua vizinhança. Os parâmetros do modelo incluem a distância mínima de interação h (também conhecida como “hard core”), a distância máxima de interação d , e um parâmetro κ que controla a força da interação. A função interação é dado por:

$$c(s_i, s_{i'}) = \begin{cases} 0 & \text{se } \|s_i - s_{i'}\| \leq h \\ \left(\frac{\|s_i - s_{i'}\| - h}{d - h}\right)^\kappa & \text{se } h < \|s_i - s_{i'}\| \leq d \\ 1 & \text{se } \|s_i - s_{i'}\| > d \end{cases} .$$

2.7.3.2 Modelos de interação por área

Até agora, os modelos Gibbs (Markov) que foram discutidos são conhecidos como modelos de interação par-a-par. Eles descrevem interações intraespecíficas (eventos do mesmo tipo) no caso univariado e interespecíficos espécies (eventos diferentes) no caso multivariado. No entanto, além desses modelos, existem aqueles que consideram a interação por área, os quais são mais adequados quando se considera que os eventos podem competir por recursos, como nutrientes ou alimentos, em uma determinada região.

Conforme descrito em Diggle (2013), Baddeley, Rubak & Turner (2015), Nightingale et al. (2019), os modelos de interação por área, também conhecidos como modelos de esfera penetrável de equilíbrio líquido-vapor de Widom-Rowlinson, foram propostos por Baddeley & Lieshout (1995), visando identificar eventos que possam exibir tanto inibição quanto agregação.

Segundo Baddeley & Lieshout (1995), a distinção fundamental entre os modelos de interação par-a-par e de área reside na especificação da função de interação para cada tipo de modelo.

Nos modelos de interação par-a-par, a função de interação é definida como uma função da distância euclidiana ($\|\cdot\|$) entre cada par de eventos no padrão em estudo. Em contrapartida, a função de interação em um modelo de interação de área é descrita como a área da união de círculos associados a cada evento em um processo pontual.

A função densidade de probabilidade para este modelo é expressa, por

$$f(x) = \alpha \beta^{n(s)} \gamma^{-A(s,r)}, \text{ onde } A(s,r) = \left| B \cap \bigcup_{i=1}^{n(s)} b(s_i, r) \right|, \quad (2.11)$$

onde α é uma constante, $\beta > 0$ é um parâmetro de intensidade, $\gamma > 0$ é um parâmetro de interação, e $A(s,r)$ denota a área da região obtida ao desenhar um círculo de raio r centrado em cada evento s_i e unir esses círculos.

É importante observar que a função de densidade para o modelo de interação por área é integrável para todos os valores de γ , ao passo que, para modelos de interação par-a-par, como o modelo Strauss, para $\gamma > 1$ a função não é integrável. Porém, para ambos modelos, quando $\gamma = 1$, temos um modelo Poisson. Se $\gamma < 1$, os eventos apresentam inibição ou regularidade, enquanto para $\gamma > 1$, os eventos exibem aglomeração ou atração nos modelos de interação por área.

Além das subclasses dos modelos Gibbs mencionadas anteriormente, existem outros que surgem da combinação desses modelos em uma única classe, denominada modelos híbridos. Além disso, existem modelos conhecidos como modelos de interação tripla, nos quais três eventos interagem simultaneamente. Outra possibilidade é o modelo de saturação de Geyer, uma variação do modelo Strauss, porém com uma restrição $\gamma < 1$. No entanto, esses modelos não estão no escopo desta pesquisa, podendo consultar Baddeley, Rubak & Turner (2015) para mais detalhes.

2.8 Diagnóstico e Validação de Modelos

Nesta seção, apresentamos os principais testes estatísticos utilizados para o diagnóstico e validação dos modelos ajustados aos processos pontuais.

2.8.1 Diagnóstico

O diagnóstico dos modelos pode ser feito utilizando diversos tipos de testes estatísticos, a saber:

- a) teste da razão de verossimilhança - este teste, é útil na presença de covariáveis e avalia se há ou não influência das covariáveis no padrão de distribuição dos eventos em estudo. Suas hipóteses nulas e alternativas são, respectivamente, $H_0 : \varphi = 0$ e $H_a : \varphi \neq 0$, onde φ é o vetor de covariáveis. A estatística do teste é dada por

$$\Gamma = 2 \log \frac{L_1}{L_0} = 2 (\log L_1 - \log L_0),$$

onde L_0 e L_1 são os valores máximos da verossimilhança sob as hipóteses nula e alternativa, respectivamente. Para decidir sobre a aceitação ou rejeição da hipótese nula de não influência das covariáveis no padrão observado, utiliza-se um teste χ^2 com n graus de liberdade, onde n é a dimensão do vetor das covariáveis (φ).

- b) score test - também designado teste de multiplicação de Lagrange, requer e apenas um modelo considerado de referência, podendo ser Poisson homogêneo $k(r) = \pi r^2$, representando aleatoriedade espacial completa, com parâmetros θ associados a cada coordenada x, y , conforme descrito na seção 2.7.1.2. O teste de hipóteses para os parâmetros é: $H_0 : \theta = \theta_0$ e $H_a : \theta \neq \theta_0$, cuja estatística de teste é:

$$S = U(\theta_0) \mathbf{I}_\theta^{-1} U(\theta_0), \text{ se } \theta = \psi \text{ ou } S = U(\theta)_\varphi^\top (\mathbf{I}_\theta^{-1})_{\varphi\varphi} U(\theta)_\varphi, \text{ se } \theta = (\phi, \varphi),$$

onde $U(\theta_0)$ é vetor escore e no caso uniparamétrico é designada função escore, \mathbf{I}_θ é informação de Fisher¹ e os subscritos φ e $\varphi\varphi$, são componentes do vetor score e matriz inversa de informação de Fisher respectivamente.

¹ A informação de Fisher é uma medida estatística que quantifica a quantidade de informação que uma variável aleatória contém sobre um parâmetro desconhecido de uma distribuição, sendo definida como a variância do escore, ou seja, a derivada parcial do logaritmo natural da função de verossimilhança em relação ao parâmetro e indica o quão bem podemos estimar um parâmetro com base nos dados observados (Casella; Berger, 2001).

- c) teste χ^2 - o teste χ^2 pode ser utilizado tanto para testar a hipótese nula de aleatoriedade espacial completa quanto para avaliar a qualidade de ajuste de um modelo. No contexto de ajuste de modelos, o teste avalia a hipótese nula de que os eventos seguem um modelo espacial específico ajustado em relação à hipótese de que não o seguem. Aqui, ao utilizar o teste qui-quadrado χ^2 , $m - p$ serão os graus de liberdade, onde p é número de parâmetros do modelo ajustado e m número de *quadrats* e a estatística de teste é,

$$\chi^2 = \sum_j \frac{(n_j - \hat{\mu}_j)^2}{\hat{\mu}_j}, \hat{\mu}_j = \int_{B_j} \hat{\lambda}(s) ds,$$

onde n_j e $\hat{\mu}_j$ representam o número observado e esperado de eventos no local j da área de estudo B_j .

- d) simulações de Monte Carlo - realizadas de forma análoga ao teste contra a hipótese de completa aleatoriedade espacial, entretanto os eventos em estudo são simulados múltiplas vezes usando o modelo ajustado ou um modelo de referência. Se for difícil distinguir visualmente a diferença entre os eventos observados e simulados, o modelo é considerado ajustado ao padrão dos eventos observados (Scalon, 2024).

Para evitar equívocos entre tendência e dependência espacial, é importante incorporar envelopes de simulação. Segundo Baddeley, Rubak & Turner (2015), para a incorporação dos envelopes de simulação, primeiro calcula-se a estimativa S para os dados observados \hat{S}_{obs} , onde S representa qualquer modelo tomado como referência. Em seguida, usando a função S como referência, gera-se m padrões de eventos por simulação de Monte Carlo, e obtêm-se as suas estimativas $\{\hat{S}_1(r), \hat{S}_2(r), \dots, \hat{S}_m(r)\}$, na distância r tomada como referência. Considerando L_{inf} e L_{sup} como os limites inferior e superior das estatísticas associadas aos padrões simulados nas r -ésimas distâncias, rejeita-se a hipótese nula de que os eventos observados seguem o padrão do modelo tomado como referência se as estimativas de \hat{S}_{obs} para as mesmas distâncias estiverem totalmente ou parcialmente acima do limite superior, ou abaixo do limite inferior, ou seja,

$$\begin{aligned} L_{inf}(r) &= \max_j S_j(r) = \max\{S_1(r), \dots, S_m(r)\} \\ L_{sup}(r) &= \min_j S_j(r) = \min\{S_1(r), \dots, S_m(r)\}. \end{aligned} \tag{2.12}$$

2.8.2 Validação

Muitos dos métodos de validação de modelos usados na estatística clássica podem ser utilizados na validação de modelos para configurações pontuais, incluindo a análise de resíduos e a avaliação da independência por meio do gráfico dos quantis dos resíduos (Q-Q plot). Estes métodos são aplicáveis a todos os modelos apresentados, exceto aos modelos de Cox, que não permitem a análise dos resíduos.

Para modelos Poisson, os resíduos são dados pela seguinte expressão:

$$R(B) = n(S \cap B) - \int_B \hat{\lambda}(s) ds, \quad (2.13)$$

onde $n(S \cap B)$ representa o número de eventos existentes em toda área de estudo e $\int_B \hat{\lambda}(s) ds$ representa o número esperado de eventos na região B , sendo $\hat{\lambda}(s)$ a respectiva intensidade estimada pelo modelo ajustado ou função intensidade (modelo ajustado). Assim, se $R(B)$ for negativo, o modelo superestima a intensidade. Se $R(B)$ for positivo, o modelo subestima e se for zero, ou aproximadamente, o modelo se ajusta ao padrão dos eventos em estudo. Os resíduos podem ser estimados considerando subáreas i (*quadrats* i) da região de estudo, resultando nos conhecidos resíduos de Pearson, dados por:

$$R_i^P(B) = \frac{n(S \cap B) - \int_B \hat{\lambda}(s) ds}{\sqrt{\int_B \hat{\lambda}(s) ds}},$$

conforme descrito em Baddeley et al. (2005), Baddeley (2007), Baddeley et al. (2008), Baddeley (2010), Baddeley et al. (2013, 2019), Baddeley, Rubak & Turner (2015), podendo ser suavizados ou cumulativos.

Para modelos de Cox, como descrito por Baddeley (2010), uma vez que não é possível testar os resíduos, uma alternativa é recorrer ao método dos momentos e a funções que capturam a estrutura de dependência (F, G, J, K, L, g), juntamente com envelopes de simulação.

Segundo Baddeley et al. (2022), os métodos existentes para ajustar modelos de processos de agrupamento (modelos Cox) a dados de processos pontuais frequentemente enfrentam dificuldades em convergir, convergem para valores implausíveis dos parâmetros ou apresentam instabilidade numérica. Como resultado, vários métodos foram desenvolvidos para lidar com esses problemas, mas as dificuldades persistem. Alguns autores, como Waagepetersen & Guan (2009), Baddeley, Rubak & Turner (2015), sugerem que o problema está relacionado à fraca

estrutura de agrupamento dos dados analisados. No entanto, este problema pode ocorrer tanto em situações de forte agrupamento quanto de agrupamento fraco (Baddeley et al., 2022).

A validação dos modelos Gibbs segue a mesma lógica dos Poisson, porém substituindo a intensidade pela intensidade condicional e mantendo todas regras, ou seja,

$$R(B) = n(S \cap B) - \int_B \hat{\lambda}(s_i | s_j) ds.$$

Fora os resíduos, envelopes de simulação e funções que ajudam a identificar a estrutura de dependência espacial, pode-se utilizar o Critério de Informação de Akaike, dado por:

$$AIC = -2\log L_{max} + 2p,$$

onde $L_{max} = L(\hat{\theta})$ é função de máxima verossimilhança para o modelo em questão, e p é o número de parâmetros para este modelo. Para AIC, o melhor modelo é aquele que tiver valor mais baixo de AIC.

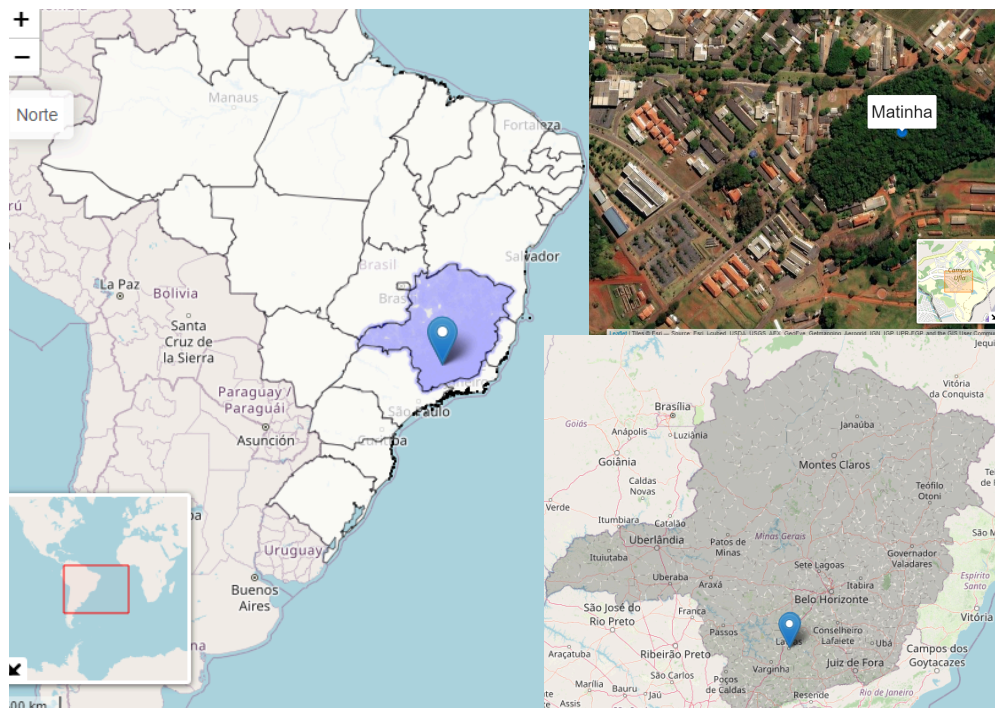
3 MATERIAL E MÉTODOS

Neste capítulo, apresentamos os materiais e métodos utilizados para a realização da pesquisa, incluindo a descrição das espécies florestais em estudo.

3.1 Material

O estudo foi conduzido utilizando dados dos inventários florestais realizados de 1987 a 2017, focando em quatro espécies arbóreas: *Siparuna guianensis*, *Copaifera langsdorffii*, *Xylopia brasiliensis* e *Myrcia splendens*. Estes levantamentos foram conduzidos em uma área de Mata Atlântica com aproximadamente 61.967,05 m² (aproximadamente 6,2 hectares), conhecida localmente como “matinha”. A modelagem, entretanto, foi realizada exclusivamente com os dados de 2017. A inclusão dos dados de 1987 a 2017 no estudo visa entender melhor como as quatro espécies arbóreas têm se comportado ao longo do tempo na área de estudo, o que permite uma análise mais completa da variação e das tendências observadas no ano de 2017. A área de estudo, ilustrada na Figura 3.1, apresenta altitude entre 920 m a 935 m, e está localizada na Universidade Federal de Lavras (UFLA), no município de Lavras, Minas Gerais, Brasil. Suas coordenadas geográficas são 21°13'43.4''S e 44°58'16.1''W. Segundo Oliveira-Filho, Mello & Scolforo (1997), Junior et al. (2017), o clima é subtropical úmido e o solo é latérico (Latossolo).

Figura 3.1 – Mapa ilustrativo da localização da área de estudo



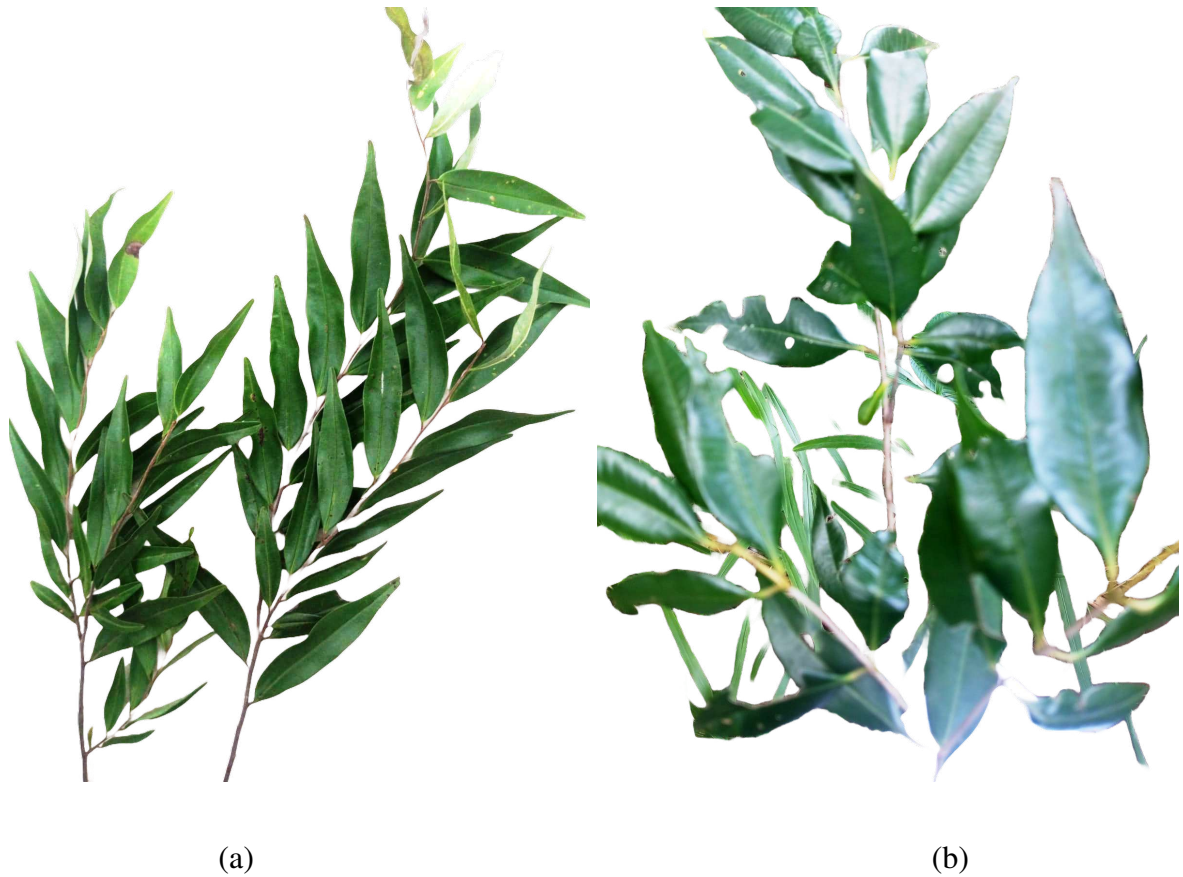
Fonte: Do Autor (2024)

3.1.1 Descrição das florestais em estudo

A seguir é apresentada uma breve descrição das espécies arbóreas *Copaifera langsdorffii*, *Xylopia brasiliensis*, *Siparuna guianensis*, e *Myrcia splendens*.

- a) *xylopia brasiliensis* (pindaíba): apresentada na Figura 3.2 (a), é uma planta da família Annonaceae, pode alcançar alturas entre 10 e 30 metros, com tronco variando de 30 a 60 centímetros de diâmetro. É encontrada em áreas de Mata Atlântica, especialmente em encostas secas, tanto em florestas nativas primárias quanto secundárias, densas ou abertas. Suas sementes são dispersas por animais (dispersão zoocórica). Além disso, sua madeira é utilizada na construção civil. *Xylopia brasiliensis* é uma espécie de sucessão ecológica secundária inicial (pioneira) ou clímax, que ocorre em solos rasos e de rápida drenagem, assim como em solos profundos, de alta fertilidade química e textura arenosa. No Brasil, pode ser encontrada nos estados de Minas Gerais, Rio de Janeiro, São Paulo, Paraná e Santa Catarina (Lorenzi et al., 1992; Carvalho, 2014; Re flora, 2024).
- b) *myrcia splendens* (guamirim): apresentada na Figura 3.2 (b), é conhecida popularmente como guamirim e pertencente à família *Myrtaceae*, apresenta um crescimento rápido a moderado (2 m em 2 anos). Encontrada na Mata Atlântica, Cerrado, Caatinga, Amazônia, Pantanal, e nas vegetações Campo Rochoso de Terras Altas, Floresta Ribeirinha e/ou de Galeria, de Terra Firme, Estacional Evergreen, Semidecidual e Ombrófila, seus frutos são fonte de alimento para humanos e animais, e suas sementes são dispersas por dispersão zoocórica. Além do valor medicinal, sua madeira é utilizada na construção civil e como fonte de lenha. É uma espécie de sucessão ecológica secundária inicial (pioneira) ou clímax. Sua preferência é por solos férteis e profundos, variando de textura arenosa a argilosa, mas demonstra tolerância à presença de camadas finas de gelo na superfície (geada) e solos alagados. No Brasil, pode ser encontrada em todos os estados (Carvalho, 2014; Santos et al., 2024).

Figura 3.2 – Espécies florestais: (a) *xylopia brasiliensis* (pindaíba) e, (b) *myrcia splendens* (guamirim)

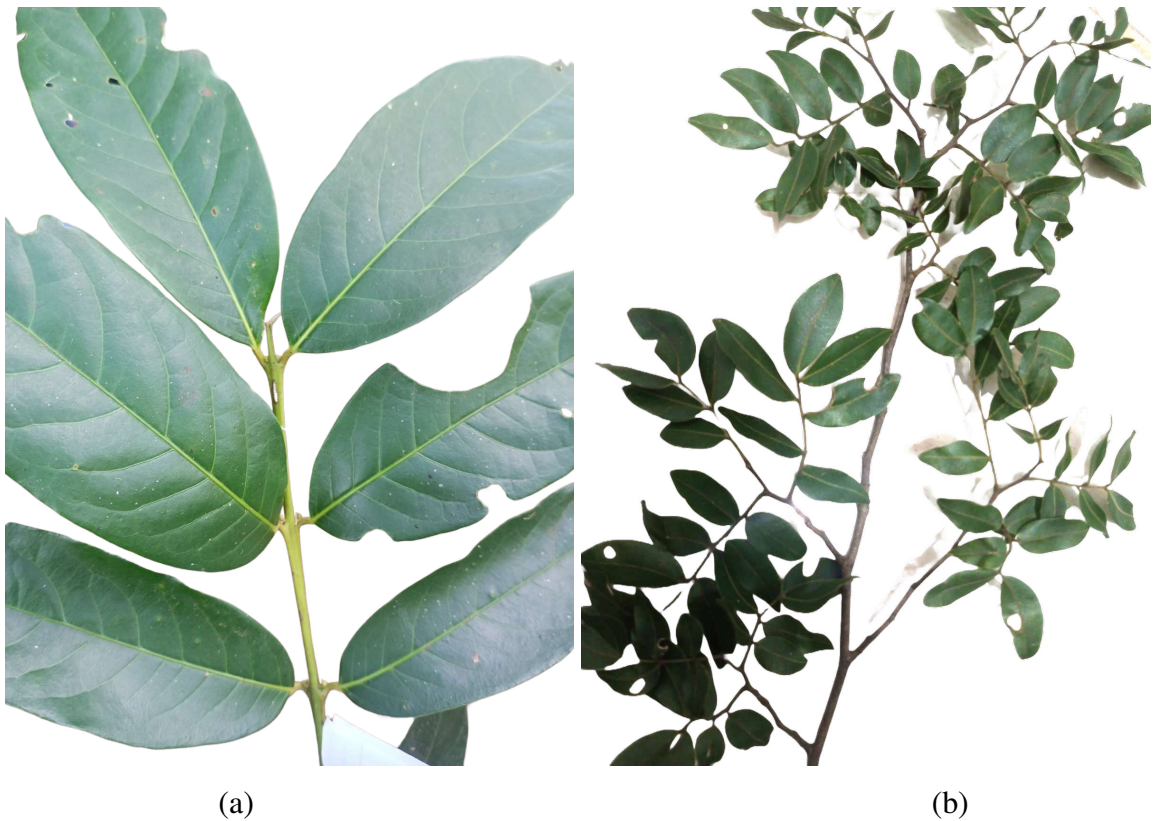


Fonte: Do Autor (2024)

c) *siparuna guianensis* (negamina): apresentada na Figura 3.3 (a), é conhecida como negamina, pertencente à família *Siparunaceae* e apresenta crescimento rápido, com altura entre 5 a 20 m. É encontrada no Cerrado, Mata Atlântica, Amazônia, Caatinga e Pantanal, e seus frutos são fonte de alimento para animais, e as sementes são dispersas por dispersão zoocórica. Além do valor medicinal, sua madeira é fonte de lenha. É de sucessão ecológica secundária inicial (pioneira), com maior ocorrência em florestas Ribeirinha e/ou de Galeria, Terra Firme, Estacional Semidecidual e Ombrófila. Sua preferência é por solos lateríticos ou latossolos (horizonte A/horizonte com forte influência da matéria orgânica ou B/horizonte de máxima expressão dos processos de formação do solo) e arenosos. No Brasil, pode ser encontrada em todos os estados, com exceção do Rio Grande do Sul e Santa Catarina (Renner; Hausner, 2005; Carvalho, 2014; Peixoto; Lirio; Pignal, 2024)

d) *copaifera langsdorffii* (copaíba): apresentada na Figura 3.3 (b), é conhecida popularmente como copaíba, pertence à família Fabaceae e possui crescimento lento. Esta espécie é encontrada na Amazônia, Caatinga, Cerrado e Mata Atlântica. Seus frutos são fonte de alimento para animais, e suas sementes são dispersas por zoocoria. Além de seu valor medicinal e ornamental, a madeira de Copaíba é utilizada como fonte de lenha e na construção civil. Copaíba é considerada uma espécie de sucessão ecológica secundária inicial (pioneira) ou clímax exigente em luz. É tolerante às condições edáficas do solo e terrenos alagados, com preferência por solos bem drenados, variando de textura franca argilosa a argilosa. A espécie ocorre em diversos habitats como Campo Rupestre de Terras Altas, Florestas Ribeirinha e/ou Mata de Galeria, Estacional Semidecidual e Ombrófila. No Brasil, Copaíba está presente em todos os estados, com exceção de Roraima, Pará, Amapá, Alagoas, Sergipe e Santa Catarina (Carvalho, 2014; Lorenzi et al., 2021; Costa, 2024).

Figura 3.3 – Espécies florestais: (a) *siparuna guianensis* (negamina) e, (b) *copaifera langsdorffii* (copaíba)



Fonte: Do Autor (2024)

3.2 Métodos

A análise de dados foi realizada no ambiente de programação R (R Core Team, 2023), utilizando o pacote *spatstat* (Baddeley; Turner, 2005) e *ggplot2* (Wickham, 2016), e todos os *scripts* constam no anexo. A análise descritiva concentrou-se em caracterizar a distribuição intraespecífica, aplicando o método não-paramétrico de Kernel para avaliar tendências na configuração espacial, juntamente com as funções $J_{inhom}(r)$ e $L_{inhom}(r)$ para investigar padrões de distribuição espacial. O nível de significância considerado é de 5%.

Para identificar potenciais fatores associados aos padrões espaciais observados, modelos Poisson foram ajustados para as espécies que não exibiram dependência espacial, enquanto modelos Cox foram ajustados para as espécies que mostraram dependência espacial. As seguintes covariáveis foram utilizadas: pH, potássio (K), fósforo (P), cálcio (Ca), magnésio (Mg), Alumínio (Al), acidez potencial (H+Al), Soma de Bases (SB), Capacidade de Troca de Cátions Efetiva (t), Capacidade de Troca de Cátions Total (T), Saturação de Bases (V), Saturação de Alumínio (m), Matéria Orgânica (MO), proporção de Argila, Silte e Areia. Para incorporar essas variáveis na análise, foram criados mapas que mostram a variação de cada uma delas por meio de suavização espacial, análoga à krigagem ordinária utilizada em geoestatística. Apenas esses mapas foram utilizados na análise.

Nos modelos Poisson, utilizou-se o método de máxima pseudo-verossimilhança, definido por:

$$PL(\theta; S) = \prod_{i=1}^n \lambda_{\theta}(s_i, S) \exp\left(-\int_B \lambda_{\theta}(z, S)\right),$$

onde n é o número de eventos $s_i = (x, y)$, S representa o conjunto de eventos s_i , z representa qualquer ponto da região de estudo B , e θ é o parâmetro que maximiza $PL(\theta; S)$.

Para os modelos Cox, foi utilizado o método de máxima verossimilhança de Palm definida por:

$$\log L_p(\theta) = \sum_i \sum_{j \neq i} Q(s_i, s_j) \log \lambda_p(s_j | s_i; \theta) - \sum_i \int_B Q(s_i, u) \lambda_p(u | x_i; \theta) du,$$

onde $Q(\cdot) = \mathbb{I}[\|\cdot\| \leq r]$.

A avaliação dos modelos foi realizada através da análise de resíduos, gráfico Quantil-Quantil e envelopes de simulação de Monte Carlo (1000 iterações).

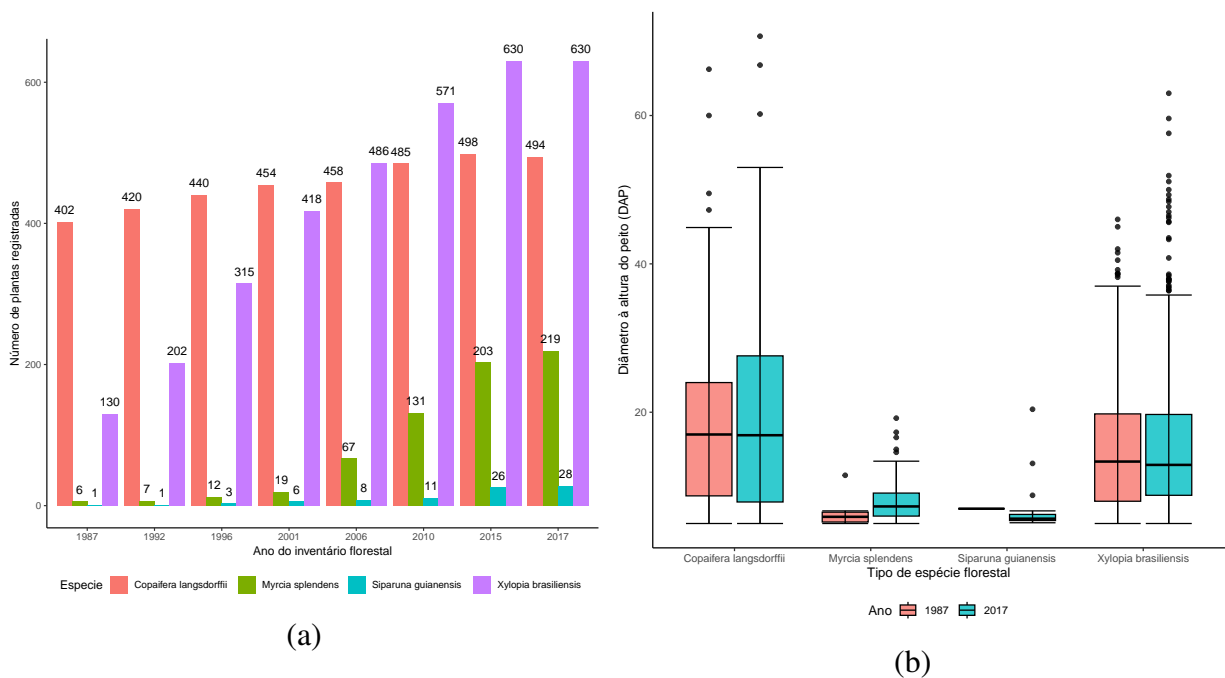
4 RESULTADOS E DISCUSSÕES

Neste capítulo, apresentamos os resultados e discussão dos mesmos, com base em teorias já existentes. De forma sequencial, apresenta-se a dinâmica espaço-temporal das espécies florestais, seguida da identificação dos padrões de distribuição que as espécies apresentam na área de estudo e por fim a modelagem dos processos subjacentes aos padrões observados.

4.1 Dinâmica espaço-temporal das espécies arbóreas em estudo

O conhecimento do passado em ecologia é crucial para compreender os processos ecológicos que moldam os padrões espaciais das espécies arbóreas, investigar as causas do estágio atual e antecipar o futuro (Crawley, 2009). Como mencionado por Schulze, Beck & Müller-Hohenstein (2005), as plantas passam por desenvolvimento ao longo do tempo, exibindo dinâmicas direcionadas ou cíclicas, que só podem ser explicadas historicamente com base no conhecimento da evolução e distribuição.

Figura 4.1 – Dinâmica espaço-temporal das espécies arbóreas de 1987 a 2017.

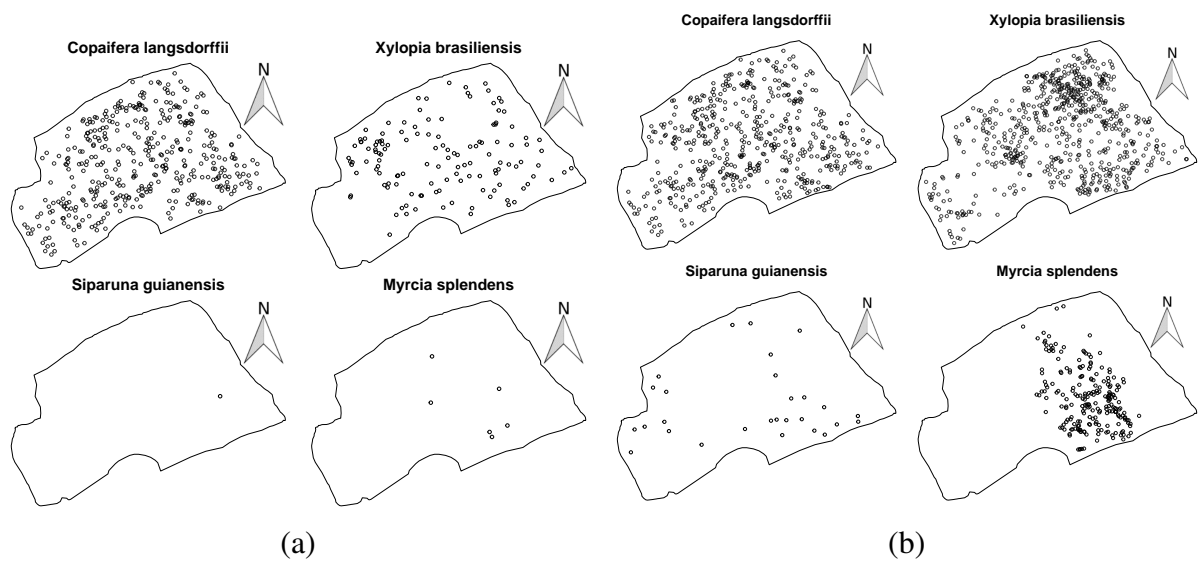


Fonte: Do Autor (2024)

A Figura 4.1 (a) revela que em 1987, quando o registro dessas espécies começou, *Copaifera langsdorfii* era dominante com 402 plantas, seguida por *Xylopia brasiliensis* com 130 plantas, enquanto *Siparuna guianensis* possuía apenas um indivíduo. Com o tempo, houve um aumento significativo no número de novas plantas de *Xylopia brasiliensis*, tornando-se predo-

minante de 2006 a 2017. Essa mudança pode ser atribuída à adaptação de *Xylopia brasiliensis* às condições ambientais locais, bem como às diferenças de crescimento e competição entre *Xylopia brasiliensis* e *Copaifera langsdorffii*. O preocupante é que *Siparuna guianensis* tenha apresentado crescimento populacional e diâmetro à altura do peito abaixo das expectativas, apesar de estar em condições de solo favoráveis para sua espécie. *Siparuna guianensis* prefere latossolos e a área de estudo possui esse tipo de solo. Examinando os outliers na Figura 4.1 (b), observa-se que os outliers de *Copaifera langsdorffii* permaneceram relativamente estáveis, enquanto os de *Xylopia brasiliensis* aumentaram expressivamente. O desenvolvimento mais rápido de *Xylopia brasiliensis* em comparação com *Copaifera langsdorffii* sugere que a espécie *Xylopia brasiliensis* apresenta maior acumulação de biomassa e capacidade competitiva por território. Em relação à mortalidade, de 1987 a 2015, todas as espécies estudadas, exceto *Siparuna guianensis* (que experimentou uma perda em 1996), não registraram perdas em plantas monitoradas. No entanto, entre 2015 e 2017, todas as espécies de árvores experimentaram mortalidade, sendo, 11 mortes para *Copaifera langsdorffii* e *Myrcia splendens*, 12 mortes para *Xylopia brasiliensis* e 4 mortes para *Siparuna guianensis*. Essas perdas podem estar ligadas ao fenômeno El Niño¹, que ocorreu nos anos de 2015 a 2016, conforme descrito por Bennett et al. (2023). A Figura 4.2 mostra a disposição destas espécies na área de estudo nos anos 1987 e 2017.

Figura 4.2 – Distribuição espacial das espécies florestais nos anos 1987 (a) e 2017 (b).



Fonte: Do Autor (2024)

¹ O fenômeno El Niño é um fenômeno climático caracterizado pelo aquecimento das temperaturas da superfície do mar nas partes central e oriental do Oceano Pacífico equatorial. <<https://www.nasa.gov/technology/how-nasa-sees-el-nino-effects-from-space/>>.

Tabela 4.1 – Estatísticas descritivas das espécies arbóreas estudadas, registradas no inventário florestal de 2017.

Tipo de Espécie	Frequência absoluta	Frequência relativa	Intensidade $\lambda(s)$
<i>Xylopia brasiliensis</i>	630	0,45951860	0,010166690
<i>Copaifera langsdorffii</i>	494	0,36032090	0,007971978
<i>Myrcia splendens</i>	219	0,15973740	0,003534136
<i>Siparuna guianensis</i>	28	0,02042305	0,000451853
Total	1371	1	0,02212466

Fonte: Do Autor (2024)

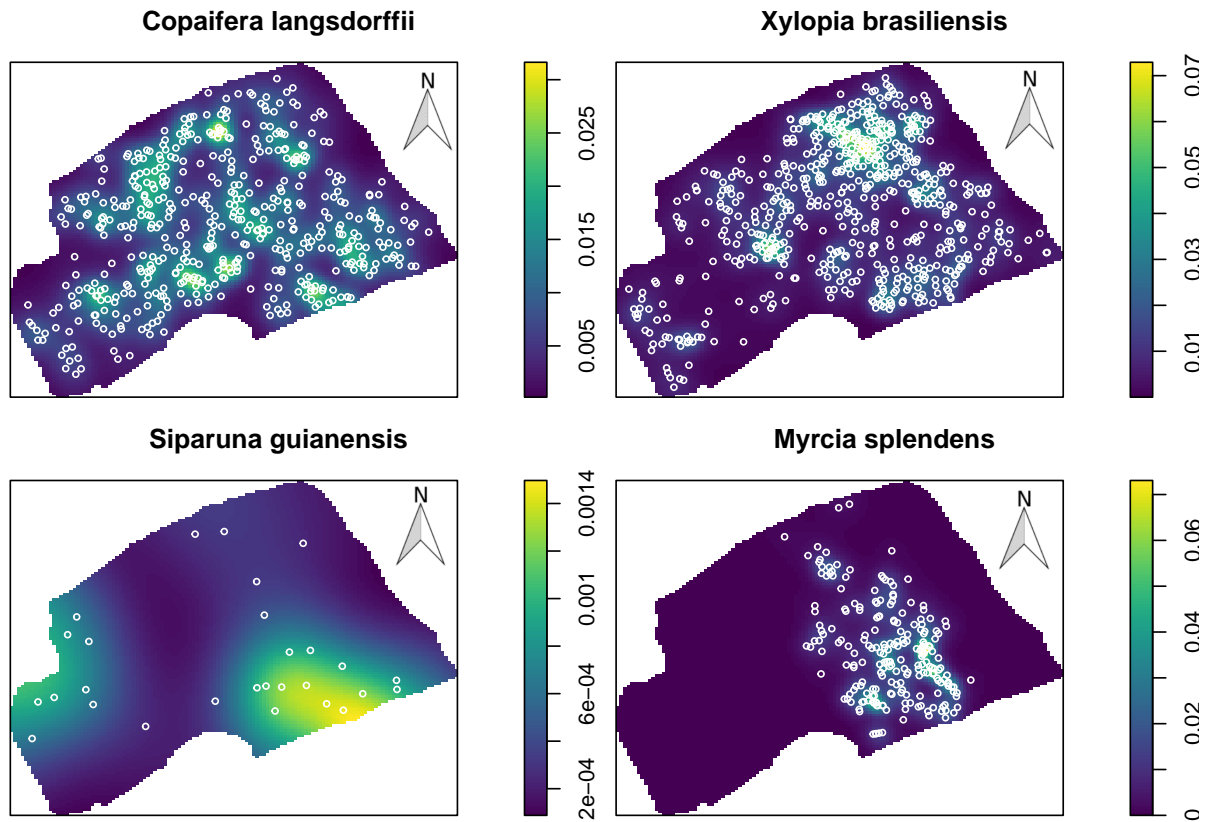
Com base na Tabela 4.1 foram identificadas 1371 árvores com DAP (Diâmetro à Altura do Peito) igual ou superior a 5 cm na área de estudo, correspondendo a uma intensidade de primeira ordem $\lambda(s)$ de 0,02212466 árvores por metro quadrado (1371/61967,05).

Esta intensidade equivale aproximadamente a 222 árvores por ha (0,02212466/0,0001), das quais, 102 são de *Xylopia brasiliensis*, 80 de *Copaifera langsdorffii*, 36 de *Myrcia splendens* e 4 de *Siparuna guianensis*.

Estas intensidades foram calculadas assumindo homogeneidade, ou seja, as plantas de diferentes espécies arbóreas estão distribuídos aleatoriamente (sem dependência espacial) e não apresentam preferência por sub-regiões específicas dentro da área de estudo (sem tendência espacial). No entanto, na prática, assumir homogeneidade implica que as plantas de cada espécie ocorram nas mesmas proporções em toda a área de estudo, o que pode ser uma suposição equivocada, já que diversos fatores ambientais e ecológicos podem resultar em uma distribuição espacial não homogênea das plantas. Assim, uma análise mais refinada requer verificar a existência de variabilidade espacial na distribuição das espécies, que pode ser alcançado estimando a densidade, que corresponde à intensidade de segunda ordem, conforme ilustrado na Figura 4.3.

Com base na Figura 4.3, observa-se que as plantas da espécie *Xylopia brasiliensis* estão concentrados de maneira mais significativa na parte norte da área de estudo, enquanto os de *Myrcia splendens* tendem a ocorrer mais no sudeste. As plantas de *Siparuna guianensis* estão mais concentrados tanto no sudeste quanto no oeste. Por outro lado, as plantas de *Copaifera langsdorffii* estão distribuídos por quase toda a área de estudo, aparentemente em quantidades semelhantes, sugerindo uma distribuição uniforme (homogeneidade). No entanto, há áreas pequenas, ainda que insignificantes, onde essa espécie apresenta um número ligeiramente maior de plantas.

Figura 4.3 – Densidade das espécies *Xylopia brasiliensis*, *Copaifera langsdorffii*, *Myrcia splendens* e *Siparuna guianensis*.

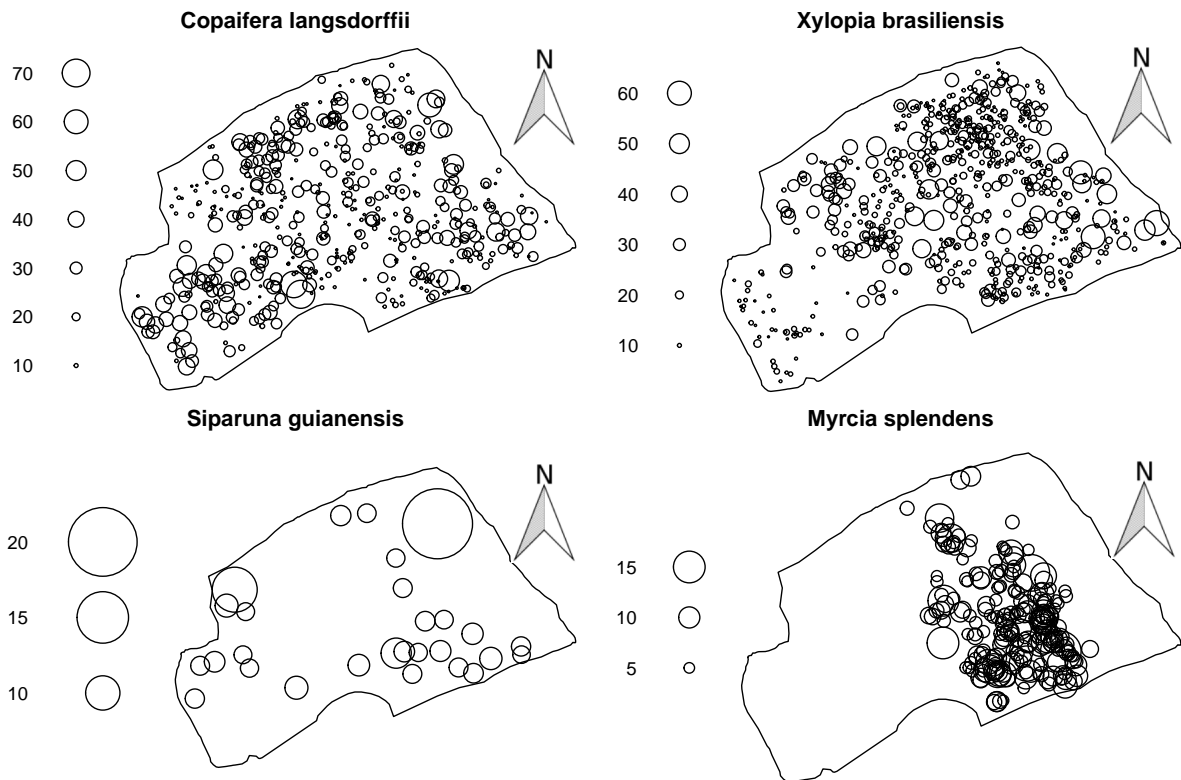


Fonte: Do Autor (2024)

Por se tratar de espécies florestais, uma análise muito importante antes de recorrer às funções que identificam o padrão espacial, é verificar se a distribuição das espécies florestais depende do DAP, como apresentado na Figura 4.4.

Ao analisar a Figura 4.4 é crucial não focar no tamanho dos círculos, mas sim nos números associados a eles para evitar interpretações equivocadas. Com base na Figura 4.4, fica evidente que as plantas das espécies estudadas, que apresentam menor Diâmetro à Altura do Peito (DAP) não estão distribuídos ao redor das plantas com maior DAP (DAP corresponde os números associados a cada círculo). Este fato, pode ser explicado pelo tipo de dispersão de sementes que as espécies apresentam (zoocórica). Em virtude destas observações, foram utilizadas funções não homogêneas descritas na seção 2.6.2, para investigar as estruturas de dependência espacial, conforme ilustrado na Figura 4.5.

Figura 4.4 – Distribuição das espécies *Xylopia brasiliensis* e *Copaifera langsdorffii* por DAP (o) no ano 2017.

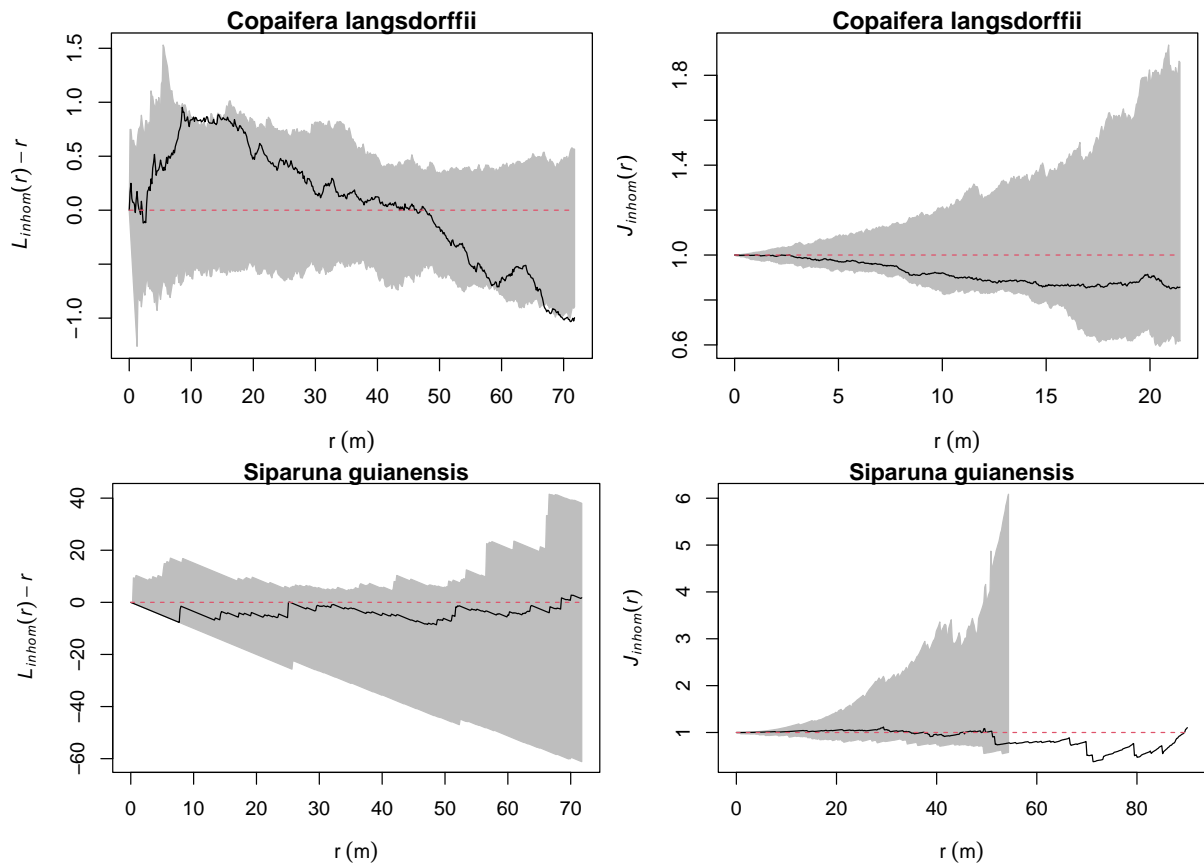


Fonte: Do Autor (2024)

A Figura 4.5 apresenta evidências que as espécies *Copaifera langsdorffii* e *Siparuna guianensis* não apresentam um padrão de distribuição espacial agregado ou regular (sem dependência espacial), porque as funções $L_{inhom}(r)$ e $J_{inhom}(r)$ estão completamente contidas nos envelopes de simulação de Monte Carlo. Este resultado sugere que os padrões observados na Figura 4.3 não são devido a interações espaciais entre eventos, mas sim a uma tendência espacial. Com base nessas observações, a distribuição das plantas dessas espécies pode ser modelada por modelos Poisson.

Na Figura 4.6 é possível ver evidências que para as espécies *Xylopia brasiliensis* e *Myrcia splendens* existe dependência espacial, com um padrão de distribuição espacial agregado ocorrendo aproximadamente de 5 metros até 35 e 40 metros respectivamente, com uma transição para padrão regular a partir de 50 metros (conforme indicado pela função $L_{inhom}(r)$). No entanto, a função $J_{inhom}(r)$ tende a mostrar resultados contraditórios para *Myrcia splendens*, o que é comum porque $J_{inhom}(r)$ é sensível à distância entre eventos. Com base nestas observações, a distribuição das plantas destas duas espécies pode ser modelada por modelos Cox.

Figura 4.5 – Padrão de distribuição espacial das espécies *Copaifera langsdorffii* e *Siparuna guianensis*

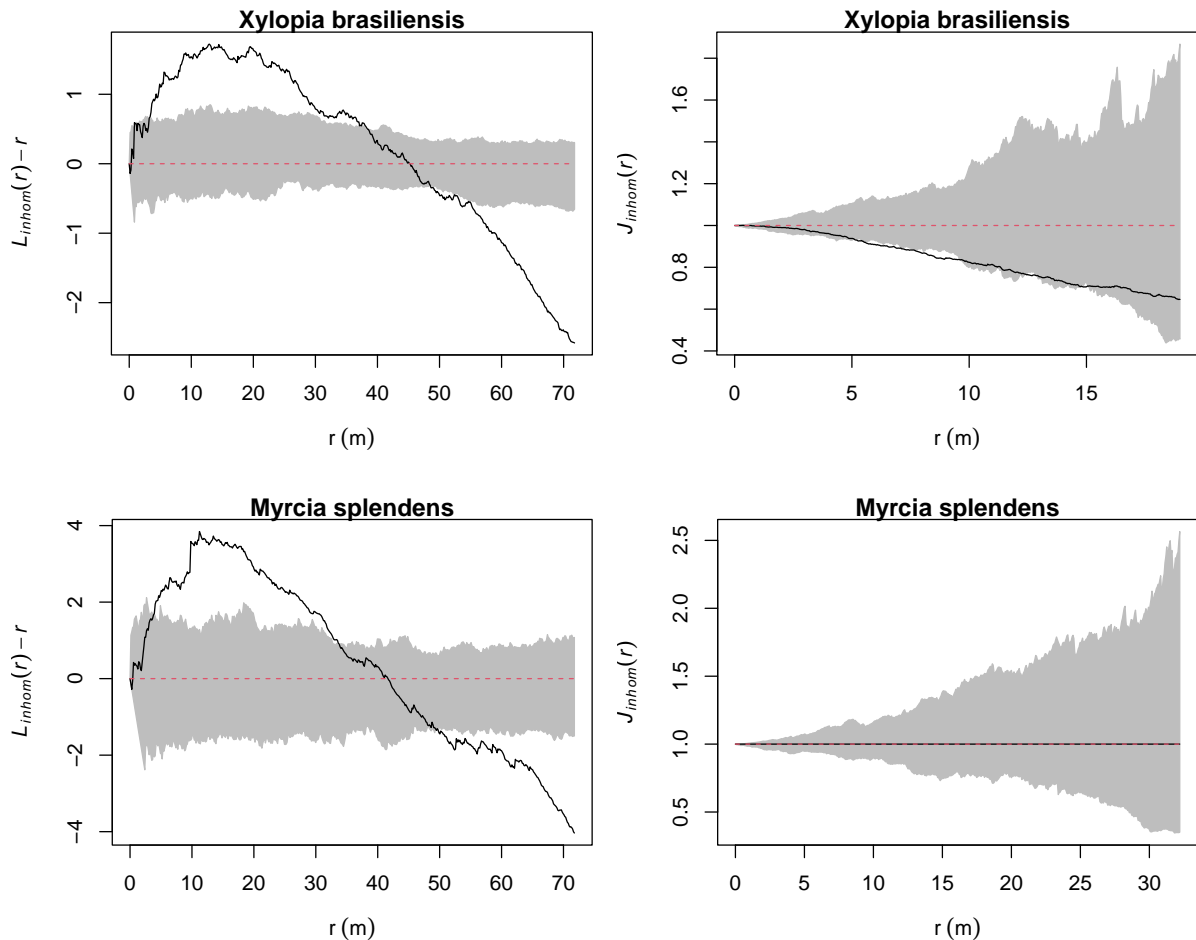


Fonte: Do Autor (2024)

Um fato curioso é que o padrão agregado observado nas espécies *Xylopia brasiliensis* e *Myrcia splendens* ocorre em escalas espaciais menores (distâncias r). De acordo com Schulze, Beck & Müller-Hohenstein (2005), Seidler & Plotkin (2006), Horáčková, Řehouňková & Prach (2016), um padrão de distribuição agregada em escalas menores em espécies florestais cujas sementes são dispersas via zoocoria indica que os animais não viajam longas distâncias antes de excretar ou liberar as sementes.

Os padrões observados na Figura 4.5 e 4.6 também podem ser atribuídos a outros fatores tais como mecanismos de dispersão de sementes, condições ambientais (umidade, características do solo, pH, temperatura, precipitação), eventos extremos, atividades humanas e outros processos espaciais, conforme discutido por Tilman (1996), Seidler & Plotkin (2006), Crawley (2009), Rai (2013), Stein, Gerstner & Kreft (2014), Horáčková, Řehouňková & Prach (2016), Pansonato (2019), Loke & Chisholm (2022).

Figura 4.6 – Padrão de distribuição espacial das espécies *Xylopia brasiliensis* e *Myrcia splendens*



Fonte: Autor (2024)

No entanto, como descrito no capítulo 3, o tipo de solo presente na área de estudo coincide com as preferências de *Siparuna guianensis*, sugerindo que o padrão observado para esta espécie não é necessariamente influenciado pelo tipo de solo, eventos extremos ou ações humanas, pois houve poucos casos registrados de mortalidade de 1987 a 2017, mesmo durante o fenômeno El Niño ocorrido entre 2015 e 2016. Em vez disso, o padrão pode estar relacionado a fatores como o tipo de dispersão de sementes, capacidade de colonização, entre outros.

Um estudo desenvolvido por Davide, Faria & Botelho (1995) com 100 sementes de *Siparuna guianensis* revelou uma taxa de germinação inferior a 30% e um peso médio de sementes de 1,6 g, indicando uma capacidade potencialmente limitada para colonizar novos territórios. Outro estudo realizado por Junior et al. (2017) na mesma área de estudo (matinha da UFLA), constatou que a região sudeste apresenta maior declive e menor umidade, enquanto a região oeste possui maior umidade. Para *Myrcia splendens*, que ocorre exclusivamente no sudeste da área de estudo, a umidade pode ser considerada um dos fatores que influenciam a tendência

espacial observada na Figura 4.3. No entanto, para *Siparuna guianensis*, *Copaifera langsdorffii* e *Xylopia brasiliensis*, a umidade pode não ser o principal fator por trás dos padrões observados na Figura 4.3 porque estas espécies florestais existem tanto em locais com menor umidade (sudeste), bem como em locais com maior umidade.

4.2 Modelagem da distribuição espacial das árvores das espécies florestais

Para ajustar modelos nas configurações pontuais das localizações das árvores das espécies observadas na região de estudo foram consideradas as seguintes covariáveis: pH, potássio (K), fósforo (P), cálcio (Ca), magnésio (Mg), alumínio (Al), acidez potencial (H+Al), soma de bases (SB), capacidade de troca Catiônica efetiva (t), capacidade de troca catiônica total (T), saturação por bases (V), saturação por alumínio (m), matéria orgânica (MO), proporção argila, silte e areia.

Para incorporar essas covariáveis nos modelos, foram criados mapas de suavização para cada variável, os quais ilustram a variação de cada uma delas separadamente na área de estudo. Este procedimento é análogo à interpolação utilizada na Geoestatística, conforme discutido por Nhancololo et al. (2024).

As covariáveis incluídas nos modelos foram selecionadas utilizando o método *stepwise* e o Critério de Informação de Akaike (AIC).

Os resultados do ajuste do modelo de Poisson não-homogêneo para a espécie *Copaifera langsdorffii* são apresentados na Tabela 4.2 a seguir.

Tabela 4.2 – Covariáveis selecionadas pelo método *stepwise* para a espécie *Copaifera langsdorffii*.

	Estimativa	S.E.	CI95.inf	CI95.sup	Teste Z	Valor Z
Intercepto	17,85431308	6,875180507	4,37920690	31,329419265	**	2,596923
P	0,45026083	0,187065923	0,08361836	0,816903306	*	2,406963
Al	54,93589638	14,566316550	26,38644055	83,485352204	***	3,771434
H+Al	-8,71623727	2,564656937	-13,74287250	-3,689602043	***	-3,398598
t	-9,37043438	2,721855778	-14,70517367	-4,035695082	***	-3,442664
V	0,98526511	0,305420451	0,38665203	1,583878193	**	3,225930
MO	-0,60086646	0,161516074	-0,91743215	-0,284300774	***	-3,720165
Argila	-0,02466321	0,008634952	-0,04158741	-0,007739017	**	-2,856207

Níveis de significância: *** para p-valor < 0.001, ** para p-valor < 0.01 e * para p-valor < 0.05
S.E. (Erro Padrão), CI95.inf e CI95.sup (Intervalo de Confiança de 95% Inferior e Superior)

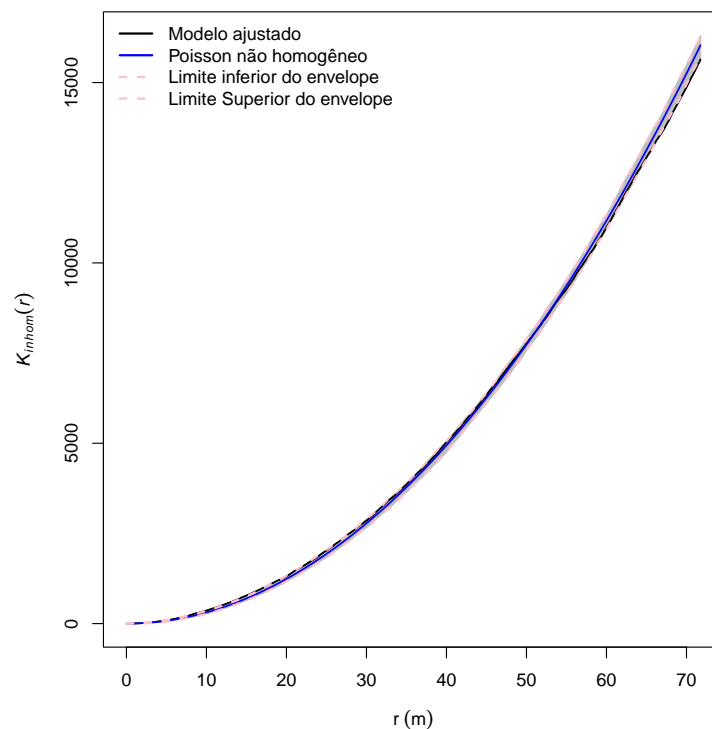
Fonte: Do Autor (2024)

Assim, o modelo de Poisson não homogêneo ajustado para a espécie *Copaifera langsdorffii* é o seguinte: $\lambda(s) = \exp(17,85431308 + 0,45026083 \cdot P + 54,93589638 \cdot Al - 8,71623727 \cdot$

(*acidez potencial*) $-9,37043438 \cdot t + 0,98526511 \cdot V - 0,60086646 \cdot MO - 0,02466321 \cdot argila$),
 $AIC = 5747,411$.

Este modelo indica que a intensidade (número de plantas da espécie *Copaifera langsdorffii* por área) varia expressivamente com base nas covariáveis selecionadas. Quando todas as covariáveis selecionadas são zero, o logaritmo do número (intensidade) de plantas da espécie *Copaifera langsdorffii* por área é 17,85431308 plantas por m^2 . Para cada unidade adicional de Fósforo (P), Alumínio (Al) e Saturação por Bases (V), o logaritmo da função intensidade aumenta em numa taxa de 0,45026083; 54,93589638 e 0,98526511, respectivamente. Para cada unidade adicional de acidez potencial (H+Al), Capacidade de Troca Catiônica Efetiva (t), Matéria Orgânica (MO) e teor de Argila, a intensidade de árvores diminui em 8,71623727; 9,37043438; 0,60086646 e 0,02466321, respectivamente. Uma visualização do modelo ajustado é apresentada na Figura 4.7

Figura 4.7 – Modelo ajustado para a espécie *Copaifera langsdorffii*

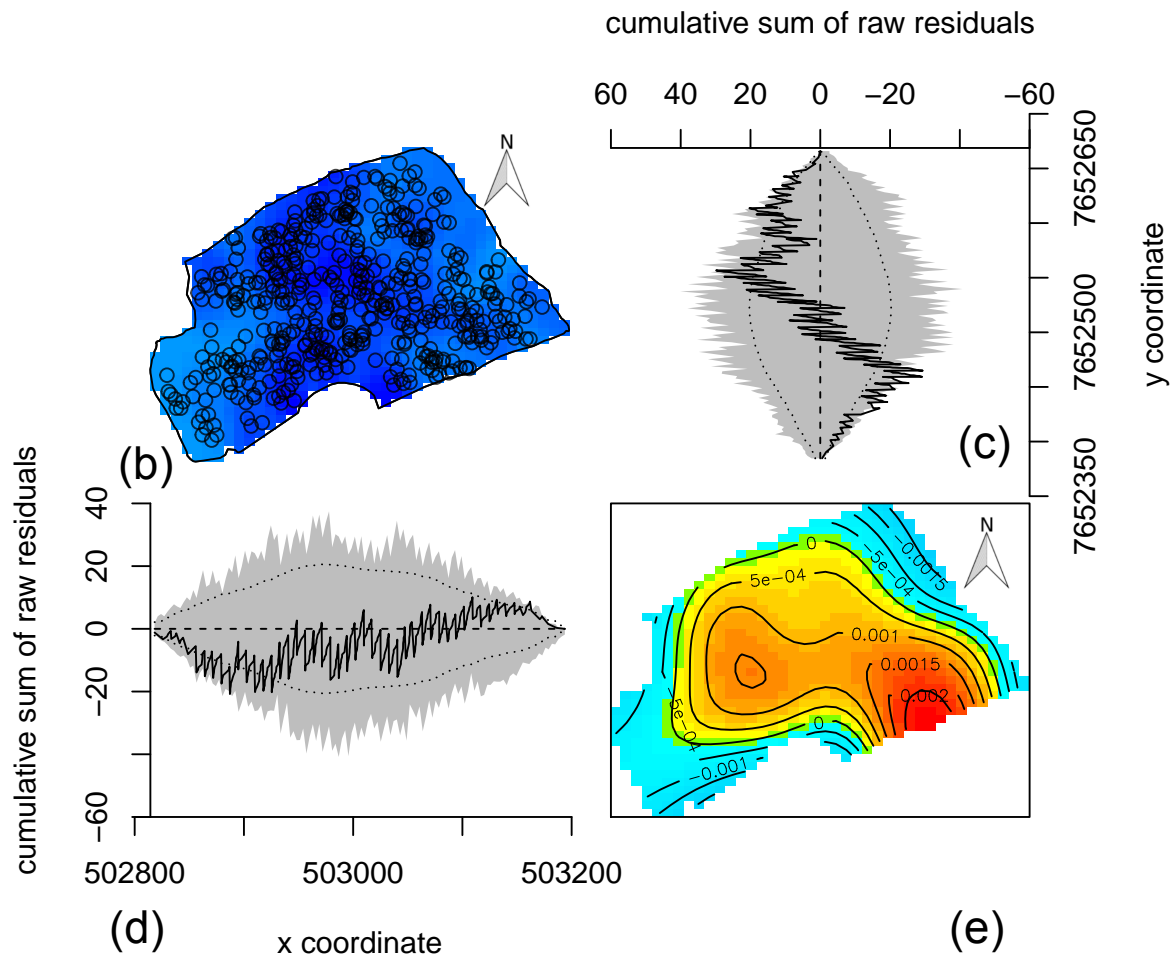


Fonte: Do Autor (2024)

A Figura 4.8 (b) mostra os dados observados, enquanto a Figura 4.8 (e) exibe o mapa dos resíduos suavizados, calculados como a diferença entre os dados observados e os gerados pelo modelo ajustado. As cores no mapa indicam os níveis de superestimação e subestimação, e apesar de pequenas diferenças nos resíduos suavizados, os níveis de superestimação e subestimação são aproximadamente zero. As Figuras 4.8 (c) e 4.8 (d) mostram se há superestimação

ou subestimação de variáveis latentes (não medidas) ao longo das coordenadas x e y , respectivamente. O fato de o modelo ajustado estar ligeiramente fora do envelope indica que, embora as covariáveis utilizadas sejam significativas, elas não explicam completamente a intensidade da distribuição de árvores da espécie. Pode ser que existam variáveis latentes (variáveis não observadas) que poderiam ser incluídas no modelo.

Figura 4.8 – Resíduos do modelo ajustado à espécie *Copaifera langsdorffii*



Fonte: Do Autor (2024)

Os resultados do ajuste do modelo de Poisson não-homogêneo para a espécie *Siparuna guianensis* são apresentados na Tabela 4.3. A interpretação dos parâmetros apresentados na Tabela 4.3 é idêntica à apresentada para os parâmetros na Tabela 4.2. É importante notar que Al e t no modelo geral foram significativos ao nível de 5%, mas perderam significância após o procedimento stepwise ao nível de 5%. No entanto, foram mantidos porque mostraram significância no modelo geral.

Tabela 4.3 – Covariáveis selecionadas pelo método stepwise para a espécie *Siparuna guianensis*.

	Estimativa	S.E.	IC95.inf	IC95.sup	Teste Z	Valor Z
Intercepto	-304,3477380	125,107955	-549,5548232	-59,140653	*	-2,4326810
Al	-1,0454193	1,791749	-4,5571822	2,466344	ns	-0,5834631
t	0,4300871	1,621355	-2,7477105	3,607885	ns	0,2652640
Argila	2,9566876	1,245035	0,5164646	5,396911	*	2,3747835
Silte	3,0052131	1,250725	0,5538363	5,456590	*	2,4027761
Areia	2,9943229	1,243861	0,5564001	5,432246	*	2,4072809

* - estatística significativa a 5% e, ns - estatística não significativa a 5%

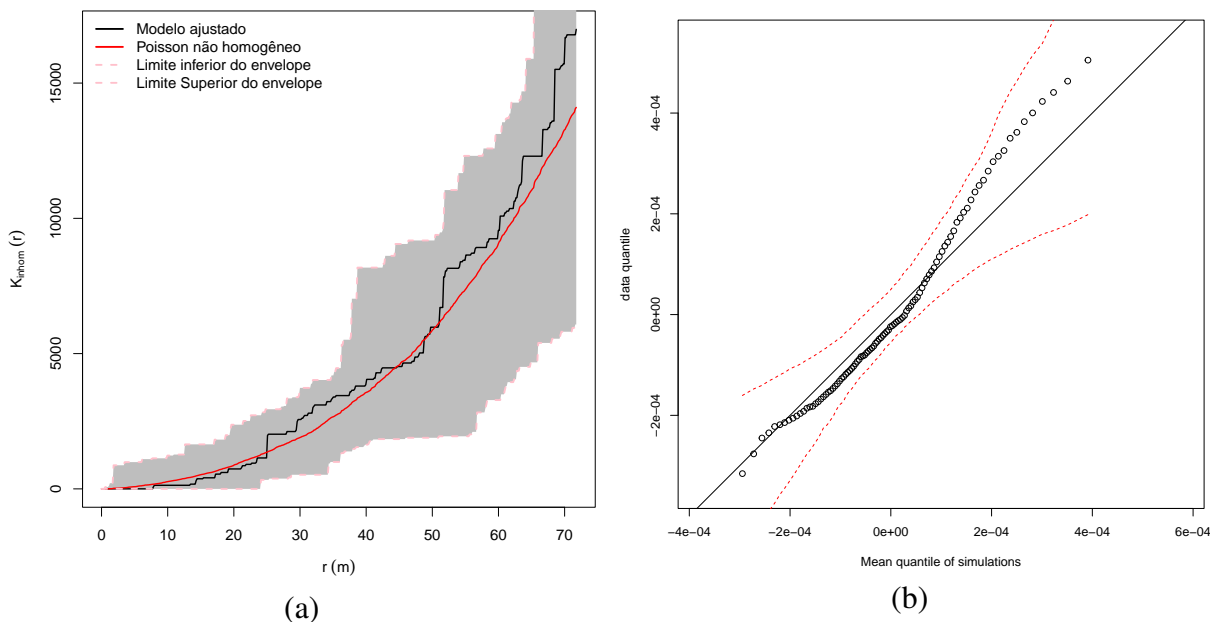
Fonte: Do Autor (2024)

O modelo de Poisson não-homôgeneo ajustado para a espécie *Siparuna guianensis* é dado pela função

$$\lambda(s) = \exp(-304,3477380 - 1,0454193 \cdot Al + 0,4300871 \cdot t + 2,9566876 \cdot Argila + 3,0052131 \cdot Silte + 2,9943229 \cdot Areia), \text{ AIC}=488,8257.$$

Os envelopes de simulação do modelo ajustado usando a função K e o Q-Q plot dos resíduos são apresentados na Figura 4.9.

Figura 4.9 – *Siparuna guianensis*: (a) modelo ajustado e (b) q-q plot

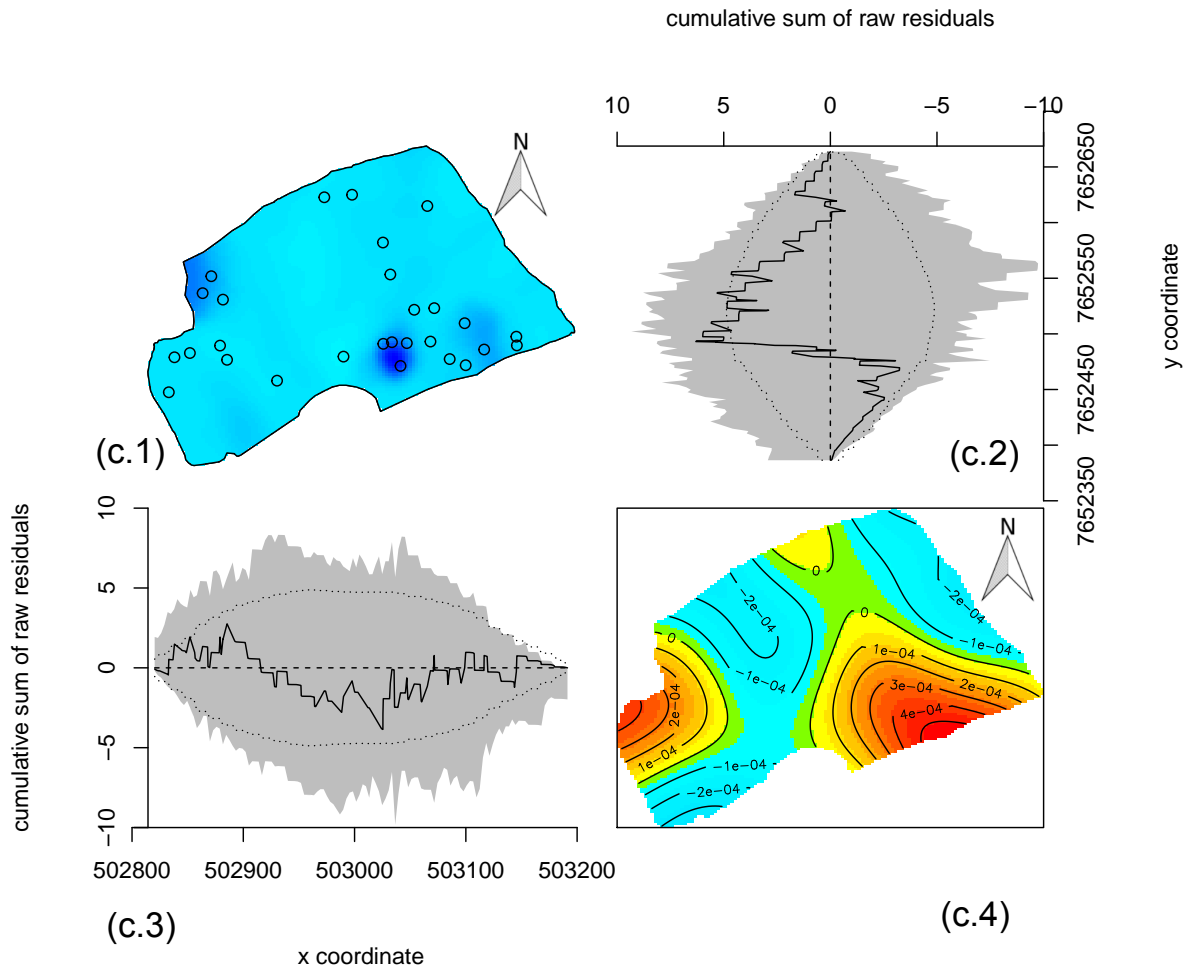


Fonte: Do Autor (2024)

Pode-se observar na Figura 4.9 (a) que o modelo teórico e os dados ajustados são equivalentes. A Figura 4.9 (b) mostra que os resíduos estão dentro das bandas de confiança (envelope), sugerindo que os resíduos seguem uma distribuição normal e, portanto, o modelo captura adequadamente a variação dos dados (tendência espacial), sem padrões sistemáticos não explicados pelo modelo, mesmo na ausência de covariáveis.

Além do Q-Q plot, outra forma plausível de verificar a adequação do modelo é construindo o gráfico da soma dos resíduos cumulativos (Figura 4.10), no qual é possível verificar se o modelo explica perfeitamente as variáveis latentes (variáveis não mensuradas).

Figura 4.10 – Resíduos do modelo ajustado à espécie *Siparuna guianensis*.



Fonte: Do Autor (2024)

A Figura 4.10 (c.1) exibe os dados observados, enquanto a Figura 4.10 (c.4) apresenta o mapa de resíduos suavizados, calculados como a diferença entre os dados observados e os previstos pelo modelo ajustado. As cores indicam níveis de superestimação e subestimação, e embora haja pequenas diferenças no mapa de resíduos suavizados, os níveis de superestimação e subestimação observados são aproximadamente zero. As Figuras 4.10 (c.3) e 4.10 (c.4) mostram se há superestimação ou subestimação de variáveis latentes ao longo das coordenadas x e y , respectivamente, na área de estudo. Como o modelo ajustado está totalmente dentro do envelope, isso indica que as variáveis usadas no modelo de Poisson não-homogêneo foram capazes de capturar perfeitamente (sem superestimação ou subestimação) a intensidade espacial

observada na Figura 4.3. Cinco modelos de Cox foram ajustados para a espécie *Xylopia brasiliensis* conforme ilustrado na Figura 4.11 (a). Os valores de AIC obtidos foram: 390018,5 para o Modelo Thomas, 390014,7 para o Modelo Matérn, 390015,4 para o Modelo Cauchy, 390014,7 para o Modelo Variância Gama, e 390015 para o Modelo Log-Cox Gaussiano. Assim, o Modelo Matérn foi selecionado devido ao menor AIC e à sua praticidade.

Os resultados do ajuste do modelo de Matérn para a espécie *Xylopia brasiliensis* é apresentado na Tabela 4.4

Tabela 4.4 – Covariáveis selecionadas pelo método stepwise para a espécie *Xylopia brasiliensis*.

	Estimativa	S.E.	IC95.inf	IC95.sup	Teste Z	Valor Z
Intercepto	-18,055665603	6,081884813	-29,97594079	-6,135390412	**	-2,968762
pH	-1,382769388	0,433587515	-2,23258530	-0,532953474	**	-3,189136
K	-0,008327173	0,005535509	-0,01917657	0,002522226	ns	-1,504319
Al	-38,967109375	9,600290624	-57,78333324	-20,150885511	***	-4,058951
H+Al	8,772211538	1,938097747	4,97360975	12,570813321	***	4,526197
T	-1,083406579	0,361861648	-1,79264238	-0,374170780	**	-2,993980
MO	0,343474659	0,136052771	0,07681613	0,610133190	*	2,524569

Fonte: Do Autor (2024)

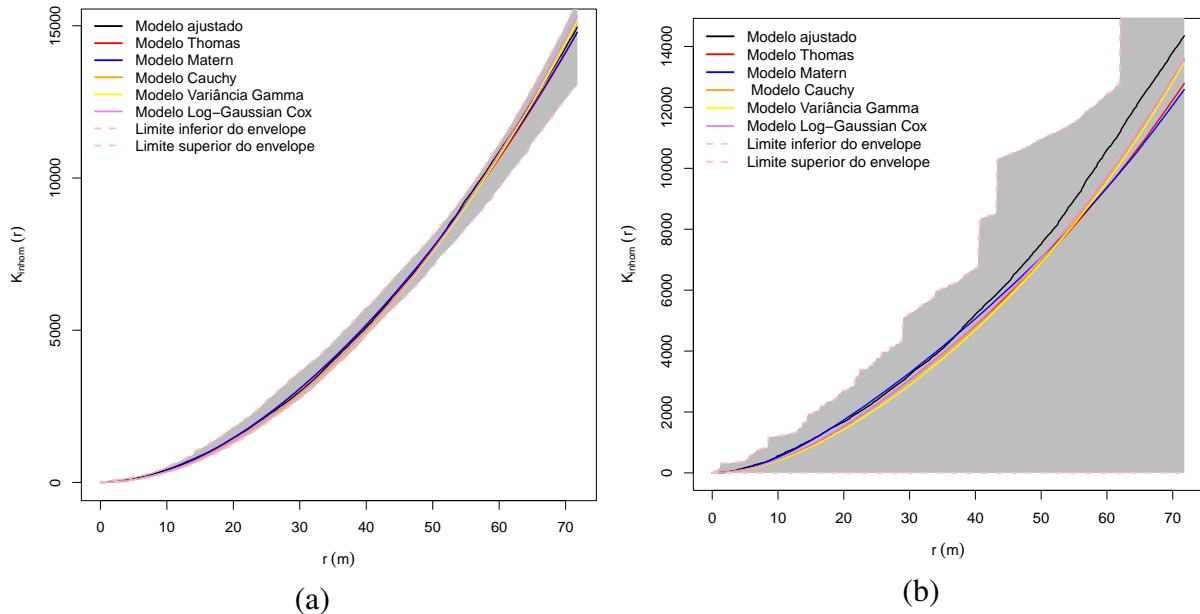
Com base nos resultados apresentados na Tabela 4.4, o modelo de Matérn ajustado para a espécie *Xylopia brasiliensis* é o seguinte:

$$\Lambda(s) = \exp(-18,055665603 - 1,382769388 \cdot pH - 0,008327173 \cdot K - 38,967109375 \cdot Al + 8,772211538 \cdot (\text{acidez potencial}) - 1,083406579 \cdot T + 0,343474659 \cdot MO + \Psi(s)), \text{ onde } \Psi(s) \text{ é o campo aleatório associado.}$$

A interpretação destes parâmetros é idêntica à apresentada para os parâmetros na Tabela 4.2. É importante notar que o potássio (K) no modelo geral foi significativo ao nível de 5%, mas perdeu significância após o procedimento stepwise ao nível de 5%. No entanto, foi mantido porque mostrou significância no modelo geral. Considerando o modelo Matérn, assume-se que plantas da espécie *Xylopia brasiliensis* exibem uma distribuição espacial uniforme dentro da área de estudo, com uma intensidade de plantas progenitoras equivalente a 0,0003251107 plantas por metro quadrado, aproximadamente 4 plantas por hectare. Cada planta progenitora tem entre 13 e 79 descendentes (mudas) distribuídos ao redor dela em um raio de 39 metros. A probabilidade de dois descendentes (mudas) serem do mesmo progenitor é de 39,02%, e a distância média entre eles é de 0,64 metros.

Os mesmos cinco modelos de Cox adotados anteriormente na espécie *X. brasiliensis* também foram utilizados para a espécie *Myrcia splendens* (Figura 4.11 (b)).

Figura 4.11 – Modelos ajustados para: (a) *xylopia brasiliensis* e (b) *myrcia splendens*.



Fonte: Do Autor (2024)

Os valores de AIC obtidos foram: 115812,6 para o Modelo Thomas, 115801,7 para o Modelo Matérn, 115801,7 para o Modelo Cauchy, 115801,7 para o Modelo Variância Gama e 115797,9 para o Modelo Log-Cox Gaussiano. O Modelo Log-Cox Gaussiano foi selecionado devido ao menor AIC. Os resultados do ajuste são apresentados na Tabela 4.5.

Tabela 4.5 – Covariáveis selecionadas pelo método stepwise para a espécie *myrcia splendens*.

	Estimativa	S.E.	IC95.inf	IC95.sup	Teste Z	Valor Z
Intercepto	-290,38414412	47,64879817	-383,7740724	-196,99421581	***	-6,094260
k	-0,09582169	0,01644225	-0,1280479	-0,06359549	***	-5,827774
Ca	-22,29361332	4,52641725	-31,1652281	-13,42199854	***	-4,925223
Mg	-34,11754314	7,57603236	-48,9662937	-19,26879256	***	-4,503352
Al	-23,67233630	4,33779402	-32,1742564	-15,17041624	***	-5,457229
t	17,77594583	3,96118135	10,0121731	25,53971860	***	4,487536
MO	1,54301081	0,39427846	0,7702392	2,31578239	***	3,913505
Argila	2,88574677	0,47444343	1,9558547	3,81563882	***	6,082383
Silte	2,96883056	0,47672364	2,0344694	3,90319173	***	6,227571
Areia	2,94572375	0,47425040	2,0162100	3,87523746	***	6,211326

Fonte: Do Autor (2024)

Com base nos resultados apresentados na Tabela 4.5, o modelo de intensidade ajustado para esta espécie é o seguinte:

$$\Lambda(s) = \exp(-290,38414412 - 0,09582169 \cdot k - 22,29361332 \cdot Ca - 34,11754314 \cdot Mg - 23,67233630 \cdot Al + 17,77594583 \cdot t + 1,54301081 \cdot MO + 2,88574677 \cdot Argila + 2,96883056 \cdot Silte + 2,94572375 \cdot Areia + \Psi(s)),$$

onde $\Psi(s)$, o campo aleatório associado.

A função de covariância para o modelo Log-Cox Gaussiano é dada por

$$\gamma(r) = 0,8405556 \exp(-r/114,2383464).$$

Esta função de covariância indica que para uma distância r igual a zero, as plantas (plantas) da espécie *Myrcia splendens* muito próximas entre si têm uma covariância igual a 0,8405556. No entanto, essa covariância diminui exponencialmente à medida que a distância r aumenta. O parâmetro 114,2383464 representa a escala de decaimento, ou seja, a taxa na qual a covariância decai. Isso significa que a covariância entre as plantas diminui expressivamente quando a distância entre elas se aproxima de 114,2383464 metros.

5 CONCLUSÕES E RECOMENDAÇÕES

Esta dissertação utilizou métodos da estatística espacial, especificamente processos pontuais, para investigar a distribuição espacial de quatro espécies arbóreas *Copaifera langsdorffii*, *Myrcia splendens* e *Siparuna guianensis* localizadas em um fragmento florestal natural do bioma Mata Atlântica.

O estudo da dinâmica espaço-temporal ao longo das últimas três décadas (1987-2017) revelou que as espécies em estudo exibiram baixa mortalidade, com o período de 2015-2017 sendo o único a mostrar uma leve mortalidade associada ao fenômeno El Niño. Apesar da baixa mortalidade, todas as espécies apresentaram crescimento limitado tanto no número de novas plantas quanto no diâmetro à altura do peito.

As covariáveis: pH, potássio (K), fósforo (P), cálcio (Ca), magnésio (Mg), alumínio (Al), acidez potencial (H+Al), soma de bases (SB), capacidade de troca de cátions efetiva (t), capacidade de troca de cátions total (T), saturação de bases (V), saturação de alumínio (m), matéria orgânica (MO), argila, silte e areia foram utilizadas para estimar a intensidade local das árvores dessas espécies. Esta análise mostrou que todas as espécies apresentam configurações com efeitos de primeira ordem, ou seja, apresentam intensidades não homogêneas (tendência) dentro do fragmento florestal no ano de 2017.

A análise das propriedades de segunda ordem (dependência espacial) utilizando funções descritoras não-homogêneas (ex. Função L_{inhom}), mostraram que as espécies *Siparuna guianensis* e *Copaifera langsdorffii* não exibem dependência espacial. Entretanto, as espécies *Xylopia brasiliensis* e *Myrcia splendens* exibiram dependência espacial para agrupamentos.

O modelo de Poisson não-homogêneo foi utilizado para modelar a intensidade das espécies *Siparuna guianensis* e *Copaifera langsdorffii* pois elas exibem apenas tendência espacial. As covariáveis que contribuíram negativamente para intensidade da espécie *Copaifera langsdorffii* incluem acidez potencial, capacidade de troca catiônica efetiva, matéria orgânica e teor de argila. Para a espécie *Siparuna guianensis*, além de uma baixa capacidade de germinação das sementes (abaixo de 30%), o alumínio contribui negativamente na sua intensidade.

Em contraste, *Xylopia brasiliensis* e *Myrcia splendens* exibiram dependência espacial para agregação e, portanto, foram ajustados modelos de Cox. Para a espécie *Xylopia brasiliensis*, a acidez potencial e a matéria orgânica foram as principais influências positivas na sua intensidade. Para a espécie *Myrcia splendens*, as covariáveis que influenciaram negativamente na sua intensidade foram: umidade, cálcio, magnésio e alumínio.

Finalmente, os resultados obtidos são importantes para mostrar a importância dos métodos estatísticos utilizados para entender a distribuição espacial, não somente das espécies arbóreas *Myrcia splendens*, *Siparuna guianensis*, *Xylopia brasiliensis* e *Copaifera langsdorffii*, mas também de outras espécies arbóreas existentes em fragmentos de florestas naturais, o que pode contribuir para a preservação desses fragmentos e, conseqüentemente, dos biomas que eles representam.

Para estudos futuros, recomenda-se ampliar a investigação para incluir outros fatores que podem estar associados aos padrões e tendências observados neste estudo, como fatores genéticos e tipos de germinação de sementes. Além disso, é aconselhável integrar modelos espaço-temporais e incorporar marcas como o diâmetro à altura do peito (DAP) nos modelos. Outra sugestão importante é o georreferenciamento de todas as espécies florestais da Mata Atlântica presentes na Universidade Federal de Lavras (UFLA). Essa abordagem permitirá explorar a influência das interações entre duas ou mais espécies florestais distintas, possibilitando a investigação de possíveis relações de dependência interespecífica.

REFERÊNCIAS

- BADDELEY, A. Validation of statistical models for spatial point patterns. In: **Statistical Challenges in Modern Astronomy IV**. [S.l.: s.n.], 2007. v. 371, p. 22.
- BADDELEY, A. Multivariate and marked point processes. **Handbook of spatial statistics**, Chapman & Hall/CRC Handb. Mod. Stat. Methods, p. 371–402, 2010.
- BADDELEY, A. et al. Residual diagnostics for covariate effects in spatial point process models. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 22, n. 4, p. 886–905, 2013.
- BADDELEY, A. et al. Fundamental problems in fitting spatial cluster process models. **Spatial Statistics**, Elsevier, v. 52, p. 100709, 2022.
- BADDELEY, A. et al. Residuals and diagnostics for spatial point processes. In: CITESEER. [S.l.], 2019.
- BADDELEY, A. et al. **Analysing spatial point patterns in R**. [S.l.], 2008.
- BADDELEY, A.; RUBAK, E.; TURNER, R. **Spatial Point Patterns: Methodology and Applications with R**. London: Chapman and Hall/CRC Press, 2015. ISBN 9781482210200. Disponível em: <<https://www.routledge.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/>>.
- BADDELEY, A.; TURNER, R. spatstat: An R package for analyzing spatial point patterns. **Journal of Statistical Software**, v. 12, n. 6, p. 1–42, 2005.
- BADDELEY, A. et al. Residual analysis for spatial point processes (with discussion). **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 67, n. 5, p. 617–666, 2005.
- BADDELEY, A. J.; LIESHOUT, M. V. Area-interaction point processes. **Annals of the Institute of Statistical Mathematics**, Springer, v. 47, p. 601–619, 1995.
- BADDELEY, A. J.; MØLLER, J.; WAAGEPETERSEN, R. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. **Statistica Neerlandica**, Wiley Online Library, v. 54, n. 3, p. 329–350, 2000.
- BELCHIOR, I. B. Avaliação de queimadas por sensoriamento remoto no parque nacional de itatiaia. Universidade Federal de Lavras, 2022.
- BELL, W. W. **Special functions for scientists and engineers**. [S.l.]: Courier Corporation, 2004.
- BENNETT, A. C. et al. Sensitivity of south american tropical forests to an extreme climate anomaly. **Nature Climate Change**, Nature Publishing Group UK London, v. 13, n. 9, p. 967–974, 2023.
- BESAG, J. Discussion of dr ripley’s paper. **Journal of the Royal Statistical Society, Series B**, v. 39, n. 2, p. 193–195, 1977.

CARVALHO, P. E. R. Espécies arbóreas brasileiras. Brasília, DF: Embrapa Informação Tecnológica; Colombo: Embrapa Florestas . . . , 2014.

CASELLA, G.; BERGER, R. L. **Statistical inference**. [S.l.]: Duxbury, 2001.

CLARK, P. J.; EVANS, F. C. Distance to nearest neighbor as a measure of spatial relationships in populations. **Ecology**, JSTOR, v. 35, n. 4, p. 445–453, 1954.

COSTA, J. **Copaifera in Flora e Funga do Brasil**. 2024. Jardim Botânico do Rio de Janeiro. Available at: <<https://floradobrasil.jbrj.gov.br/FB22895>>. consulta publica.uc.citacao.acesso.em 01 mai. 2024.

CRAWLEY, M. J. **Plant ecology**. [S.l.]: John Wiley & Sons, 2009.

CRESSIE, N. **Statistics for Spatial Data: Wiley Series in Probability and Statistics**. [S.l.]: Wiley-Interscience, 1993. 908 p.

CRONIE, O.; LIESHOUT, M. V. Bandwidth selection for kernel estimators of the spatial intensity function. **arXiv preprint arXiv:1611.10221**, 2016.

DAVIDE, A.; FARIA, J.; BOTELHO, S. Propagação de espécies florestais. **Belo Horizonte: CEMIG/UFLA/FAEPE**, 1995.

DIGGLE, P. J. Statistical analysis of spatial point patterns. In: . [S.l.: s.n.], 2003.

DIGGLE, P. J. **Nonparametric methods**. [S.l.]: Chapman & Hall/CRC Handb. Mod. Stat. Methods, 2010.

DIGGLE, P. J. **Statistical analysis of spatial and spatio-temporal point patterns**. [S.l.]: CRC press, 2013.

DIGGLE, P. J.; GATES, D. J.; STIBBARD, A. A nonparametric estimator for pairwise-interaction point processes. **Biometrika**, Oxford University Press, v. 74, n. 4, p. 763–770, 1987.

DIGGLE, P. J.; GRATTON, R. J. Monte carlo methods of inference for implicit statistical models. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 46, n. 2, p. 193–212, 1984.

FAO. **Global forest resources assessment 2010-Key finding**. [S.l.]: The Food and Agriculture Organisation of the United Nations Rome, 2020.

FERRARI, G. et al. Análise de processos pontuais marcados aplicados às características genéticas em árvores. In: **Anais do II Simpósio de Geostatística aplicada em Ciências Agrárias**. [S.l.: s.n.], 2011.

FERREIRA, D. F. **Fundamentos de probabilidade**. Lavras: UFLA, 2020.

FRA. **Global forest resources assessment, 2020: Main report**. [S.l.]: Food and Agriculture Organization of the United Nations, 2020.

- FRUEHAUF, A. L.; SILVA, P. V. da; LOMBARDO, M. A. O uso da geotecnologia e análise estatística dos incêndios florestais na floresta estadual edmundo navarro de andrade, rio claro, são paulo, brasil the use of geotechnology and statistical analysis of forest fires in the state forest edmundo navarro de andrade, rio claro, são paulo. **Brazilian Journal of Development**, v. 8, n. 1, p. 449–464, 2022.
- GARCÍA-PORTUGUÉS, E. **Notes for Nonparametric Statistics**. [s.n.], 2024. Version 6.9.1. ISBN 978-84-09-29537-1. Disponível em: <<https://bookdown.org/egarpor/NP-UC3M/>>.
- GATRELL, A. C. et al. Spatial point pattern analysis and its application in geographical epidemiology. **Transactions of the Institute of British geographers**, JSTOR, p. 256–274, 1996.
- GELFAND, A. E. et al. **Handbook of spatial statistics**. [S.l.]: CRC press, 2010.
- GONZÁLEZ, J. A.; MORAGA, P. Non-parametric analysis of spatial and spatio-temporal point patterns. The R Foundation, 2023.
- HECHT, S. B. Forests lost and found in tropical latin america: the woodland ‘green revolution’. **The Journal of Peasant Studies**, Taylor & Francis, v. 41, n. 5, p. 877–909, 2014.
- HOPKINS, B.; SKELLAM, J. G. A new method for determining the type of distribution of plant individuals. **Annals of Botany**, Oxford University Press, v. 18, n. 2, p. 213–227, 1954.
- HORÁČKOVÁ, M.; ŘEHOUNKOVÁ, K.; PRACH, K. Are seed and dispersal characteristics of plants capable of predicting colonization of post-mining sites? **Environmental Science and Pollution Research**, Springer, v. 23, n. 14, p. 13617–13625, 2016.
- ILLIAN, J. et al. **Statistical analysis and modelling of spatial point patterns**. [S.l.]: John Wiley & Sons, 2008.
- ILLIAN, J. B. Spatial and spatio-temporal point processes in ecological applications. In: **Handbook of environmental and ecological statistics**. [S.l.]: Chapman and Hall/CRC, 2019. p. 97–131.
- INPE. **Instituto Nacional de Pesquisas Espaciais**. 2023. <<http://terrabrasilis.dpi.inpe.br/queimadas/aq1km/>>.
- ISHAM, V. Spatial point process models. **Handbook of spatial statistics**, CRC Press, p. 283–298, 2010.
- JALILIAN, A.; GUAN, Y.; WAAGEPETERSEN, R. Decomposition of variance for spatial cox processes. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 40, n. 1, p. 119–137, 2013.
- JALILIAN, A.; SAFARI, A.; SOHRABI, H. Modeling spatial patterns and species associations in a hyrcanian forest using a multivariate log-gaussian cox process. **Journal of Statistical Modelling: Theory and Applications**, Yazd University, v. 1, n. 2, p. 59–76, 2020.
- JUNIOR, J. J. et al. Time-stability of soil water content (swc) in an atlantic forest-latosol site. **Geoderma**, Elsevier, v. 288, p. 64–78, 2017.
- LEININGER, T. J. **Bayesian analysis of spatial point patterns**. Tese (Doutorado) — Duke University, 2014.

- LIESHOUT, M. V.; BADDELEY, A. A nonparametric measure of spatial interaction in point patterns. **Statistica Neerlandica**, Wiley Online Library, v. 50, n. 3, p. 344–361, 1996.
- LIMA, R. R. de. **Modelagem espaço temporal para dados de incidência de doenças em plantas**. Tese (Doutorado) — Tese piracicaba-SP, 2005.
- LOKE, L. H.; CHISHOLM, R. A. Measuring habitat complexity and spatial heterogeneity in ecology. **Ecology Letters**, Wiley Online Library, v. 25, n. 10, p. 2269–2288, 2022.
- LORENZI, H. et al. Plantas medicinais no brasil: nativas e exóticas. 2021.
- LORENZI, H. et al. Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do brasil. Plantarum Nova Odessa, 1992.
- MATEUS, A. L. S. S. **Proposição de novas metodologias para análise de aleatoriedade em processos pontuais no espaço-tempo**. Tese (Tese (Doutorado em Estatística e Experimentação Agropecuária)) — Universidade Federal de Lavras, Lavras, 2013. 138 p.
- MØLLER, J.; SYVERSVEEN, A. R.; WAAGEPETERSEN, R. P. Log gaussian cox processes. **Scandinavian journal of statistics**, Wiley Online Library, v. 25, n. 3, p. 451–482, 1998.
- MOLLER, J.; WAAGEPETERSEN, R. P. **Statistical inference and simulation for spatial point processes**. [S.l.]: CRC press, 2003.
- MØLLER, J.; WAAGEPETERSEN, R. P. Modern statistics for spatial point processes. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 34, n. 4, p. 643–684, 2007.
- MORAGA, P. **Spatial Statistics for Data Science: Theory and Practice with R**. [S.l.]: CRC Press, 2023.
- MORISITA, M. Measuring of the dispersion of individuals and analysis of the distributional patterns. **Memoirs of the Faculty of Science, E 2, Kyushu Univ. Ser.**, Kyushu University, 1959.
- NHANCOLOLO, A. M. et al. Comparação entre o método laboratorial e o penetrômetro de stolf na análise da densidade do solo: um estudo utilizando abordagens geoestatísticas. **Sigmae**, v. 13, n. 1, p. 63–78, 2024.
- NIGHTINGALE, G. F. et al. Area interaction point processes for bivariate point patterns in a bayesian context. **Journal of Environmental Statistics**, UCLA Department of Statistics, v. 9, n. 2, 2019.
- OKABE, M.; TANEMURA, M. Bayesian estimation of soft-core potential models for spatial point patterns. **Journal of the Japan Statistical Society**, THE JAPAN STATISTICAL SOCIETY, v. 36, n. 2, p. 121–147, 2006.
- OLIVEIRA-FILHO, A. T.; MELLO, J. Márcio de; SCOLFORO, J. R. S. Effects of past disturbance and edges on tree community structure and dynamics within a fragment of tropical semideciduous forest in south-eastern brazil over a five-year period (1987–1992). **Plant Ecology**, Springer, v. 131, p. 45–66, 1997.
- OLIVEIRA, W. A. d. **Métodos de processos pontuais para análise espacial de espécies arbóreas nativas de um fragmento florestal da região de Lavras-MG**. Dissertação (Mestrado) — UFLA, Lavras, 2022.

- PAIVA, P. D. d. O. et al. Performance of native species in urban afforestation of public pathways in lavras-mg, brazil. **Ornamental Horticulture**, SciELO Brasil, v. 28, p. 161–171, 2022.
- PANSONATO, M. P. **Fatores determinantes dos padrões de similaridade, distribuição e abundância de árvores em dois ambientes da Mata Atlântica**. Tese (Doutorado) — Universidade de São Paulo, 2019.
- PEBESMA, E.; BIVAND, R. **Spatial data science: With applications in R**. [S.l.]: Chapman and Hall/CRC, 2023.
- PEIXOTO, A.; LIRIO, E.; PIGNAL, M. **Siparunaceae in Flora e Funga do Brasil**. 2024. Jardim Botânico do Rio de Janeiro. Available at: <<https://floradobrasil.jbrj.gov.br/FB14548>>. consulta publica. uc. citacao. acesso. em 01 mai. 2024.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <<https://www.R-project.org/>>.
- RAI, V. **Spatial ecology patterns and processes**. [S.l.]: Bentham Science Publishers, 2013. v. 1.
- REFLORA. **Xylopia in Flora e Funga do Brasil**. 2024. Jardim Botânico do Rio de Janeiro. Available at: <<https://floradobrasil.jbrj.gov.br/FB110560>>. consulta publica. uc. citacao. acesso. em 01 mai. 2024.
- RENNER, S. S.; HAUSNER, G. **Siparunaceae**. [S.l.: s.n.], 2005.
- SANTOS, M. et al. **Myrcia in Flora e Funga do Brasil**. 2024. Jardim Botânico do Rio de Janeiro. Available at: <<https://floradobrasil.jbrj.gov.br/FB10759>>. consulta publica. uc. citacao. acesso. em 01 mai. 2024.
- SCALON, J. D. **Análise de Dados Espaciais com Aplicações em R**. Lavras: Ed. UFLA, 2024. 275 p.
- SCALON, J. D. et al. Statistical characterization of spatial and size distributions of particles in composite materials used in the manufacturing of biomedical instruments. **Brazilian Journal of Biometrics**, v. 40, n. 4, p. 428–441, 2022.
- SCHELHAS, J.; BRANDEIS, T. J.; RUDEL, T. K. Planted forests and natural regeneration in forest transitions: patterns and implications from the us south. **Regional environmental change**, Springer, v. 21, n. 1, p. 8, 2021.
- SCHULZE, E.-D.; BECK, E.; MÜLLER-HOHENSTEIN, K. **Plant ecology**. [S.l.]: Springer Science & Business Media, 2005.
- SEIDLER, T. G.; PLOTKIN, J. B. Seed dispersal and spatial pattern in tropical trees. **PLoS biology**, Public Library of Science San Francisco, USA, v. 4, n. 11, p. e344, 2006.
- SHIMATANI, K. Multivariate point processes and spatial variation of species diversity. **Forest Ecology and Management**, Elsevier, v. 142, n. 1-3, p. 215–229, 2001.
- STEIN, A.; GERSTNER, K.; KREFT, H. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. **Ecology letters**, Wiley Online Library, v. 17, n. 7, p. 866–880, 2014.

TILMAN, D. Mechanisms of plant competition. **Plant ecology**, Wiley Online Library, p. 239–261, 1996.

VIEIRA, S. R. Investigating the driving forces of deforestation in the state of minas gerais, brazil. 2011.

WAAGEPETERSEN, R.; GUAN, Y. Two-step estimation for inhomogeneous spatial point processes. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 71, n. 3, p. 685–702, 2009.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<https://ggplot2.tidyverse.org>>.

WIEGAND, T.; MOLONEY, K. A. **Handbook of spatial point-pattern analysis in ecology**. [S.l.]: CRC press, 2014.

A ANEXO - Códigos em R

```
#===== PACOTES =====#
library(spatstat);library(readxl);library(xlsx);library(dplyr);
library(geobr);library(sf);library(sp);library(ggmap);library(maps);
library(tmap);library(writexl) library(ggplot2);library(dplyr);
library(viridis);library(sparr);library(viridis);library(segregation)
#=====Leitura e filtragem dos dados=====#
setwd("E:/PROGRAMA DE MESTRADO UFLA/3o SEMESTRE/DISSERTACAO")
Dados <- read_xlsx("DADOS_MATINHA (1).xlsx")
DAP1987e2017 <- read_excel('DAP1987e2017.xlsx')
Todas <- filter(Dados, Dados$DAP2017!=0)
#DAP>=5cm,Se DAP=0 a esp cie morreu
DAP2017 <- Todas[,c(4,14)]; nomes(Todas)
#===== area de estudo=====#
M <- st_read("Shape matinha/Matinha.shp")
sf_data <- st_zm(M) # transformando para V.Bidimenssionais eliminando z
janela<- as.owin(sf_data) # criando janela
#=====1987=====#
#Separando as especies para o ano 1987 (inicio do inventario florestal)=#
Dados1987 <- filter(Dados, Dados$DAP1987!=0)
#===== Esp cie Copaifera langsdorffii
Copaifera1987<-filter(Dados1987,Dados1987$Esp cie=="Copaifera langsdorffii")
#=====Transformando em objecto ppp (point pattern process)
Copaifera1987.ppp<-ppp(Copaifera1987$X,Copaifera1987$Y>window = janela)
any(duplicated.ppp(Copaifera1987.ppp)) #vendo se tem mais de uma planta na
#mesma posicao)
#===== Xylophia brasiliensis
Xylophia1987<-filter(Dados1987,Dados1987$Esp cie=="Xylophia brasiliensis")
Xylophia1987.ppp<-ppp(Xylophia1987$X,Xylophia1987$Y>window=janela)
any(duplicated.ppp(Xylophia1987.ppp))
#===== Siparuna guianensis
Siparuna1987<-filter(Dados1987,Dados1987$Esp cie=="Siparuna guianensis")
Siparuna1987.ppp<-ppp(Siparuna1987$X,Siparuna1987$Y>window=janela)
any(duplicated.ppp(Siparuna1987.ppp))
```

```

#===== Myrcia splendens
Myrcial1987<-filter(Dados1987,Dados1987$Esp cie=="Myrcia splendens")
Myrcial1987.ppp <-ppp(Myrcial1987$X,Myrcial1987$Y,window=janela)
any(duplicated.ppp(Myrcial1987.ppp))
#Construindo mapas com as respectivas esp cies florestais (ponto)
par(mfrow=c(2,2), mar = c(0,0,1,0),oma=c(0.0001, 0.0001,0.0001,0.0001))
plot(Copaiferal1987.ppp, main="Copaifera langsdorffii", cex=0.6);
plot(Xylopial1987.ppp, main="Xylopia brasiliensis", cex=0.6);
plot(Siparuna1987.ppp, main="Siparuna guianensis", cex=0.6);
plot(Myrcial1987.ppp,main="Myrcia splendens", cex=0.6 )
#====Filtrando dados para identificar esp cies mortas e vivas=====#
#=====Novos individuos em 1992
Dados1992nov <- filter(Dadoss, Dadoss$DAP1987==0 & Dadoss$DAP1992!=0)
table(Dados1992nov$Esp cie)
#=====individuos mortos at 1992
Dados1992mort <- filter(Dadoss, Dadoss$DAP1987!=0 & Dadoss$DAP1992==0)
table(Dados1992mort$Esp cie)
#=====Novos individuos em 1996
Dados1996nov <- filter(Dadoss, Dadoss$DAP1992==0 & Dadoss$DAP1996!=0)
table(Dados1996nov$Esp cie)
#====Novos individuos mortos at 1996
Dados1996mort <- filter(Dadoss, Dadoss$DAP1992!=0 & Dadoss$DAP1996==0)
table(Dados1996mort$Esp cie) #DAP=9.4
#=====Novos individuos em 2001
Dados2001nov <- filter(Dadoss, Dadoss$DAP1996==0 & Dadoss$DAP2001!=0)
table(Dados2001nov$Especie)
#Novos individuos mortos ate 2001
Dados2001mort <- filter(Dadoss, Dadoss$DAP1996!=0 & Dadoss$DAP2001==0)
table(Dados2001mort$Esp cie)
#=====Novos individuos em 2006
Dados2006nov <- filter(Dadoss, Dadoss$DAP2001==0 & Dadoss$DAP2006!=0)
table(Dados2006nov$Esp cie)
#Novos individuos mortos at 2006
Dados2006mort <- filter(Dadoss, Dadoss$DAP2001!=0 & Dadoss$DAP2006==0)

```

```

table(Dados2006mort$Esp cie)
#====Novos individuos em 2010
Dados2010nov <- filter(Dadoss, Dadoss$DAP2006==0 & Dadoss$DAP2010!=0)
table(Dados2010nov$Esp cie)
#Novos individuos mortos at 2010
Dados2010mort <- filter(Dadoss, Dadoss$DAP2006!=0 & Dadoss$DAP2010==0)
table(Dados2010mort$Esp cie)
Dados2010mort #Nao morreu houve erro no lan amento pq em 2006 tinha
#DAP = 6.2 em 2015 teve 6.4 e 2017 6.9
#====Novos individuos em 2015
Dados2015nov <- filter(Dadoss, Dadoss$DAP2010==0 & Dadoss$DAP2015!=0)
table(Dados2015nov$Esp cie)
#Novos individuos mortos at 2015
Dados2015mort <- filter(Dadoss, Dadoss$DAP2010!=0 & Dadoss$DAP2015==0)
table(Dados2015mort$Esp cie)
#====Novos individuos em 2017
Dados2017nov <- filter(Dadoss, Dadoss$DAP2015==0 & Dadoss$DAP2017!=0)
table(Dados2017nov$Esp cie)
#Novos individuos mortos at 2017
Dados2017mort <- filter(Dadoss, Dadoss$DAP2015!=0 & Dadoss$DAP2017==0)
dados2017<-filter(Dados2017mort,
Dados2017mort$Esp cie=='Siparuna guianensis')
table(Dados2017mort$Esp cie)#muita mortalidade, fenomeno El nino
#====Base de dados novas e novas registradas por ano=====#
Especies1987a2017 <- read_excel('Especiesano.xlsx')
#====Dados descritivos das esp cies por anos
par(mfrow=c(1,2),mar=c(0,0,1,0),oma=c(0.0001,0.0001,0.0001,0.0001),cex=0.4)
Especies1987a2017$Ano <- factor(Especies1987a2017$Ano)
ggplot(data = Especies1987a2017,aes(x =Ano,y = Ocorrencia,fill=Especie)) +
  geom_bar(stat = 'identity', position = "dodge") +
  geom_text(aes(label=Ocorrencia), position = position_dodge(width=0.9),
  vjust = -0.9), theme_classic()+
  labs(x="Ano do invent rio florestal",y="N mero de plantas registradas")+
  theme(legend.position = "bottom")

```

```

#==== Gráfico Variacao do DAP para 1987 e 2017
DAP1987a2017 <- read_excel('DAP1987e2017.xlsx')
ggplot(data =DAP1987a2017,aes(x = Esp cie ,y=DAP,fill= as.factor(Ano))) +
  geom_boxplot(alpha = 0.8, color = "black",size = 0.5, coef = 1.5)+
  # Adicionando as whiskers
stat_boxplot(geom = 'errorbar')+#Definindo as cores preenchimento manualmente
  labs(x = "Type of forest species",y = "Diameter at breast height (DBH)",
  fill = "Ano") +# Adicionando r tulos aos eixos e legenda de cores
  theme_classic()+
  theme(legend.position = "bottom")
#=====Esp cies de 2017 apenas=====#
Todas <- filter(Dadoss, Dadoss$DAP2017!=0)
Todas.Especies <- ppp(Todas$X, Todas$Y,
marks=as.factor(Todas$Esp cie), window=janela)
#===== Esp cie Copaifera langsdorffii
Copaifera <- filter(Todas, Todas$Esp cie == "Copaifera langsdorffii")
Copaifera.ppp <- ppp(Copaifera$X, Copaifera$Y, window = janela)
#===== Xylopia brasiliensis
Xylopia <- filter(Todas, Todas$Esp cie == "Xylopia brasiliensis")
Xylopia.ppp <- ppp(Xylopia$X, Xylopia$Y, window = janela)
#===== Siparuna guianensis
Siparuna <- filter(Todas, Todas$Esp cie == "Siparuna guianensis")
Siparuna.ppp <- ppp(Siparuna$X, Siparuna$Y, window = janela)
#===== Myrcia splendens
Myrcia <- filter(Todas, Todas$Esp cie == "Myrcia splendens")
Myrcia.ppp <- ppp(Myrcia$X, Myrcia$Y, window = janela)
#Construindo mapas com as respectivas esp cies florestais(ponto)
par(mfrow=c(2,2), mar = c(0,0,1,0),oma=c(0.0001, 0.0001, 0.0001, 0.0001))
plot(Copaifera.ppp, main="Copaifera langsdorffii", cex=0.6);
plot(Xylopia.ppp, main="Xylopia brasiliensis", cex=0.6);
plot(Siparuna.ppp, main="Siparuna guianensis", cex=0.6);
plot(Myrcia.ppp,main="Myrcia splendens", cex=0.6 )
#====Distribui o espacial das esp cies por DBH=====#
Dados1987 <- filter(Dadoss, Dadoss$DAP1987!=0)

```

```

DAP1987 <- Dados1987[,c(4,7)]
#===== Copaifera langsdorffii
Dados1987cop <- filter(Dados1987,
Dados1987$Esp cie=='Copaifera langsdorffii')
DAP1987c<- Dados1987cop$DAP1987
DAP1987ppp <- ppp(Dados1987cop$X , Dados1987cop$Y ,
window = janela , marks = DAP1987c)
DAP2017c<- Copaifera$DAP2017
DAP2017ppp <- ppp(Copaifera$X,Copaifera$Y,window=janela,marks=DAP2017c)
#===== Xylophia brasiliensis
Dados1987xyl <-filter(Dados1987, Dados1987$Esp cie=='Xylophia brasiliensis')
DAP1987xyl<- Dados1987xyl$DAP1987
DAP1987xylppp<-ppp(Dados1987xyl$X,Dados1987xyl$Y,window=janela,
marks=DAP1987xyl)
DAP2017xyl<- Xylophia$DAP2017
DAP2017pppxyl <- ppp(Xylophia$X,Xylophia$Y,window=janela,marks=DAP2017xyl)
#===== Siparuna guianensis
Dados1987sivr <- filter(Dados1987,Dados1987$Esp cie=='Siparuna guianensis')
DAP1987sivr<- Dados1987sivr$DAP1987
DAP1987sivrppp <-ppp(Dados1987sivr$X,Dados1987sivr$Y,window=janela,
marks = DAP1987sivr)
DAP2017sivr<-Siparuna$DAP2017
DAP2017pppsivr <- ppp (Siparuna$X,Siparuna$Y, window = janela ,
marks = DAP2017sivr)
#===== Myrcia splendens
Dados1987Myr <- filter(Dados1987,Dados1987$Esp cie=='Myrcia splendens')
DAP1987Myr<- Dados1987Myr$DAP1987
DAP1987Myrppp <- ppp(Dados1987Myr$X,Dados1987Myr$Y,window = janela ,
marks = DAP1987Myr)
DAP2017Myr<-Myrcia$DAP2017
DAP2017pppMyr <- ppp (Myrcia$X,Myrcia$Y,window=janela,marks = DAP2017Myr)
#=====Plots
par(mfrow=c(2,2),mar = c(0,0,1,0),oma=c(0.0001,0.0001,0.0001,0.0001))
plot(DAP1987ppp, main='Copaifera langsdorffii')

```

```

plot(DAP2017ppp , main='Copaifera langsdorffii')
plot(DAP1987xylppp, main='Xylopiya brasiliensis')
plot(DAP2017pppxyl, main='Xylopiya brasiliensis')
plot(DAP1987sivrppp, main='Siparuna guianensis')
plot(DAP2017pppsivr, main='Siparuna guianensis')
plot(DAP1987Myrppp, main='Myrcia splendens')
plot(DAP2017pppMyr, main='Myrcia splendens')
#=====Estadísticas descriptivas=====#
#=====Intensidade
summary(Todas.Especies)
#=====Adaptative density
?adaptive.density()
par(mfrow=c(2,2), mar = c(0,0,1,1), oma=c(0.0001, 0.0001,0.0001,0.0001))
plot (adaptive.density(Copaifera.ppp,method="kernel"),
      main = "Copaifera langsdorffii",col = viridis(100))
plot (Copaifera.ppp, add = TRUE, col="white",cex = 0.6)
plot(adaptive.density(Xylopiya.ppp,method="kernel"),
      main = "Xylopiya brasiliensis",col = viridis(100))
plot (Xylopiya.ppp, add = TRUE, col="white",cex = 0.6)
plot(adaptive.density(Siparuna.ppp,method="kernel"),
      main = "Siparuna guianensis",col = viridis(100))
plot (Siparuna.ppp, add = TRUE, col="white",cex = 0.6)
plot(adaptive.density(Myrcia.ppp,method="kernel"),
      main = "Myrcia splendens",col = viridis(100))
plot (Myrcia.ppp, add = TRUE, col="white",cex = 0.6)
#=====Estrutura de dependência =====#
par(mfrow=c(1,2), mai=c(1, 1, 0.1, 0.1))
#par(mfrow=c(1,3), mar = c(4,4,.1,.1), oma=c(0.1, 0.1, 0.1, 0.1))
plot(envelope(Copaifera.ppp, fun = "Jinhom", diggle = T,
             nsim=200),main='',legend=F, ylab=expression(J~('r')))
plot(envelope(Copaifera.ppp, fun = "Jest", diggle = T,
             nsim=100),main='',legend=F,border = "white")
#=====Copaifera langsdorffii
par(mfrow=c(1,2), mai=c(0.8, 0.8, 0.3, 0.1))

```



```

#=====Modelagem espacial=====#
#=====Smoothing of the covariates=====#
#=====pH
pH <- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  pH= Todas$pH)
min(Todas$pH)
max(Todas$pH)
sd(Todas$pH)
pH_ppp <- as.ppp(pH, W = janela )
pHcolfun <- spatstat.options("image.colfun")
pHcolmap <- colourmap(pHcolfun(100), range = c(3.9, 4.4))
                                     #n n mero de cores no mapa
pHsymmap <- symbolmap(cols=pHcolmap,size =.8,pch=19,range = c(3.9,4.4))
plot(pH_ppp, symap = pHsymmap, main = "pH at locations",legend = FALSE)
#Agora suavizamos os dados de pH e os plotamos
phsigma <- bw.ppl(pH_ppp)
pHsmooth <- Smooth(pH_ppp, sigma =phsigma)
plot(pHsmooth, cols = pHcolmap, main = "Smoothed pH")
contour(pHsmooth, add = TRUE)
#=====k
k<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  k= Todas$K
)
min(Todas$K)
max(Todas$K)
sd(Todas$K)
k_ppp <- as.ppp(k, W = janela )
kcolfun <- spatstat.options("image.colfun")
kcolmap <- colourmap(kcolfun(100), range = c(19, 69))

```

```

#n numero de cores no mapa
ksymmap <- symbolmap(cols=kcolmap, size =.8,pch =19,range =c(19, 69))
plot(k_ppp, symap = ksymmap, main = "k at locations", legend = FALSE)
#Agora suavizamos os dados de k e os plotamos:
ksigma <- bw.ppl(k_ppp)
ksmooth <- Smooth(k_ppp, sigma = ksigma)
plot(ksmooth, cols = kcolmap, main = "Smoothed k")
contour(ksmooth, add = TRUE)
#===== P
P<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  P= Todas$P
)
min(Todas$P)
max(Todas$P)
sd(Todas$P)
P_ppp <- as.ppp(P, W = janela )
pcolfun <- spatstat.options("image.colfun")
pcolmap <- colourmap(pcolfun(100), range = c(1, 3))
#n numero de cores no mapa
psymmap <- symbolmap(cols=pcolmap, size = .8, pch = 19, range = c(1, 3))
plot(P_ppp, symap = psymmap, main = "P at locations", legend = FALSE)
#=====Agora suavizamos os dados de P e os plotamos
psigma <- bw.ppl(P_ppp)
psmooth <- Smooth(P_ppp, sigma = psigma)
plot(psmooth, cols = pcolmap , main = "Smoothed p")
contour(psmooth, add = TRUE)
#=====Ca
Ca<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  Ca= Todas$Ca
)

```

```

min(Todas$Ca)
max(Todas$Ca)
sd(Todas$Ca)
Ca_ppp <- as.ppp(Ca, W = janela )
Cacolfun <- spatstat.options("image.colfun")
Cacolmap <- colourmap(Cacolfun(100), range = c(0.2, 0.7))
#n n mero de cores no mapa
Casymmap <- symbolmap(cols=Cacolmap,size = .8,pch = 19,range =c(0.2,0.7))
plot(Ca_ppp, symap = Casymmap,main = "Ca at locations",legend = FALSE)
#Agora suavizamos os dados de Ca e os plotamos
Casigma <- bw.ppl(Ca_ppp)
Casmooth <- Smooth(Ca_ppp, sigma = Casigma)
plot(Casmooth, cols = Cacolmap , main = "Smoothed Ca")
contour(Casmooth, add = TRUE)
#===== Mg
Mg<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  Mg= Todas$Mg
)
min(Todas$Mg)
max(Todas$Mg)
sd(Todas$Mg)
Mg_ppp <- as.ppp(Mg, W = janela )
Mgcolfun <- spatstat.options("image.colfun")
Mgcolmap <- colourmap(Mgcolfun(100), range = c(0.1,0.2))
#n n mero de cores no mapa
Mgsymmap <- symbolmap(cols=Mgcolmap,size =.8,pch=19,range =c(0.1,0.2))
plot(Mg_ppp, symap = Mgsymmap, main = "Mg at locations", legend = FALSE)
#Agora suavizamos os dados de Mg e os plotamos :
Mgsigma <- bw.ppl(Mg_ppp)
Mgsmooth <- Smooth(Mg_ppp, sigma = Mgsigma)
plot(Mgsmooth, cols = Mgcolmap,main = "Smoothed Mg")
contour(Mgsmooth, add = TRUE)

```

```

#===== Al
Al<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  Al= Todas$Al
)
min(Todas$Al)
max(Todas$Al)
sd(Todas$Al)
Al_ppp <- as.ppp(Al, W = janela )
Alcolfun <- spatstat.options("image.colfun")
Alcolmap <- colourmap(Alcolfun(100), range = c(1.1, 2.5))
#n nmero de cores no mapa
Alsymmap <- symbolmap(cols=Alcolmap,size =.8,pch =19,range=c(1.1,2.5))
plot(Al_ppp, symap=Alsymmap,main="Al at locations",legend = FALSE)
#Agora suavizamos os dados de Al e os plotamos:
Alsigma <- bw.ppl(Al_ppp)
Alsmooth <- Smooth(Al_ppp, sigma = Alsigma)
plot(Alsmooth, cols = Alcolmap , main = "Smoothed Al")
contour(Alsmooth, add = TRUE)
#=====H+Al
H.Al<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  H.Al= Todas$`H+Al`
)
min(Todas$`H+Al`)
max(Todas$`H+Al`)
sd(Todas$`H+Al`)
H.Al_ppp <- as.ppp(H.Al, W = janela )
H.Alcolfun <- spatstat.options("image.colfun")
H.Alcolmap <- colourmap(H.Alcolfun(100), range = c(8.1, 14.9))
#n nmero de cores no mapa
H.Alsymmap<-symbolmap(cols=H.Alcolmap,size =.8,pch =19,range=c(8.1,14.9))

```

```

plot(H.Al_ppp, symap=H.Alsymmap, main="H.Al at locations", legend = FALSE)
#Agora suavizamos os dados de H+Al e os plotamos:
H.Alsigma <- bw.ppl(H.Al_ppp)
H.Alsmooth <- Smooth(H.Al_ppp, sigma = H.Alsigma)
plot(H.Alsmooth, cols = H.Alcolmap , main = "Smoothed H.Al")
contour(H.Alsmooth, add = TRUE)
#=====B
SB<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  SB= Todas$SB
)
min(Todas$SB)
max(Todas$SB)
sd(Todas$SB)
SB_ppp <- as.ppp(SB, W = janela )
SBcolfun <- spatstat.options("image.colfun")
SBcolmap <- colourmap(SBcolfun(100), range=c(0.3,1))
#n n mero de cores no mapa
SBsymmap <- symbolmap(cols=SBcolmap, size =.8, pch=19, range = c(0.3, 1))
plot(SB_ppp, symap = SBsymmap, main = "SB at locations", legend = FALSE)
#Agora suavizamos os dados de SB e os plotamos:
SBsigma <- bw.ppl(SB_ppp)
SBsmooth <- Smooth(SB_ppp, sigma = SBsigma)
plot(SBsmooth, cols = SBcolmap , main = "Smoothed SB")
contour(H.Alsmooth, add = TRUE)
#===== t
t<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  t= Todas$t
)
min(Todas$t)
max(Todas$t)

```

```

sd(Todas$t)
t_ppp <- as.ppp(t, W = janela )
tcolfun <- spatstat.options("image.colfun")
tcolmap <- colourmap(tcolfun(100), range = c(2, 3.3))
#n n mero de cores no mapa
tsymmap <- symbolmap(cols=tcolmap,size=.8, pch=19,range = c(2,3.3))
plot(t_ppp, symap = tsymmap,main = "t at locations",legend=FALSE)
#Agora suavizamos os dados de t e os plotamos:
tsigma <- bw.ppl(t_ppp)
tsmooth <- Smooth(t_ppp, sigma = tsigma)
plot(tsmooth, cols = tcolmap , main = "Smoothed t")
contour(tsmooth, add = TRUE)
#=====T
T<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  T= Todas$T
)
min(Todas$T)
max(Todas$T)
sd(Todas$T)
T_ppp <- as.ppp(T, W = janela )
Tcolfun <- spatstat.options("image.colfun")
Tcolmap <-colourmap(Tcolfun(100),range = c(9.1,15.7))
#n numero de cores no mapa
Tsymmap <- symbolmap(cols=Tcolmap,size =.8,pch=19,range = c(9.1,15.7))
plot(T_ppp, symap = Tsymmap, main = "T at locations",legend = FALSE)
#Agora suavizamos os dados de T e os plotamos:
Tsigma <- bw.ppl(T_ppp)
Tsmooth <- Smooth(T_ppp, sigma = Tsigma)
plot(Tsmooth, cols = Tcolmap , main = "Smoothed T")
contour(Tsmooth, add = TRUE)
#=====V
V<- data.frame(

```

```

X=Todas$X,
Y=Todas$Y,
V= Todas$V)
min(Todas$V)
max(Todas$V)
sd(Todas$V)
V_ppp <- as.ppp(V, W = janela )
Vcolfun <- spatstat.options("image.colfun")
Vcolmap <- colourmap(Vcolfun(100), range = c(2.4, 11))
#n n mero de cores no mapa
Vsymmap <- symbolmap(cols=Vcolmap,size=.8,pch=19,range=c(2.4, 11))
plot(V_ppp, symap=Vsymmap,main = "V at locations",legend=FALSE)
#Agora suavizamos os dados de V e os plotamos
Vsigma <- bw.ppl(V_ppp)
Vsmooth <- Smooth(V_ppp, sigma = Vsigma)
plot(Vsmooth, cols = Vcolmap , main = "Smoothed V")
contour(Vsmooth, add = TRUE)
#===== m
X=Todas$X,
Y=Todas$Y,
V= Todas$m)
min(Todas$m)
max(Todas$m)
sd(Todas$m)
m_ppp <- as.ppp(m, W = janela )
mcolfun <- spatstat.options("image.colfun")
mcolmap <- colourmap(mcolfun(100), range = c(52.4, 87))
#n n mero de cores no mapa
msymmap <- symbolmap(cols=mcolmap,size=.8,pch=19,range=c(52.4, 87))
plot(m_ppp, symap = msymmap, main = "m at locations",legend=FALSE)
#Agora suavizamos os dados de m e os plotamos:
msigma <- bw.ppl(m_ppp)
msmooth <- Smooth(m_ppp, sigma = msigma)
plot(msmooth, cols = mcolmap , main = "Smoothed m")

```

```

contour(msmooth, add = TRUE)
#=====OM
MO<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  MO= Todas$MO
)
min(Todas$MO)
max(Todas$MO)
sd(Todas$MO)
MO_ppp <- as.ppp(MO, W = janela )
MOcolfun <- spatstat.options("image.colfun")
MOcolmap <- colourmap(MOcolfun(100), range = c(4.3, 7.6))
#n numero de cores no mapa
MOsymmap <- symbolmap(cols=MOcolmap, size =.8, pch=19, range=c(4.3, 7.6))
plot(MO_ppp, symap=MOsymmap, main="MO at locations", legend=FALSE)
#Agora suavizamos os dados de OM e os plotamos:
MOsigma <- bw.ppl(MO_ppp)
MOsmooth <- Smooth(MO_ppp, sigma = MOsigma)
plot(MOsmooth, cols = MOcolmap , main = "Smoothed MO")
contour(MOsmooth, add = TRUE)
#=====Argila
Argila<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  Argila= Todas$Argila
)
min(Todas$Argila)
max(Todas$Argila)
sd(Todas$Argila)
Argila_ppp <- as.ppp(Argila, W = janela )
Argilacolfun <- spatstat.options("image.colfun")
Argilacolmap <- colourmap(Argilacolfun(100), range = c(35, 74))
#n numero de cores no mapa

```

```

Argilasymmap <- symbolmap(cols=Argilacolmap, size = .8, pch = 19,
range = c(35, 74))
plot(Argila_ppp, symap = Argilasymmap, main = "Argila at locations",
legend = FALSE)
#==Agora suavizamos os dados de Argila e os plotamos:
Argilasigma <- bw.ppl(Argila_ppp)
Argilasmooth <- Smooth(Argila_ppp, sigma = Argilasigma)
plot(Argilasmooth, cols = Argilacolmap , main = "Smoothed Argila")
contour(Argilasmooth, add = TRUE)
#=====Silte
Silte<- data.frame(
  X=Todas$X,
  Y=Todas$Y,
  Silte= Todas$Silte
)
min(Todas$Silte)
max(Todas$Silte)
sd(Todas$Silte)
Silte_ppp <- as.ppp(Silte, W = janela )
Siltecolfun <- spatstat.options("image.colfun")
Siltecolmap <- colourmap(Siltecolfun(100), range = c(2, 43))
#n numero de cores no mapa
Siltesymmap <- symbolmap(cols=Siltecolmap, size = .8, pch = 19,
range = c(2, 43))
plot(Silte_ppp, symap = Siltesymmap, main = "Silte at locations",
legend = FALSE)
#===Agora suavizamos os dados de Silte e os plotamos :
Siltesigma <- bw.ppl(Silte_ppp)
Siltesmooth <- Smooth(Silte_ppp, sigma = Siltesigma)
plot(Siltesmooth, cols = Siltecolmap , main = "Smoothed Silte")
contour(Siltesmooth, add = TRUE)
#=====Areia
Areia<- data.frame(
  X=Todas$X,

```

```

Y=Todas$Y,
Areia= Todas$Areia
)
min(Todas$Areia)
max(Todas$Areia)
sd(Todas$Areia)
Areia_ppp <- as.ppp(Areia, W = janela )
Areiacolfun <- spatstat.options("image.colfun")
Areiacolmap <- colourmap(Areiacolfun(100), range = c(14, 41))
#n numero de cores no mapa
Areiasymmap <- symbolmap(cols=Areiacolmap, size = .8, pch = 19,
range = c(14, 41))
plot(Areia_ppp, symap = Areiasymmap, main = "Areia at locations",
legend = FALSE)
#====Agora suavizamos os dados de Areia e os plotamos
Areiasigma <- bw.ppl(Areia_ppp)
Areiasmooth <- Smooth(Areia_ppp, sigma = Areiasigma)
plot(Areiasmooth, cols = Areiacolmap , main = "Smoothed Areia")
contour(Areiasmooth, add = TRUE)
#====Modelos ajustados=====#
#==== Esp cie Copaifera langsdorffii====
fitCop1 <- ppm(Copaifera.ppp~pHsmooth+ksmooth+psmooth+Casmooth+
Mgsmooth+Alsmooth+H.Alsmooth+SBsmooth+
          tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+Argilasmooth+
          Siltesmooth+Areiasmooth)
fitCop1_step <- step(fitCop1)
summary(fitCop1_step)
fitCopval <- ppm(Copaifera.ppp~psmooth + Alsmooth + H.Alsmooth +
tsmooth + Vsmooth + MOsmooth + Argilasmooth)
summary(fitCopval)
par(mfrow=c(1,1), mar = c(0,0,.1,.1), oma=c(0.0001, 0.0001, 0.0001, 0.0001))
diagnose.ppm(fitCopval, envelope=TRUE, nsim = 1000)
par(mfrow=c(1,1), mar=c(4, 4.5, 0.2, 0.4))
qqplot.ppm(fitCopval , correction="best", nsim = 1000)

```

```

#plots and simulation
par(mfrow=c(1,2), mar = c(0,0,1,1), oma=c(0.0001, 0.0001, 0.0001, 0.0001))
plot(simulate(fitCopval))
plot(Copaifera.ppp)
#method
fitCopval$method
AIC(fitCopval)
#kfunction
par(mfrow=c(1,1), mar=c(4, 4.5, 0.2, 0.4))
E1Cop=envelope(fitCopval, fun = "Kinhom", correction="best", nsim=1000)
plot(E1Cop, main='', xlab=expression(r~("m")), legend=F)
lines(E1Cop$r, E1Cop$obs, col = "black", lty="solid", lwd=1.5)
lines(E1Cop$r, E1Cop$mmean, lty="solid", col = "blue", lwd=1.5)
lines(E1Cop$r, E1Cop$lo, lty="dashed", col = "pink", lwd=2) #Limite inferior
lines(E1Cop$r, E1Cop$hi, lty="dashed", col = "pink", lwd=2) # Limite superior
legend("topleft", legend = c("Modelo ajustado", "Poisson nao homogeneo",
"Limite inferior do envelope", "Limite Superior do envelope"),
      col = c("black", "blue", "pink", "pink"), lwd = 2, bty = "n",
      lty = c("solid", "solid", "dashed", "dashed"))

#===== Xylopia brasiliensis =====
fitxylop1 <- kppm(Xylopia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
      Alsmooth+H.Alsmooth+SBsmooth+
      tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
      Argilasmooth+Siltesmooth+Areiasmooth, "Thomas", method='p')
fitxylop1_step <- step(fitxylop1)
summary(fitxylop1_step)
fitxylopThomasvali <- kppm(Xylopia.ppp~pHsmooth + ksmooth+
      Alsmooth + H.Alsmooth +
      Tsmooth+ MOsmooth, "Thomas", method='p')
#===matclust
fitxylop2 <- kppm(Xylopia.ppp~pHsmooth+ksmooth+psmooth+
      Casmooth+Mgsmooth+Alsmooth+H.Alsmooth+SBsmooth+
      tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+

```

```

      Argilasmooth+Siltesmooth+Areiasmooth, "MatClust", method='p')
fitxylop2_step <- step(fitxylop2)
fitxylopmatclustvali <- kppm(Xylophia.ppp~pHsmooth+ksmooth+
  Alsmooth+H.Alsmooth+Tsmooth +MOsmooth, "MatClust", method='p')
summary(fitxylopmatclustvali)
min(fitxylopmatclustvali$mu)
max(fitxylopmatclustvali$mu)
#===Cauchy
fitxylop3 <- kppm(Xylophia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
  Alsmooth+H.Alsmooth+SBsmooth+
  tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
  Argilasmooth+Siltesmooth+Areiasmooth, "Cauchy", method='p')
fitxylop3_step <- step(fitxylop3)
summary(fitxylop3_step)
fitxylopCauchyvali <- kppm(Xylophia.ppp~pHsmooth+ksmooth+
  Alsmooth+H.Alsmooth+Tsmooth +MOsmooth, "Cauchy", method='p')
#===VarGamma
fitxylop4 <- kppm(Xylophia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
  Alsmooth+H.Alsmooth+SBsmooth+tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
  Argilasmooth+Siltesmooth+Areiasmooth, "VarGamma", method='p')
fitxylop4_step <- step(fitxylop4)
summary(fitxylop4_step)
fitxylopVarGammavalivali <- kppm(Xylophia.ppp~pHsmooth+ksmooth+
  Alsmooth+H.Alsmooth+Tsmooth +MOsmooth, "VarGamma", method='p')
#===LGCP
fitxylop5 <- kppm(Xylophia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
  Alsmooth+H.Alsmooth+SBsmooth+tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
  Argilasmooth+Siltesmooth+Areiasmooth, "LGCP", method='p')
fitxylop5_step <- step(fitxylop5)
summary(fitxylop5_step)
fitxylopLGCPvali <- kppm(Xylophia.ppp~pHsmooth+ksmooth+
  Alsmooth+H.Alsmooth+Tsmooth +MOsmooth, "LGCP", method='p')
summary(fitxylopLGCPvali)
#===== Envelopes

```

```

AIC(fitxylopThomasvali)
AIC(fitxylopmatclustvali)
AIC(fitxylopCauchyvali)
AIC(fitxylopVarGammavali)
AIC(fitxylopLGCPvali)
plot(fitxylop1$mu)
E1xy=envelope(fitxylopThomasvali,Kinhom, nsim=1000,correction="best")
E2xy=envelope(fitxylopmatclustvali,Kinhom, nsim=1000,correction="best")
E3xy=envelope(fitxylopCauchyvali,Kinhom, nsim=1000,correction="best")
E4xy=envelope(fitxylopVarGammavali,Kinhom, nsim=1000,correction="best")
E5xy=envelope(fitxylopLGCPvali,Kinhom, nsim=1000,correction="best")
#===== Gr fico dos modelos ajustados=====#
par(mfrow=c(1,1),mai=c(0.8, 0.8, 0.3, 0.2))
plot(E1xy$r, E1xy$obs, type="n", ylab=expression(K[inhom]~(r)),
xlab=expression(r~("m")),main="")
polygon(c(rev(E1xy$r),E1xy$r),c(rev(E1xy$lo),E1xy$hi),col="gray",border=NA) #
lines(E1xy$r, E1xy$obs, col = "black",lty="solid", lwd=1.5)
# Tra ando o modelo para dados observada
lines(E1xy$r, E1xy$mmean, lty="solid", col = "red", lwd=1.5)
# Tra ando o modelo para fit
lines(E2xy$r, E2xy$mmean, lty="solid", col="blue", lwd=1.5)
# Tra ando o modelo para fit
lines(E3xy$r, E3xy$mmean, lty="solid", col="orange", lwd=1.5)
# Tra ando o modelo para fit
lines(E4xy$r, E4xy$mmean, lty="solid", col="yellow", lwd=1.5)
# Tra ando o modelo para fit
lines(E5xy$r, E5xy$mmean, lty="solid", col="violet", lwd=1.5)
# Tra ando o modelo para fit
#===== Tra ando as envelopes inferior e superior =====#
lines(E1xy$r, E1xy$lo, lty="dashed", col = "pink", lwd=2) #Limite inferior
lines(E1xy$r, E1xy$hi, lty="dashed", col = "pink", lwd=2) # Limite superior
#===== Adicionando a legenda
legend("topleft", legend = c("Modelo ajustado","Modelo Thomas",
"Modelo Matern","Modelo Cauchy ",

```

```

      "Modelo Variância Gamma ", "Modelo Log-Gaussian Cox ",
      "Limite inferior do envelope", "Limite superior do envelope"),
col = c("black", "red", "blue", "orange", "yellow", "violet",
"pink", "pink"), lwd = 2, bty = "n",
lty = c("solid", "solid", "solid", "solid", "solid",
"solid", "dashed", "dashed"))
#====Myrcia splendens=====#
fitMyr1 <- kppm(Myrcia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
      Alsmooth+H.Alsmooth+SBsmooth+
      tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
      Argilasmooth+Siltesmooth+Areiasmooth, "Thomas", method='p')
fitMyr1_step <- step(fitMyr1)
summary(fitMyr1_step)
fitMyr1Thomasvali <- kppm(Myrcia.ppp~ksmooth+Casmooth+Mgsmooth+
      Alsmooth+tsmooth +MOsmooth+
      Argilasmooth+Siltesmooth+Areiasmooth, "Thomas", method='p')
#===matclust
fitMyr2 <- kppm(Myrcia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
      Alsmooth+H.Alsmooth+SBsmooth+
      tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
      Argilasmooth+Siltesmooth+Areiasmooth, "MatClust", method='p')
fitMyr2_step <- step(fitMyr2)
summary(fitMyr2_step)
fitMyr2matclustvali <- kppm(Myrcia.ppp~ksmooth+Casmooth+Mgsmooth+
      Alsmooth+tsmooth +MOsmooth+
      Argilasmooth+Siltesmooth+Areiasmooth, "MatClust", method='p')
#===Cauchy
fitMyr3 <- kppm(Myrcia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
      Alsmooth+H.Alsmooth+SBsmooth+
      tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+
      Argilasmooth+Siltesmooth+Areiasmooth, "Cauchy", method='p')
fitMyr3_step <- step(fitMyr3)
summary(fitMyr3_step)
fitMyr3Cauchyvali <- kppm(Myrcia.ppp~ksmooth+Casmooth+Mgsmooth+

```

```

                                Alsmooth+tsmooth +M0smooth+
                                Argilasmooth+Siltesmooth+Areiasmooth, "MatClust", method='p')
#===VarGamma
fitMyr4 <- kppm(Myrcia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
                Alsmooth+H.Alsmooth+SBsmooth+
                tsmooth+Tsmooth +Vsmooth+msmooth+M0smooth+
                Argilasmooth+Siltesmooth+Areiasmooth, "VarGamma", method='p')
fitMyr4_step <- step(fitMyr4)
summary(fitMyr4_step)
fitMyrVarGammavali <- kppm(Myrcia.ppp~ksmooth+Casmooth+Mgsmooth+
                             Alsmooth+tsmooth +M0smooth+
                             Argilasmooth+Siltesmooth+Areiasmooth, "MatClust", method='p')
#===LGCP
fitMyr5 <- kppm(Myrcia.ppp~pHsmooth+ksmooth+psmooth+Casmooth+Mgsmooth+
                Alsmooth+H.Alsmooth+SBsmooth+tsmooth+Tsmooth +Vsmooth+msmooth+M0smooth+
                Argilasmooth+Siltesmooth+Areiasmooth, "LGCP", method='p')
fitMyr5_step <- step(fitMyr5)
fitMyr5_step
fitMyrLGCPvali <- kppm(Myrcia.ppp~ksmooth+Casmooth+Mgsmooth+
                       Alsmooth+tsmooth +M0smooth+
                       Argilasmooth+Siltesmooth+Areiasmooth, "LGCP", method='p')
summary(fitMyrLGCPvali)
min(fitMyrLGCPvali$mu)
max(fitMyrLGCPvali$mu)
AIC(fitMyr1Thomasvali)
AIC(fitMyr2matclustvali)
AIC(fitMyr3Cauchyvali)
AIC(fitMyrVarGammavali)
AIC(fitMyrLGCPvali)
#===== Envelopes
E1Myr=envelope(fitMyr1, Kinhom, nsim=100, correction="best")
E2Myr=envelope(fitMyr2, Kinhom, nsim=100, correction="best")
E3Myr=envelope(fitMyr3, Kinhom, nsim=100, correction="best")
E4Myr=envelope(fitMyr4, Kinhom, nsim=100, correction="best")

```

```

E5Myr=envelope(fitMyr5,Kinhom,nsim=100,correction="best")
#===== Gr fico dos modelos ajustados=====#
par(mfrow=c(1,1),mai=c(0.8, 0.8, 0.3, 0.2))
plot(E1Myr$r, E1Myr$obs, type="n", ylab=expression(K[inhom]~(r)),
xlab=expression(r~("m")),main="")
polygon(c(rev(E1Myr$r),E1Myr$r),c(rev(E1Myr$lo),E1Myr$hi),
col="gray",border=NA)
lines(E1Myr$r, E1Myr$obs, col="black",lty="solid",lwd=1.5)
# Tra ando o modelo para dados observada
lines(E1Myr$r, E1Myr$mmean, lty="solid", col="red",lwd=1.5)
# Tra ando o modelo para fit
lines(E2Myr$r, E2Myr$mmean, lty="solid", col="blue",lwd=1.5)
# Tra ando o modelo para fit
lines(E3Myr$r, E3Myr$mmean, lty="solid", col="orange",lwd=1.5)
# Tra ando o modelo para fit
lines(E4Myr$r, E4Myr$mmean, lty="solid", col="yellow",lwd=1.5)
# Tra ando o modelo para fit
lines(E5Myr$r, E5Myr$mmean, lty="solid", col="violet",lwd=1.5)
# Tra ando o modelo para fit
#===== Tra ando as envelopes inferior e superior =====#
lines(E1Myr$r, E1Myr$lo, lty="dashed",col="pink",lwd=2)#Limite inferior
lines(E1Myr$r, E1Myr$hi, lty="dashed",col="pink",lwd=2)# Limite superior
#=Adicionando a legenda
legend("topleft",legend=c("Modelo ajustado","Modelo Thomas ",
"Modelo Matern ","Modelo Cauchy",
"Modelo Vari ncia Gamma ","Modelo Log-Gaussian Cox",
"Limite inferior do envelope", "Limite superior do envelope"),
col=c("black","red", "blue","orange","yellow","violet","pink",
"pink"),lwd=2,bty="n",
lty=c("solid","solid", "solid","solid","solid", "solid",
"dashed", "dashed"))
#===== Siparuna guianensis
#===== Modelo Poisson n o homogeneo
par(mfrow=c(1,1),mar=c(0,0,0,0),oma=c(0.0001, 0.0001, 0.0001, 0.0001))

```

```

fitstr1 <- ppm(Siparuna.ppp~pHsmooth+ksmooth+psmooth+Casmooth+
Mgsmooth+Alsmooth+H.Alsmooth+SBsmooth+
tsmooth+Tsmooth +Vsmooth+msmooth+MOsmooth+Argilasmooth+
Siltesmooth+Areiasmooth)
fitstr1_step <- step(fitstr1)
summary(fitstr1_step)
fitstrval <-ppm(Siparuna.ppp~Alsmooth+tsmooth +Argilasmooth+ Siltesmooth+
Areiasmooth)
summary(fitstrval)
par(mfrow=c(1,1), mar=c(0.2,0.2,0.2, 0.2))
diagnose.ppm(fitstrval, envelope=TRUE, nsim = 1000)
par(mfrow=c(1,1), mar=c(4, 4.5, 0.2, 0.4))
qqplot.ppm(fitstrval, correction="best", nsim = 1000)
#qqplot.ppm(fitstr1 , type = 'eem', correction="best", diggle=T)
AIC(fitstrval)
#plots and simulation
plot(fitstrval)
par(mfrow=c(1,2), mar = c(0,0,1,1), oma=c(0.0001, 0.0001, 0.0001, 0.0001))
plot(simulate(fitstrval))
plot(Siparuna.ppp)
par(mfrow=c(1,1), mar=c(4, 4.5, 0.2, 0.4))
ElCo=envelope(fitstrval, fun = "Kinhom", correction="best")
plot(ElCo, main='', legend=F)
par(mfrow=c(1,1), mai=c(0.8, 0.8, 0.3, 0.2))
plot(ElCo$r, ElCo$obs, type="n", ylab=expression(K[inhom]~(r)),
      xlab=expression(r~("m")), main="")
      # Criando a rea de plotagem sem tra ar nada
polygon(c(rev(ElCo$r), ElCo$r), c(rev(ElCo$lo), ElCo$hi), col = "gray", border =NA)
lines(ElCo$r, ElCo$obs, col = "black", lty="solid", lwd=1.5)
lines(ElCo$r, ElCo$mmean, lty="solid", col = "red", lwd=1.5)
lines(ElCo$r, ElCo$lo, lty="dashed", col = "pink", lwd=2) #Limite inferior
lines(ElCo$r, ElCo$hi, lty="dashed", col = "pink", lwd=2) # Limite superior
legend("topleft", legend = c("Modelo ajustado", "Poisson n o homog neo",
"Limite inferior do envelope", "Limite Superior do envelope"),

```

```

    col = c("black", "red", "pink", "pink"), lwd = 2, bty = "n",
    lty = c("solid", "solid", "dashed", "dashed"))
#=====Mapa do Brasil destacando Minas gerais=====
# Carregar pacotes necessarios
library(sf); library(ggplot2); library(geobr); library(ggspatial)
library(grid); library(leaflet)

# Carregar dados do Brasil (por estados)
brasil <- read_state(
  code_state = "all",
  year = 2020,
  simplified = TRUE,
  showProgress = TRUE
)
# Calcular centroide dos estados e adicionar abreviaturas
brasil_centroids <- brasil %>%
  st_centroid() %>%
  st_coordinates() %>%
  as.data.frame() %>%
  rename(longitude = X, latitude = Y) %>%
  bind_cols(brasil) %>%
  mutate(sigla = as.character(code_state)) # Adicionar codigo do estado
  como 'sigla'
# Carregar dados de Minas Gerais (por municipios)
minas_gerais <- read_municipality(code_muni = "MG", year = 2020,
  simplified = TRUE)
# Destacar Lavras em Minas Gerais
lavras <- minas_gerais %>%
  filter(name_muni == "Lavras")
# Ponto da localiza o de Lavras (coordenadas aproximadas)
lavras_coords <- data.frame(
  longitude = -44.9965,
  latitude = -21.248,
  label = "Lavras"

```

```

)
# Ponto da rea de estudo em Lavras (coordenadas corrigidas)
estudo_coords <- data.frame(
  longitude = -44.971138,
  latitude = -21.228722,
  label = "Matinha"
)
# Criar mapa do Brasil com Lavras e siglas dos estados
brasil_map <- brasil %>%
  ggplot() +
  geom_sf(fill = "white", color = "black") +
  geom_sf(data = filter(brasil, code_state == "MG"),
    fill = "grey90", color = "grey50") + # Destacar MG
  geom_point(data = lavras_coords, aes(x = longitude,
    y = latitude),
    color = "red", size = 2) +
  geom_text_repel(data = lavras_coords, aes(x = longitude,
    y = latitude, label = label),
    nudge_y = 0.5, size = 3, fontface = "bold",
    color = "red") +
  geom_text(data = brasil_centroids, aes(x = longitude,
    y = latitude, label = sigla),
    size = 2.5, color = "black",
    nudge_y = ifelse(brasil_centroids$longitude < -55, -1, 1),
    fontface = "bold") + # Siglas em negrito
  geom_rect(data = lavras_coords,
    aes(xmin = longitude - 0.5, xmax = longitude + 0.5,
    ymin = latitude - 0.5, ymax = latitude + 0.5),
    color = "blue", fill = NA, linewidth = 0.8) +
  labs(x = NULL, y = NULL) +
  theme_minimal() +
  theme(axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()) +

```

```

annotation_scale(location = "bl", width_hint = 0.5) +
annotation_north_arrow(location = "tl",
which_north = "true",
  style = north_arrow_fancy_orienteering()) +
coord_sf(xlim = c(-75, -30), ylim = c(-35, 5))
# Ajuste para focar no Brasil

# Criar o mapa detalhado de Lavras usando leaflet
lavras_map <- leaflet() %>%
  addTiles() %>%
  addPolygons(data = minas_gerais, fillColor = "grey90",
color = "blue") %>%
  addMarkers(lng = lavras_coords$longitude,
lat = lavras_coords$latitude, popup = "Lavras",
  label = "Lavras") %>%
  addMarkers(lng = estudo_coords$longitude,
lat = estudo_coords$latitude, popup = " rea de Estudo",
  label = " rea de Estudo") %>%
  addScaleBar(position = "bottomleft") %>%
  addControl(html = "<div style='background: white;
padding: 5px;'>Norte</div>", position = "topleft") %>%
  addMiniMap(
  tiles = providers$OpenStreetMap.Mapnik, # Camada de base para o mini-mapa
  toggleDisplay = TRUE,
  position = "bottomleft",
  zoomLevelOffset = -6, # Ajustar zoom para focar no Brasil
  aimingRectOptions = list(color = "red", weight = 1,
  fillOpacity = 0.02)
) %>%
  addPolygons(data = brasil, # Adiciona o mapa do Brasil ao mini-mapa
fillColor = ~ifelse(code_state == "MG", "grey", "white"),
color = "black", weight = 1, fillOpacity = 0.7,
  group = "miniMap") # Atribui o grupo "miniMap" para manipula o
# Plotar ambos os mapas

```

```

grid.newpage()
pushViewport(viewport(layout = grid.layout(2, 1)))
print(brasil_map, vp = viewport(layout.pos.row = 1, layout.pos.col = 1,
width = 0.8, height = 0.4))
print(lavras_map, vp = viewport(layout.pos.row = 2, layout.pos.col = 1,
width = 0.8, height = 0.6))

#=====Criando mapa da matinha=====
# Coordenadas do ponto de estudo
estudo_coords <- data.frame(
  longitude = -44.971138,
  latitude = -21.228722,
  label = "Matinha" # Palavra a ser exibida no mapa
)
# Criar o mapa principal com leaflet
map <- leaflet(estudo_coords) %>%
  addProviderTiles(providers$Esri.WorldImagery) %>% # Adiciona imagens
de sat lite detalhadas
  addMarkers(
    ~longitude, ~latitude,
    popup = ~label, # Adiciona um popup com o label
    label = ~label, # Adiciona uma label para o ponto
    labelOptions = labelOptions(
      noHide = TRUE,
      direction = "top",
      style = list("font-size" = "32px") # Aumenta o tamanho da fonte
para 32
    )
  ) %>%
  addCircleMarkers(
    ~longitude, ~latitude,
    radius = 8,
    color = "#007bff",

```

```

    fillColor = "#007bff",
    fillOpacity = 0.7,
    popup = ~label
  ) %>%
  setView(lng = -44.971138, lat = -21.228722, zoom = 17) # Centraliza
  e define o n vel de zoom
# Adicionar minimapa
map <- map %>%
  addMiniMap(
tiles = providers$OpenStreetMap.Mapnik, # Provedor para o minimapa
position = "bottomright",             # Posi o do minimapa
zoomLevelOffset = -4,                 # Define o nivel de zoom do
#minimapa
toggleDisplay = TRUE                   # Permite ocultar/exibir o
#minimapa
  )
# Exibir o mapa
map
#Nota: Os mapas da Figura 3.1 foram salvos em PNG e juntados no Canva.

```