**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Relações entre fatores ambientais e espécies florestais por
metodologias de processos pontuais**

Djair Durand Ramalho Frade

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2014**

Djair Durand Ramalho Frade
Bacharel em Estatística

**Relações entre fatores ambientais e espécies florestais por metodologias de
processos pontuais**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **PAULO JUSTINIANO RIBEIRO JUNIOR**

Dissertação apresentado para obtenção do título de Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2014**

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP

Frade, Djair Durand Ramalho

Relações entre fatores ambientais e espécies florestais por metodologias de processos pontuais / Djair Durand Ramalho Frade. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.

94 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.
Bibliografia.

1. Log-Gaussiano Cox 2. Função K de Ripley 3. Poisson não homogêneo 4. Estimador de Kernel I. Título

CDD 634.97
F799r

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

A minha mãe Darlene, a qual tenho muito respeito.

Aos meus irmãos Jordânia e Darley

À minha namorada Natália, por seu amor

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Paulo Justiniano Ribeiro Junior, pela orientação, dedicação, incentivo e paciência ao longo do desenvolvimento desta pesquisa;

A Professora Dr^a Silvia Emiko Shimakura, pelo apoio e valiosas sugestões;

A CAPES pela concessão da bolsa de mestrado para a realização da pesquisa e ao projeto temático “Diversidade, dinâmica e conservação em Florestas do Estado de São Paulo: 40,96 ha de parcelas permanentes” (Processo no. 99/09635-0) do programa BIOTA-FAPESP, que disponibilizou a base de dados utilizada neste trabalho;

Aos meus familiares;

Aos amigos Ana Júlia, José Nilton, Kuang, Maria Cristina, Pedro, Rafael, Ricardo, Thiago Gentil e Thiago Oliveira pela ajuda em momentos de dificuldade e, principalmente, por vossas amizades;

Estendo meus agradecimentos à Natália, minha namorada, companheira e incentivadora;

Aos professores do Departamento de Ciências Exatas - LCE;

A todos aqueles que indiretamente colaboraram para a conclusão desta dissertação;

A equipe do Biota Parcelas Permanentes.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
LISTA DE ABREVIATURAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Processos Pontuais Espaciais	24
2.2 Processo de Poisson	25
2.2.1 Processo Poisson Homogêneo - PPH	25
2.2.2 Processo Poisson Não-Homogêneo - PPNH	27
2.3 Processo Log Gaussiano de Cox	28
2.4 Pressupostos para análise de processos pontuais espaciais	28
2.4.1 Completa Aleatoriedade Espacial (CAE)	28
2.4.2 Estacionariedade	29
2.4.3 Isotropia	29
2.4.4 Propriedade de primeira ordem	30
2.4.5 Propriedade de segunda ordem	30
2.5 Análise de processos pontuais	31
2.5.1 Estimação da propriedade de primeira ordem	31
2.5.1.1 Estimação por kernel	31
2.5.2 Estimação da propriedade de segunda ordem	34
2.5.2.1 Função K para processos pontuais homogêneos	34
2.5.2.2 Função K para processos pontuais não-homogêneos	38
2.6 Estudos caso-controle	38
3 METODOLOGIA	41
3.1 Material	41
3.2 Método	43
4 RESULTADOS E DISCUSSÃO	49
4.1 Efeitos das covariáveis espaciais na falta de homogeneidade das três espécies mais abundantes na Estação Ecológica de Assis	52

4.1.1 Variáveis originais	52
4.1.2 Análise Fatorial	61
4.2 Distribuição espacial do risco	68
5 CONCLUSÕES	73
REFERÊNCIAS	75
APÊNDICE	81

RESUMO

Relações entre fatores ambientais e espécies florestais por metodologias de processos pontuais

O padrão espacial de espécies em florestas nativas pode fornecer evidências sobre a estrutura da comunidade vegetal. Fatores ambientais podem influenciar o padrão espacial das espécies, como as características edáficas e processos que dependem da densidade, como competição intra e interespecífica. Desse modo, a pesquisa da relação entre as características ambientais e o padrão espacial de espécies florestais pode ajudar a entender a dinâmica de florestas. O objetivo deste estudo foi empregar técnicas da análise de processos pontuais para verificar o efeito de fatores ambientais sobre a ocorrência de espécies florestais. A área de estudo foi a Estação Ecológica de Assis (EEA), da unidade de Conservação do Estado de São Paulo em parcelas permanentes, dentro do projeto “Diversidade, dinâmica e conservação em florestas do Estado de São Paulo: 40 *ha* de parcelas permanentes” do programa Biotá da FAPESP. A descrição do padrão espacial das espécies mais abundantes na área de estudo foi avaliada pela função K proposta por Ripley e suas extensões para processo não homogêneos, por meio das coordenadas geográficas das espécies com circunferência na altura do peito igual ou superior a 15 cm. Modelos do Processo Poisson Homogêneo, Processo Poisson Não Homogêneos e do Processo Log Gaussiano de Cox foram ajustados para cada espécie. Foi utilizado o critério de AIC para selecionar o modelo que melhor se ajusta aos dados. Testes de diagnósticos dos modelos foram feitos utilizando a função K não homogênea sob a hipótese de Completa Aleatoriedade Espacial. Os resultados indicaram que as espécies mais abundantes na EEA apresentam um padrão de distribuição agregado, ou seja, o número esperado de indivíduos próximos de um evento qualquer é maior do que esperado para uma distribuição aleatória. Conforme esperado, os fatores ambientais desempenharam um importante papel para explicar a distribuição espacial das espécies, porém, os resultados indicaram que existe uma variação espacialmente estruturada que não foi incluída na análise que é imprescindível para um bom ajuste dos modelos. Portanto os resultados sugerem que outros fatores não incluídos nos modelos e dados disponíveis podem estar determinando os padrões espaciais além das (co)variáveis medidas.

Palavras-chave: log-gaussiano Cox; Função K de Ripley; Poisson não homogêneo; Estimador de Kernel

ABSTRACT

Relationship between environmental factors and forest species using points process methodologies

The spatial pattern of species in native forests may provide evidence on the structure of the plant community. Environmental factors may influence the species' spatial patterns, as well as soil characteristics and processes which depend on the density, as intraspecific and interspecific competition. Therefore, researching the relationship among the environmental features and the spatial pattern of the forest species may aid in understanding forest dynamics. The goal of this study was to apply point process techniques to verify the effect of environmental factors on the occurrence of forest species. The study area was the "Assis's Ecological Station" (AES), of the "Unit of conservation of the state of São Paulo in permanent plots". The data was collected as part of the project entitled "Diversity, dynamics and conservation in forests of São Paulo state: 40 *ha* of permanent plots", from FAPESP's Biota program. The description of the spatial pattern of the most abundant species in the study area was assessed using Ripley's K function, using the species' geographic coordinates with circumference at chest height equal or larger than 15 cm. Homogeneous and Non-Homogeneous Poisson Process models, as well as Cox Log Gaussian Process models were fitted to each species. Model selection was made using the Akaike information criterion. Diagnostics tests were made using the non-homogeneous K function under the hypothesis of complete spatial randomness. Results suggested that the most abundant species in the AES present an aggregate distribution pattern, i.e., the expected number of individuals next to any event is larger than the expected by a random distribution. As it was expected, environmental factors played a major role in explaining the spatial distribution of the species. However, results suggested that there is a spatially structured variation that was not included in the analysis and is needed to a good model fit. Therefore, further studies are needed to assess which environmental feature which was not considered in this study presents an effect on the occurrence of these forest species

Keywords: Log-Gaussian Cox; Ripley K function; Inhomogeneous Poisson; Kernel estimate

LISTA DE FIGURAS

Figura 1 - Os tipos básicos de distribuição dos pontos amostrais. Da esquerda para a direita: aglomerado, regular e aleatório	24
Figura 2 - Uma realização do Processo Poisson Homogêneo com média 70	26
Figura 3 - Estimador de intensidade pontual do processo em toda região de estudo (kernel)	32
Figura 4 - Representação gráfica da função $\hat{L}(h)$	36
Figura 5 - Representação gráfica da função $\hat{L}(h)$ com envelope	37
Figura 6 - (a) Localização das quatro Unidades de Conservação no Estado de São Paulo e (b) Parcela permanente subdividida em 256 subparcelas contíguas de 20×20 ($400 m^2$)	41
Figura 7 - Fluxograma da abordagem geral do trabalho	48
Figura 8 - Distribuição espacial das três espécies mais abundantes na Estação Ecológica de Assis. (a) <i>Copaifera langsdorffii</i> ; (b) <i>Vochysia tucanorum</i> e (c) <i>Xylopia aromatica</i>	49
Figura 9 - Distribuição espacial das três espécies mais abundantes na Estação Ecológica de Assis por meio da suavização de Kernel, em que a escala (lado direito) representa a intensidade de pontos por unidade de área. (a) <i>Copaifera langsdorffii</i> ; (b) <i>Vochysia tucanorum</i> e (c) <i>Xylopia aromatica</i> . . .	50
Figura 10 - Estimativas da função L (linha cheia) e envelopes de simulação, contra a distribuição teórica sob hipótese de CAE (linha tracejada) referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) <i>Copaifera langsdorffii</i> ; (b) <i>Vochysia tucanorum</i> e (c) <i>Xylopia aromatica</i>	51
Figura 11 - Estimativas da função L não homogênea (linha cheia) e envelopes de simulação dos modelos com menor valor de AIC referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) <i>Copaifera langsdorffii</i> ; (b) <i>Vochysia tucanorum</i> e (c) <i>Xylopia aromatica</i>	59
Figura 12 - Estimativas da função L não homogênea (linha cheia) e envelopes de simulação dos modelos com menor valor de AIC considerando um componente espacial estimado por meio do LGCP referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) <i>Copaifera langsdorffii</i> ; (b) <i>Vochysia tucanorum</i> e (c) <i>Xylopia aromatica</i>	61

- Figura 13 - Estimativas da função L não homogênea (linha cheia) e envelopes de simulação utilizando os fatores extraídos das variáveis originais, referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica* . . . 67
- Figura 14 - Estimativas da função L não homogênea (linha cheia) e envelopes de simulação dos modelos com menor valor de AIC utilizando os fatores extraídos das variáveis originais considerando um componente espacial estimado por meio do LGCP referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica* 68
- Figura 15 - Mapa de risco para a propensão de presença das espécie *Copaifera langsdorffii*, controlando para fatores ambientais 70
- Figura 16 - Mapa de risco para a propensão de presença das espécie *Vochysia tucanorum*, controlando para fatores ambientais 70
- Figura 17 - Mapa de risco para a propensão de presença das espécie *Xylopia aromatica*, controlando para fatores ambientais 72

LISTA DE TABELAS

Tabela 1 - Diferentes tipos de função de estimação $k(h)$	33
Tabela 2 - Comparação entre os valores de $L(h)$ e $K(h)$	36
Tabela 3 - Espécies com maior densidade na Estação Ecológica de Assis. As espécies estudadas neste trabalho estão em negrito	42
Tabela 4 - Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie <i>Copaifera langsdorffii</i> e os coeficientes das variáveis preditoras dos modelos selecionados	53
Tabela 5 - Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie <i>Vochysia tucanorum</i> e os coeficientes das variáveis preditoras dos modelos selecionados	55
Tabela 6 - Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie <i>Copaifera langsdorffii</i> e os coeficientes das variáveis preditoras dos modelos selecionados	56
Tabela 7 - Parâmetros estimados a partir do LGCP aplicado aos dados de ocorrência de árvores na Estação Ecológica de Assis	60
Tabela 8 - Medidas de avaliação da Adequação da Análise Fatorial	62
Tabela 9 - Resultado da Análise Fatorial (AF)	63
Tabela 10 - Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie <i>Copaifera langsdorffii</i> considerando como covariáveis os fatores extraídos dos dados originais e os coeficientes das variáveis preditoras dos modelos selecionados	64

Tabela 11 -Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie <i>Vochysia tucanorum</i> considerando como covariáveis os fatores extraídos dos dados originais e os coeficientes das variáveis preditoras dos modelos selecionados	65
Tabela 12 -Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie <i>Xylopia aromatica</i> considerando como covariáveis os fatores extraídos dos dados originais e os coeficientes das variáveis preditoras dos modelos selecionados	66
Tabela 13 -Parâmetros estimados a partir do LGCP aplicado aos dados de ocorrência de árvores na Estação Ecológica de Assis, utilizando os fatores principais extraídos dos dados originais como covariáveis	67
Tabela 14 -Estimativas dos efeitos dos fatores ambientais para a <i>Copaifera langsdorffii</i>	69
Tabela 15 -Estimativas dos efeitos dos fatores ambientais para a <i>Vochysia tucanorum</i>	71
Tabela 16 -Estimativas dos efeitos dos fatores ambientais para a <i>Vochysia tucanorum</i>	72

LISTA DE ABREVIATURAS

ESALQ – Escola Superior de Agricultura “Luiz de Queiroz”

PEIC – Parque Estadual da Ilha do Cardoso

PECB – Parque Estadual de Carlos Botelho

EEA – Estação Ecológica de Assis

EEC – Estação Ecológica de Caetetus

GPS – Sistema de Posicionamento Global

S – Ponto de inflexão da curva de retenção de água

pH – potencial Hidrogeniônico

Al – Alumínio

Ca – Cálcio

Mg – Magnésio

K – Potássio

P – Fósforo

Na – Sódio

MO – teor de matéria orgânica

V – saturação por bases

m – saturação por alumínio

CTC – Capacidade de Troca Catiônica

ACP – Análise de Componentes Principais

AF – Análise Fatorial

CAE – Completa Aleatoriedade Espacial

CSR – Complete Spatial Randomness

KMO – medida de adequacidade da amostra de Kaiser-Meyer-Olkin

LGCP – Log Gaussian Cox Process

GAM – Generalized Additive Models

1 INTRODUÇÃO

Conhecer a maneira com que os indivíduos de uma espécie distribuem-se em uma comunidade é um dos primeiros passos para se entender a dinâmica de populações vegetais (HAY et al., 2000). Esse conhecimento é de suma importância para entender como determinada espécie utiliza os recursos disponíveis e de que forma estes recursos são importantes no sucesso do estabelecimento e reprodução dessa espécie (CONDIT et al., 2000).

Segundo Capretz et al. (2012), diversos fatores abióticos são frequentemente discutidos ao explicar a biodiversidade de florestas tropicais, e como consequência, o arranjo espacial das árvores. São citados principalmente a disponibilidade de habitats, nutrientes e água no solo, relevo e incidência de luz.

Nas últimas décadas, muita informação vem sendo acumulada sobre a composição e a estrutura das remanescentes florestais paulistas, com destaque para os estudos de florística e aspectos da estrutura da vegetação. No entanto, o conhecimento acumulado nestes estudos ainda é insuficiente para explicar a influência das características abióticas na distribuição geográfica das espécies arbóreas e, portanto, dos tipos vegetacionais, o que prejudica o avanço de hipóteses sobre a atuação conjunta dos fatores abióticos e dos processos bióticos na manutenção da diversidade nesses fragmentos (GANDOLFI; LEITÃO; BEZERRA, 1995; SALIS; SHEPHERD; JOLY, 1995; TORRES; MARTINS; KINOSHITA, 1997).

Investigar as relações existentes entre a distribuição das populações de espécies arbóreas e as variáveis ambientais que podem de forma direta ou indireta afetar tal distribuição é um dos principais objetivos do estudo da ecologia vegetal de florestas tropicais. Baseados nesta problemática, Berg e Santos (2003), verificaram quanto da diversidade de espécies arbóreas está associada à heterogeneidade ambiental e como as variáveis ambientais influenciam na distribuição das populações das espécies arbóreas. Para os autores, as propriedades físicas e químicas do solo e a topografia, em conjunto com os níveis de umidade e distribuição das chuvas, exercem importante papel na distribuição e composição de espécies.

Korman (2008) afirma que os fatores bióticos e abióticos agem muitas vezes de forma simultânea e fortemente correlacionadas, determinando a distribuição dos tipos

florestais e influenciando na distribuição das espécies dentro de uma mesma formação florestal. O conhecimento dos processos espaciais é fundamental para revelar as causas, modos e taxas de processos de dispersão das espécies. Pesquisas voltadas para as formações florestais podem servir para explicar de que modo uma grande quantidade de indivíduos de várias espécies estão aparentemente aptos a coexistirem num mesmo local (NEWBERY; RENSHAW; BRÜNIG, 1986) e como estas espécies distribuem-se espacialmente condicionadas às características do solo. Essas informações são relevantes uma vez que possibilitam melhorar manejos tradicionais, orientar pesquisadores na elaboração de modelos de manejo sustentável (LEVIN, 1992; MOEUR, 1993) e subsidiar estratégias de conservação de ecossistemas naturais (REIS, 1996).

A metodologia de processos pontuais é aplicada em diversas áreas do conhecimento como ecologia, climatologia, epidemiologia, dentre outras. Em ciências florestais, é fundamental para se revelar as causas, modos e taxas do processo de dispersão (REIS, 1995). Segundo Cunha (2003), com a análise de padrões espaciais é possível estimar o efeito da competição entre árvores na dinâmica da floresta, permitindo também conhecer melhor os processos espaciais relacionados com a reprodução, estabelecimento e crescimento das espécies a serem manejadas.

A análise de dados do processo pontual pode ser bastante útil para modelar o padrão de distribuição espacial das espécies de uma determinada área. Os processos pontuais e suas respectivas análises são utilizados em diversas áreas do conhecimento, inclusive na análise espacial para descrever o padrão das espécies na Estação ecológica de Assis que é alvo deste estudo.

Um das situações mais importantes na análise espacial de pontos é a possibilidade de comparação entre dois processos espaciais. Tipicamente, um dos processos representa os casos em estudo, e o outro os casos de um processo pontual que representa um conjunto de controle (CÂMARA et al., 2004). Diversos trabalhos que tem como objetivo estimar o risco espacial são encontrados na área de Epidemiologia ambiental (LAWSON; WILLIAMS, 1993; KELSALL; DIGGLE, 1995; SHIMAKURA et al., 2001)

Neste estudo objetiva-se investigar as relações entre os fatores ambientais e a distribuição das espécies arbóreas mais abundantes, bem como definir uma medida de risco para então investigar a significância da variação espacial e assim determinar e produzir mapas que identifiquem áreas de risco para as espécies em análises na Estação Ecológica

de Assis, que faz parte do projeto “Diversidade, dinâmica e conservação em florestas do Estado de São Paulo: 40 *ha* de parcelas permanentes”, por meio do uso dos dados originais e dos dados submetidos a análise multivariada pelo método da Análise Fatorial, utilizando tanto uma abordagem não-paramétrica quanto a análise estatística paramétrica de modo a modelar e descrever o padrão de ocorrência das espécies mais abundantes.

2 REVISÃO BIBLIOGRÁFICA

O objetivo da análise espacial é mensurar propriedades e relacionamentos, levando em consideração a localização espacial do fenômeno em estudo de forma explícita. Segundo Lopes (2005), pode-se definir a análise espacial como o estudo, exploração e modelagem de processo que se expressam através de uma distribuição no espaço.

Para Cressie (1993), a classificação mais utilizada para caracterizar problemas de análise espacial no que diz respeito à natureza estocástica da observação se divide em três grandes áreas:

- *Padrões de pontos* - Na análise de Padrões de Pontos lidamos com fenômenos que podem ser modelados segundo Processos Pontuais. Neste caso, o interesse do pesquisador reside na análise da intensidade de ocorrência de determinados eventos e na avaliação da existência de padrões de agregação ou repulsão espacial. A localizações de espécies, ocorrências de crimes e doenças são exemplos de fenômenos modelados via Processos Pontuais;
- *Geoestatística* - Diferentemente do eventos discretos que estão associados a ocorrências pontuais, os estudos sobre esses tipos de dados recorrem a amostras da superfície, isto é, valores representativos do fenômeno na área de estudo. Segundo Bailey e Gatrell (1995), os fenômenos naturais, tais como temperatura, pressão atmosférica, características de solo podem ser, em princípio, observados e medidos em qualquer parte da superfície terrestre;
- *Dados de Área* - No caso da análise de Dados de Área lidamos com fenômenos que, apesar de poderem se distribuir de forma contínua no espaço, são observados em sub-regiões que compõem uma partição da região de interesse. Neste caso, o pesquisador dispõe de medições individuais associadas à cada sub-região e possui o interesse em detectar padrões subjacentes ou quantificar o grau de correlação espacial na região. Dados agregados por bairro, municípios e Estados são exemplos de processos analisados via Dados de Área.

Pela estrutura dos dados utilizados nessa dissertação, as três abordagens podem ser utilizadas para verificar a associação entre os fatores ambientais e a distribuição

espacial das espécies. Um exemplo disto são os estudos de Grego¹ (em fase de elaboração) que contemplam duas das três abordagens: a abordagem geoestatística e a análise de dados de área. Esta dissertação no entanto, está concentrada apenas nas análises de processos pontuais.

2.1 Processos Pontuais Espaciais

Fenômenos expressos por meio de ocorrências identificadas como pontos localizados no espaço são denominados de *processos pontuais espaciais*. Um processo pontual espacial $Z(x)$, $x \in A$, em que A é um conjunto aleatório de índices em uma determinada região, cuja localização e número de realizações foi gerada por um mecanismo estocástico. Tal processo espacial difere-se da análise de Dados de Área e da Geoestatística pelo fato de que o comportamento estocástico primário é a própria localização espacial das observações. O conjunto das localizações espaciais observadas $Z = \{x_1, \dots, x_n\}$ é chamado de *arranjo pontual* e cada uma delas é usualmente chamada de *evento*, para distingui-las de pontos arbitrários no plano, denotados por x .

Basicamente são considerados três tipos básicos de distribuição dos pontos: aglomerado, regular e aleatório. A Figura 1 mostra uma realização simulada de cada um destes tipos.

Figura 1 – Os tipos básicos de distribuição dos pontos amostrais. Da esquerda para a direita: aglomerado, regular e aleatório

Quando os eventos estão concentrados no espaço, espera-se um padrão de pontos que mostre uma aglomeração (Figura 1). Esse padrão também é conhecido como **agrupado** ou **clusterizado**. O oposto direto do aglomerado é a distribuição regular, sendo

¹Grego, S. (Escola Superior de Agricultura “Luiz de Queiroz”). **Modelos não espaciais e espaciais para relacionar variáveis de solos e número de espécies em uma área de vegetação natural**

observada a existência de uma distância média entre os pontos que tende a ser constante. Os pontos estão espalhados no espaço, mas existe uma regularidade entre eles. No arranjo aleatório os eventos se distribuem no espaço de maneira completamente ao acaso.

Segundo Reis (2008), é possível que surjam distribuições pontuais heterogêneas como, por exemplo, da observação das posição de plantas onde a fertilidade do solo exibe uma variação espacial. Assumindo-se que a fertilidade é um campo aleatório que varia espacialmente, mas que, condicionada à fertilidade do solo (e possivelmente a outros fatores ambientais) as localizações das plantas são independentes, o modelo apropriado é um processo com intensidade (de primeira ordem) variando no espaço e com intensidade de segunda ordem nula. Por outro lado, se há dependência entre as plantas (competição), é natural pensar em um processo que tenha termos de interação.

2.2 Processo de Poisson

2.2.1 Processo Poisson Homogêneo - PPH

Autores como Baddeley, Møller e Waagepetersen (2000) e Illian et al. (2008), consideram o Processo de Poisson Homogêneo (ou estacionário) o modelo mais simples e importante dos processos pontuais. Por se tratar de um modelo com distribuição totalmente aleatória no espaço, sem qualquer interação, não há efeitos de primeira nem de segunda ordem. O PPH desempenha um papel importante como modelo nulo ou de referência que pode ser utilizado para distinguir se o padrão dos pontos exibe um comportamento aglomerado ou regular.

Um PPH é caracterizado pelo processo em que todos os eventos apresentam-se de forma independente e são uniformemente distribuídos na região onde ocorre o ponto. Isto significa que a localização do evento não afeta a probabilidade de outros eventos aparecerem nas proximidades, demonstrando ainda que não existem regiões onde os eventos são mais prováveis de ocorrerem (BIVAND; PEBESMA; GÓMES-RUBIO, 2008). Mais formalmente, Diggle (2003) descreve uma PPH em uma região A caracterizado por duas propriedades fundamentais:

PP1 - Para algum $\lambda > 0$, e qualquer região plana finita A , $N(A)$ (número de eventos em A), segue uma distribuição de Poisson com média $\lambda(|A|)$, sendo $|A|$ correspondente à área de A .

PP2 - Dado $N(A) = n$, os n eventos em A formam uma amostra aleatória independente com distribuição uniforme em A .

A constante λ representa a intensidade do processo, ou seja, o número esperado de pontos por unidade de área. Assim,

$$E[N(A)] = \lambda|A|$$

As propriedades PP1 e PP2 implicam em estacionariedade e isotropia. Caracteriza-se como estacionária uma vez que a intensidade é constante e a função intensidade de segunda ordem depende apenas das posições relativas dos dois pontos (isto é, direção e distância). Além disso, é isotrópica, porque a intensidade de segunda ordem é invariante em rotação. Assim, o processo não tem ponto de intensidade constante e a sua função intensidade de segunda ordem depende apenas da distância entre os dois pontos, independentemente das posições relativas dos pontos. Para simular uma realização de um processo de Poisson em A dado um valor de $N(A)$, é necessário gerar eventos independentemente e uniformemente distribuídos sobre A (DIGGLE, 2003). A Figura 2 apresenta uma amostra simulada de um processo Poisson homogêneo.

Figura 2 – Uma realização do Processo Poisson Homogêneo com média 70

Para um PPH com intensidade λ tem-se que $N(A) \sim \text{Poisson}(\lambda|A|)$, portanto, a função intensidade de primeira ordem e segunda ordem são respectivamente $\lambda(x) = \lambda$ e $\lambda_2(x_i, x_j) = \lambda(\mathbf{h}) = \lambda^2$.

O processo de Poisson homogêneo é útil como base de comparação mas pouco realístico para aplicações usuais. Ainda que não haja interação espacial entre os eventos, raramente se tem homogeneidade na intensidade. Assumindo eventos independentes, mas com a intensidade $\lambda(x)$ variando no espaço, um padrão espacial pode ser modelado através do *processo de Poisson não-homogêneo*.

2.2.2 Processo Poisson Não-Homogêneo - PPNH

Na maioria dos casos, assumir que o processo em estudo é homogêneo não é realista. Um exemplo é a localização das árvores em uma floresta em que diferentes fatores ambientais tais como a umidade, a qualidade do solo, a declividade, entre outros afetam a distribuição espacial.

Segundo Bivand, Pebesma e Gómez-Rubio (2008), o Processo Poisson Não-Homogêneo é uma generalização simples do Processo Poisson Homogêneo, no qual é introduzido heterogeneidade, o que permite uma intensidade não constante. Isto significa que a intensidade constante de um PPH é substituída por uma função de intensidade $\lambda(x)$, cujo valor varia de acordo com a localização x . Dessa forma, temos as seguintes propriedades fundamentais:

PPNH1 - $N(A)$ tem distribuição de Poisson com média $\int_A \lambda(x)dx$

PPNH2 - Dado $N(A) = n$, os n eventos em A formam uma amostra aleatória independente e identicamente distribuída, com função densidade de probabilidade $f(x) = \lambda(x)/I$, em que $I = \int_A \lambda(x)dx$.

No processo de Poisson não homogêneo o número de eventos em uma região $A \subset \mathbb{R}^2$ tem distribuição de Poisson com média $\mu(A) = \int_A \lambda(x)dx$ e, para regiões disjuntas, as contagens de eventos são independentes. Este é um processo não-estacionário, mas tem apenas efeitos de primeira ordem: aglomeração de eventos é resultante da heterogeneidade da intensidade, não da atração entre eventos. Também é possível modelar a intensidade por meio de covariáveis espaciais bem como covariáveis $Z(x)$ medida em cada localização x . O ajuste de um modelo de Poisson Não-Homogêneo para a função de intensidade em que intensidade seja uma função linear log da covariável é dado por

$$\hat{\lambda}(x) = \exp(\beta_0 + \beta_1 Z(x)) \quad (1)$$

2.3 Processo Log Gaussiano de Cox

Um processo de Cox $\mathbf{X}$ é uma classe de modelos que deriva do processo Poisson não homogêneo. (MØLLER; SYVERSVEEN; WAAGEPETERSEN, 1998) define o processo de Cox pelas seguintes propriedades:

- $\{\Lambda(x) : x \in A\}$ é um processo estocástico que assume apenas valores não negativos;
- Para $\mathbf{X}$ condicional apenas a uma realização de $\Lambda(x)$, $\{\Lambda(x) = \lambda(x) : x \in A\}$, os acontecimentos formam um processo de Poisson não homogêneo com função intensidade $\lambda(x)$.

Uma classe especialmente útil quando se pretende modelar a intensidade com o recurso de covariáveis é a do processo Log Gaussiano de Cox. Neste caso, modela-se o logaritmo da intensidade $\lambda(x)$ recorrendo a covariáveis e acrescentando-se um erro que é um processo gaussiano. O modelo é dado por

$$\log \lambda(x) = \boldsymbol{\beta}' \mathbf{d}(x) + \phi(x), \quad \text{com} \quad \phi(\cdot) \sim PG[0; \sigma^2; \rho(\cdot; \theta)],$$

em que, $\boldsymbol{\beta}$ é o vetor de coeficientes desconhecidos da regressão, σ^2 é a variância do processo espacial $\rho(h; \theta)$ é uma função de correlação espacial com parâmetro θ que depende apenas da distância h entre as localizações no espaço. Os efeitos espaciais $\phi(x)$ levam em conta a variação espacial não explicada pelas covariáveis e a incerteza na estimação da intensidade da população.

2.4 Pressupostos para análise de processos pontuais espaciais

2.4.1 Completa Aleatoriedade Espacial (CAE)

Segundo Diggle (2003), a análise estatística dos padrões de distribuição de pontos requer um modelo teórico de referência, base para o desenvolvimento de métodos formais que chegam à significância dos resultados exploratórios. O modelo teórico mais simples é conhecido como completa aleatoriedade espacial (ou *complete spatial randomness* - *CSR*), que corresponde a uma situação na qual é igualmente provável que um evento aconteça em qualquer local dentro da área de estudo, embora os locais sejam de outros eventos. No modelo *CAE*, considera-se que as ocorrências em cada sub-área sejam

não-correlacionadas, estando associadas à mesma distribuição de probabilidade de Poisson. Numa visão intuitiva, pode-se considerar que a posição dos eventos é independente e que os eventos têm igual probabilidade de ocorrência em toda a região A .

Câmara et al. (2004) afirma que é possível estabelecer uma base de comparação entre uma distribuição completamente aleatória (que seria gerada por um processo de Poisson) e os dados coletados em campo. O procedimento mais usual para estimar a probabilidade associada ao padrão encontrado é a produção de uma *simulação* do processo aleatório na região de estudo. Dado um número fixo de eventos medidos em campo (denotado por n), pode-se determinar um retângulo envolvente da região A . Os eventos são gerados a partir de abscissas x , obtidas de uma distribuição uniforme em (y_1, y_2) . Pontos que caem fora da região são rejeitados. Este processo é repetido até que n eventos tenham sido obtidos na região. Segundo Cressie (1993), rejeitada a hipótese de CAE a próxima etapa é ajustar algum modelo alternativo para os dados.

2.4.2 Estacionariedade

Para Illian et al. (2008) um processo é dito estacionário se Z (processo pontual) e Z_x (processo pontual transladado) possuem a mesma distribuição por meio de todas as translações de x . Ou seja,

$$Z \equiv Z_x$$

em que Z_x é o processo pontual resultante de uma translação de todos os pontos de Z pelo mesmo vetor $\mathbf{x}$: $Z = \{x_1, x_2, \dots\}$, então $Z_x = \{x_1 + x, x_2 + x, \dots\}$.

Considerando $Z(x)$ uma variável aleatória que representa o número de eventos em uma região plana A , então $Z(x) = \#(x_i \in A)$ em que $\#$ significa contagem.

Assim, o processo será estacionário se, para qualquer inteiro n e regiões A_i : $i = 1, \dots, n$, a distribuição conjunta de $Z(x_1), \dots, Z(x_n)$ for invariante sob translação para um ponto x .

2.4.3 Isotropia

Conforme Illian et al. (2008), o conceito de isotropia é análogo ao de estacionariedade, em que ao invés de translações por meio de vetores, as rotações são descritas por um ângulo α entre 0° e 360° , tal que, sendo $\mathbf{x} = (x, y)$, então, as coordenadas do ponto

transladado são dadas por

$$x_\alpha = x \cos(\alpha) + y \sin(\alpha) \quad \text{e} \quad y_\alpha = -x \sin(\alpha) + y \cos(\alpha).$$

Dessa forma, o processo será dito isotrópico se, para qualquer inteiro n e regiões $A_i : i = 1, \dots, n$, a distribuição conjunta de $Z(x_1), \dots, Z(x_n)$ for invariante sob rotação por meio do ângulo arbitrário α .

2.4.4 Propriedade de primeira ordem

O momento de primeira ordem ou propriedade de primeira ordem, é representado por uma função $\lambda(x)$, que é a função intensidade de primeira ordem. Ela indica a intensidade do processo na localização x . No caso do plano ($\mathbb{R}^2$) uma localização é definida por um sistema de coordenadas simples: $x = (x_i, x_j)$ (CAPRETZ et al., 2012). Segundo Câmara et al. (2002), os efeitos de primeira ordem, considerados globais ou de larga escala, correspondem à variações no valor médio do processo no espaço.

Considerando-se $Z = \{x_1, x_2, \dots, x_n\}$ um conjunto de pontos em uma determinada região A , S uma sub-região de A e dx uma pequena região em torno do ponto x . Assim, a função intensidade que caracteriza as propriedades de primeira ordem de um processo pontual é dada por

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[Z(dx)]}{|dx|} \right\},$$

em que $E[\]$ denota o valor esperado, dx é uma pequena região ao redor do ponto x e $|dx|$ é a área desta região.

Segundo Nunes (2011), para um processo pontual estacionário a medida da intensidade é invariante sob translação. Assim, $\lambda(x) = \lambda = \frac{E[A]}{|A|}$ é a intensidade de X e a interpretação é da mesma é dada como o número esperado de acontecimentos por unidade de área.

2.4.5 Propriedade de segunda ordem

Efeitos de segunda ordem, denominados locais ou de pequena escala, resultam da estrutura de correlação espacial ou dependência espacial no processo pontual e são formalmente descritos pela *função intensidade de segunda ordem* que é uma medida de

estrutura de dependência entre as localizações x_i e x_j (DIGGLE, 2003):

$$\lambda_2(x_i, x_j) = \lim_{|dx_i|, |dx_j| \rightarrow 0} \frac{E[Z(dx_i)Z(dx_j)]}{|dx_i||dx_j|}. \quad (2)$$

Um processo pontual espacial é denominado fracamente estacionário se o processo é invariante em localização, isto é,

$$\lambda(x) = \lambda \quad \forall x \in A \quad \text{e} \quad \lambda_x(x_i, x_j) = \lambda_2(\mathbf{h}) \quad \forall x_i, x_j \in A,$$

em que, $\mathbf{h} = x_i - x_j$ é o vetor bidimensional da mudança em localização espacial do ponto x_i ao ponto x_j (DIGGLE, 2003). Câmara et al. (2004) aponta que este processo equivale a dizer que o número esperado de eventos em uma localização arbitrária é constante e a dependência entre os eventos em duas localizações quaisquer depende apenas do vetor diferença $\mathbf{h}$ e não das localizações específicas x_i e x_j .

Conforme Maia (2005), quando o processo é estacionário, $\lambda(x)$ é uma constante, ou seja, $\lambda(x) = \lambda$; no caso de também ser isotrópico, $\lambda(x_i, x_j)$ se reduz à $\lambda(|h|)$, sendo $|h|$ a distância entre os dois pontos. Em caso de processo não estacionário, ou seja, quando a intensidade média varia na região A , a modelagem da estrutura de dependência $\lambda(x_i, x_j)$ deve incorporar a variação de $\lambda(x)$.

2.5 Análise de processos pontuais

Segundo Cressie (1993), a análise de processos pontuais caracteriza-se pela distribuição espacial da ocorrência dos indivíduos em termos da estimação das propriedades de primeira e segunda ordem. A seguir, serão discutidos os métodos para estimação das propriedades de primeira e segunda ordem.

2.5.1 Estimação da propriedade de primeira ordem

2.5.1.1 Estimação por kernel

Dentre as diversas técnicas existentes para representar a forma da distribuição dos pontos, destaca-se a estimação por kernel, a qual consiste num estimador de intensidade do processo pontual por meio da função kernel. A *Estimação por kernel* foi originalmente desenvolvida para obter uma estimativa suave de uma densidade de probabilidade univa-

riada ou bivariada a partir de uma amostra observada (BAILEY; GATRELL, 1995); em outras palavras, um histograma suave.

Segundo Câmara et al. (2004), essa função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse, conforme é apresentado na Figura 3.

Figura 3 – Estimador de intensidade pontual do processo em toda região de estudo (kernel)

Conforme Bailey e Gatrell (1995), x representa uma localização qualquer em A , e $x_1, \dots, x_n$ são localizações dos n eventos observados, então um estimador para λ , em x é dado por:

$$\hat{\lambda}_\tau(x) = \frac{1}{\delta_\tau(x)} \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{d(x_i; x)}{\tau}\right) \quad (3)$$

As distâncias $d(x_i; x)$ do ponto avaliando x a cada evento observado x_i , em unidades de τ , são medidas que contribuem para a estimativa de $\lambda(x)$ de acordo com a proximidade de cada evento x_i a x . $k(\cdot)$ é uma função de densidade de probabilidade bivariada escolhida adequadamente, conhecida como *kernel*, a qual é simétrica em relação à origem. O parâmetro $\tau > 0$ é conhecido como largura da banda (*bandwidth*) e determina a quantidade de suavização - essencialmente é o raio de um disco centrado em x dentro do qual os pontos x_i contribuirão significativamente para $\hat{\lambda}_\tau(x)$. O fator

$$\delta_\tau(x) = \int_A \frac{1}{\tau^2} k\left(\frac{x-u}{\tau}\right) dx \quad (4)$$

é uma correção de borda - a proporção do volume sob o kernel encontrado em x que está contido em A .

A técnica suaviza a superfície, calculando a densidade para cada região da área de estudo, utilizando interpolação. Isto permite a construção de uma superfície contínua de ocorrências das variáveis, inferindo para toda a área de estudo a variação espacial da variável, mesmo nas regiões onde o processo não tenha gerado nenhuma ocorrência real, permitindo verificar, em escala global, possíveis tendências de dados (KAWAMOTO, 2012). Segundo Illian et al. (2008), adotando-se um raio pequeno, é gerada uma superfície muito descontínua; se o raio for grande, a superfície poderá ficar muito suavizada. Silverman (1986) apresenta uma visão geral dos métodos para seleção da largura de banda para o estimador de kernel. No contexto da estatística espacial, várias metodologias vem sendo propostas até agora, mas ainda não foi definido qual o melhor método para a escolha da largura de banda ideal. Diggle (1985) e Berman e Diggle (1989), propuseram um critério baseado na minimização do erro quadrático médio do estimador de kernel.

No caso da função de interpolação $k(\cdot)$, é comum usar diferentes tipos de funções de estimação, conforme será mostrado na Tabela 1.

Tabela 1 – Diferentes tipos de função de estimação $k(h^a)$

kernel	$k(h)$
Kernel Uniforme	$\frac{1}{2}$
Kernel Triangular	$(1 - h)$
Kernel Quártico	$\frac{3}{\pi}(1 - h^2)^2$
Kernel Exponencial Negativo	$\frac{1}{2\pi}e^{-h}$
Kernel Gaussiano ou normal	$\frac{1}{2\pi\tau^b}\exp\left(-\frac{h^2}{2\tau^2}\right)$

^a h representa a distância euclidiana entre a localização, em que deseja-se calcular a função e o evento observado.

^b τ a largura da banda

A distribuição normal pondera os pontos dentro do círculo de forma que pontos mais próximos são pesados mais intensamente se comparados aos mais afastados. A distribuição uniforme pesa todos os pontos dentro do círculo igualmente. A função quártica pesa pontos próximos mais do que pontos distantes, todavia o decrescimento é gradual. A função triangular pesa pontos próximos mais do que pontos distantes dentro do círculo, mas o decrescimento é mais rápido. A função exponencial negativa pesa pontos próximos muito mais intensamente do que pontos distantes.

Segundo Olinda e Scalón (2010), o estimador de intensidade de kernel é muito útil uma vez que fornece uma visão geral da distribuição de primeira ordem dos eventos.

2.5.2 Estimação da propriedade de segunda ordem

Segundo Bailey e Gatrell (1995), as propriedades de segunda ordem ou dependência espacial em um processo espacial de pontos envolvem o relacionamento entre os eventos localizados em uma região. Uma abordagem prática é procurar relações de distâncias entre eventos e propriedades de segunda ordem. Os métodos de distância ao vizinho mais próximo e a função K são exemplos de procedimentos que consideram a distância entre eventos. Neste trabalho, apenas a função K será abordada.

2.5.2.1 Função K para processos pontuais homogêneos

A função K de Ripley (1977) é uma estatística descritiva do padrão espacial, também conhecida por “análise de segunda ordem reduzida” por se basear na medida de segunda ordem (RIPLEY, 1981). Segundo Anjos et al. (2004), a função K pode auxiliar no reconhecimento de distribuições espaciais, bem como nas interações entre os eventos, entre características de eventos, quando se considera mais de um tipo de indivíduo e variáveis medidas no indivíduo.

A função teórica K de um processo pontual estacionário é definida como (CRESSIE, 1993):

$$K(h) = \frac{2\pi}{\lambda^2} \int_0^h \lambda_2(x) x dx.$$

A medida de primeira ordem é representada por uma função λ , que é a função de intensidade de primeira ordem. Ela indica a intensidade do processo na localização x . No caso do plano $\mathbb{R}^2$, uma localização é definida por um sistema de coordenadas (x, y) .

Capretz et al. (2012) menciona que na função K , ao serem combinadas as funções de intensidade de primeira ordem e segunda ordem, resulta numa função descritiva do padrão espacial com interpretações mais diretas. Assim,

$$K(h) = \lambda^{-1} E[N_0(h)],$$

em que $N_0(h)$ é o número de eventos mais próximos dentro da distância h de um evento

arbitrário e λ é a intensidade.

Na CAE, também conhecida como Processo Poisson Homogêneo, as funções de intensidade são constantes, ou seja,

$$\lambda_1(x) \equiv \lambda \quad \text{e} \quad \lambda_2(x_i, x_j) = \lambda_2(\|x_i - x_j\|) \equiv \lambda_2(x)$$

em que $\|x_i - x_j\|$ é a distância euclidiana entre os pontos x_i e x_j .

Como consequência, a função K sobre a hipótese de CAE será:

$$\begin{aligned} K(h) &= \frac{2\pi}{\lambda^2} \int_0^h \lambda_2(x) x dx. \\ &= \frac{2\pi}{\lambda^2} \int_0^h \lambda_2(x) x dx. \\ &= \pi h^2 \end{aligned}$$

Portanto, a função K pode ser facilmente utilizada para testar a hipótese de CAE. Assim, para um processo aleatório teríamos $K(h) = \pi h^2$, para um processo agregado os pontos têm em média mais vizinhos do que o esperado sob a hipótese de CAE, consequentemente, $K(h) > \pi h^2$ e, para um processo regular os pontos têm em média menos pontos do que esperado sob a hipótese de CAE, consequentemente, $K(h) < \pi h^2$.

O estimador da função K com correção de borda proposta por Ripley (1977), discutido com detalhes em Cunha (2003) é dado por:

$$\hat{K}(h) = \frac{|A|}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_h(d_{ij})}{w_{ij}}$$

em que $|A|$ é a área da região, n é o número de eventos observados, $I_h(d_{ij})$ é uma função indicatriz cujo valor é 1 se $(d_{ij}) \leq h$ e 0 caso contrário, e w_{ij} é a proporção da circunferência do círculo centrado no evento i em que está dentro da região (correção devido ao efeito da borda).

O gráfico da função K nem sempre é fácil de interpretar, uma vez que a linha obtida para a hipótese nula é parabólica. Por essa razão, Besag e Diggle (1977) propuseram

o estimador modificado (5).

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \quad (5)$$

A função $L(h)$ é mais fácil de interpretar do que $K(h)$ pois a linha obtida para a hipótese nula é uma reta com abscissa igual a zero, ou seja, $L(h) = 0$ em todas as distâncias h . Observe na Tabela 2 um comparativo entre os valores de $K(h)$ e $L(h)$:

Tabela 2 – Comparação entre os valores de $L(h)$ e $K(h)$

Tipo de Processo	$L(h)$	$K(h)$
Aleatório	$= 0$	$= \pi h^2$
Regular	< 0	$< \pi h^2$
Aglomerado	> 0	$> \pi h^2$

O gráfico de $\hat{L}(h)$ em função da distância h indica atração espacial entre eventos ou agrupamentos para valores positivos, sendo o agrupamento mais forte em picos positivos, o que indica repulsão espacial ou regularidade em pontos de valores negativos. Uma representação gráfica da função $\hat{L}(h)$ esta representada na Figura 4.

Figura 4 – Representação gráfica da função $\hat{L}(h)$

Segundo Câmara et al. (2004), é possível utilizar a função $\hat{L}(h)$ como método formal para se comparar estatisticamente a distribuição dos eventos observados com o que

se esperaria na hipótese de completa aleatoriedade espacial (CAE). Para isso, são construídos envelopes de simulação para a distribuição de CAE, afim de acessar a significância dos desvios. Os envelopes de simulação inferior e superior são construídos a partir de k simulações independentes de n eventos na região A .

$$\begin{aligned} L(h) &= \min_{i=1,\dots,k} \{\hat{L}_i(h)\} \\ U(h) &= \max_{i=1,\dots,k} \{\hat{L}_i(h)\} \end{aligned}$$

$L(h)$ e $U(h)$ são os mínimos e os máximos da função L , respectivamente, calculadas a partir de k simulações independentes de processo aleatórios com n eventos. Na análise do gráfico com a distribuição e os envelopes (Figura 5), picos positivos na função estimada $\hat{L}(h)$ que estão acima do envelope superior evidenciam ocorrências de agrupamento na escala considerada, portanto, se todos os valores da função $\hat{L}(h)$ estiverem acima do envelope superior e com valores positivos, teremos agrupamentos em todas as escalas. Depressões negativas na função estimada $\hat{L}(h)$ que estiverem abaixo do envelope inferior, evidenciam regularidade nessa escala, portanto, se todos os valores de $\hat{L}(h)$ estiverem abaixo do envelope inferior e com valores negativos, tem-se regularidade em todas as escalas.

Figura 5 – Representação gráfica da função $\hat{L}(h)$ com envelope

2.5.2.2 Função K para processos pontuais não-homogêneos

A função $K(h)$ não homogênea é uma generalização da função $K(h)$ usual, uma vez que o processo pontual estacionário tem $\lambda(x) \equiv \lambda$, lado direito da função (5) reduz-se à definição de $K(h)$ disponível em Ripley (1977). Analogamente ao caso estacionário,

$$\hat{L}_{inhom} = \sqrt{\frac{\hat{K}_{inhom}}{\pi}}$$

O estimador da função $K(h)$ não-homogênea incluindo a correção do efeito de borda é obtido por (BADDELEY; MØLLER; WAAGEPETERSEN, 2000):

$$\hat{K}_{inhom}(h) = \frac{1}{|A|} \sum_i^n \sum_{i \neq j}^n \frac{w_{i,j}^{-1} I_h(d_{i,j})}{\lambda(x_i) \lambda(x_j)}, \quad h \geq 0 \quad (6)$$

em que $|A|$ é a área da região, n é o número de eventos observados, $I_h(d_{i,j})$ é uma função indicatriz cujo valor é 1 se $(d_{i,j}) \leq h$ e 0 caso contrário, e w_{ij} é a proporção da circunferência do círculo centrado no evento i em que está dentro da região (correção devido ao efeito da borda).

As propriedades da curva de $K_{inhom}(h)$ são similares às da função $K(h)$ usual. Como na função $K(h)$, o gráfico da função $K_{inhom}(h)$ também é difícil de interpretar, por essa razão, Besag e Diggle (1977) propuseram o estimador modificado da função K_{inhom} (7).

$$\hat{L}_{inhom}(h) = \sqrt{\frac{\hat{K}_{inhom}(h)}{\pi}} - h \quad (7)$$

2.6 Estudos caso-controle

Considere-se um tipo de estudo em que temos dois tipos de eventos por exemplo, neste trabalho, o número de árvores da espécie em análise, denominado de caso e as outras espécies que fazem parte da região, controles. Sendo assim, um estudo com resposta do tipo binomial (assumindo 1 para caso e 0 para o controle), dependente dos diversos fatores ambientais analisados.

Considerando $(x_1, x_2, \dots, x_{n_1})$ as localizações na região de estudo A dos n_1

casos e $(x_{n_1+1}, x_{n_1+2}, \dots, x_n)$ as localizações dos $(n - n_1)$ controles como observações de dois processos Poisson, com intensidades $\lambda_1(x)$ e $\lambda_2(x)$, respectivamente. Câmara et al. (2002) defini o risco na localidade A como

$$\rho(x) = \log \left\{ \frac{\lambda_1(x)}{\lambda_2(x)} \right\}$$

e o objetivo da análise é investigar a variação espacial de $\rho(x)$ em A .

Kelsall e Diggle (1998) consideram o uso de regressão binária não-paramétrica e apresentam uma proposta semi-paramétrica na estimação da razão de densidades conhecida como abordagem de Modelos Aditivos Generalizados (GAM). O modelo semi-paramétrico com efeito de covariáveis, sugerido pelos autores, é dado por:

$$\log \left\{ \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} \right\} = \boldsymbol{\beta}' d(\mathbf{x}) + g(\mathbf{x}), \quad p(\mathbf{x}) = p(Y = 1 | \mathbf{x})$$

em que, Y é a variável resposta e tem a forma sim/não, zero/um (espécies/outras espécies); a função de ligação da regressão é o logit, como usuais para dados binomiais; $d(\mathbf{x})$ é o vetor de covariáveis; $\boldsymbol{\beta}$ é o vetor de parâmetros estimados pelo modelo, que no caso da regressão logística é a razão de chances relacionadas a cada covariável, ou seja, é o incremento de cada covariável ao logito; e $g(\mathbf{x})$ é uma função suave desconhecida das coordenadas espaciais de $\mathbf{x}$ de casos e controles.

Segundo Shimakura et al. (2001), o ganho desse tipo de modelagem é possibilitar a estimativa da variação espacial do risco, controlando pelos fatores conhecidos de variação do risco. Os procedimentos de estimação dos parâmetros destes modelos baseia-se em métodos iterativos usuais de modelos aditivos generalizados, em que se estima os parâmetros da regressão, e sobre os resíduos estima-se a função kernel, e assim sucessivamente até que as estimativas não mais se alterem. O método permite identificar áreas de sobre ou sub risco significativamente diferente da média global (CÂMARA et al., 2002).

3 METODOLOGIA

3.1 Material

Os dados experimentais, utilizados neste trabalho, foram coletados em um estudo desenvolvido em quatro Unidades de Conservação do Estado de São Paulo (Figura 6(a)), em parcelas permanentes, dentro do projeto “Diversidade, dinâmica e conservação em florestas do Estado de São Paulo: 40 *ha* de parcelas permanentes” do programa Biota da FAPESP. As parcelas permanentes foram instaladas nas Unidades de Conservação do Parque Estadual Ilha do Cardoso (PEIC), Parque Estadual de Carlos Botelho (PECB), Estação Ecológica dos Caetetus (EEC) e a Estação Ecológica de Assis (EEA). Em cada Unidade de Conservação, foi montada uma parcela permanente de 320×320 m (Figura 6(b)), totalizando 10,24 *ha*, subdividida em 256 subparcelas contíguas de 20×20 m, totalizando 400 m^2 , onde todas as árvores com circunferência na altura do peito igual ou superior a 15 *cm* foram medidas, georreferenciadas e identificadas. Para este trabalho foram utilizadas apenas as informações obtidas na Estação Ecológica de Assis.

Figura 6 – (a) Localização das quatro Unidades de Conservação no Estado de São Paulo e (b) Parcela permanente subdividida em 256 subparcelas contíguas de 20×20 ($400 m^2$)

Área de Estudo

A Estação Ecológica de Assis, com uma área de aproximadamente 1300 *ha*, localiza-se no município de Assis na região oeste do Estado de São Paulo, entre as coordenadas geográficas $22^{\circ}33'$ a $22^{\circ}36'$ S, e $50^{\circ}23'$ e $50^{\circ}22'$ O. A precipitação anual nessa região fica em torno de 1400 *mm*, concentrada no verão (CAPRETZ et al., 2012).

Dados

Foram levantadas 22576 árvores em Assis. As dez espécies mais abundantes nesse área são apresentadas na Tabela 3. As três espécies mais abundantes, *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica* foram utilizadas na análise, uma vez que a soma de seus indivíduos corresponde a 34,8% das árvores presentes na parcela.

Tabela 3 – Espécies com maior densidade na Estação Ecológica de Assis. As espécies estudadas neste trabalho estão em negrito

Espécies	Número de indivíduos
Copaifera langsdorffii	3977
Vochysia tucanorum	2107
Xylopia aromatica	1770
Ocotea corymbosa	1555
Machaerium acutifolium	1049
Rapanea umbellata	991
Tapirira guianensis	973
Myrcia multiflora	897
Myrcia guianensis	880
Symplocos mosenii	833

A *Copaifera langsdorffii*, popularmente conhecida como “copaíba ou copaibeiras” foi a espécie escolhida para este estudo por ser a espécie com maior número de ocorrências na região de estudo. Segundo Sousa et al. (2012), as copaibas são árvores nativas da região tropical da América Latina e também da África Ocidental. Estas árvores produzem o óleo de copaíba, que tem sido utilizado pela medicina popular, por mais de 500 anos, para diversas finalidades terapêuticas (PIERI; MUSSI; MOREIRA, 2009). Este óleo também é intensamente aplicado nas indústrias de perfumes e cosméticos, como loções capilares, fixador, cremes, sabonetes, entre outros, sendo utilizado também na indústria de vernizes e tintas (VEIGA; PINTO, 2002), o que demonstra a importância econômica desta espécie.

Segundo Clemente (2010) a *Vochysia tucanorum*, pode ser encontrada em campos rupestres da Cadeia do Espinhaço em Minas Gerais e em campos de altitude da Bahia, e no cerrado, sendo mais comum no cerrado. Planta pioneira, adaptada a terrenos arenosos e pobres podendo ser empregada em reflorestamentos heterogêneos destinados à recomposição de áreas de preservação permanente. Ainda segundo o autor, esta espécie vem sendo utilizada na medicina popular, principalmente na região central do país, na prevenção e tratamento de lesões da mucosa gástrica, comumente usada na forma de chás

feitos a partir de suas folhas.

As árvores da espécie *Xylopia aromatica* apresentam altura de até 8 m. A casca do caule e as folhas apresentam propriedades anti-inflamatórias quando usadas na forma de chá. A esta espécie, também são atribuídas propriedades vermífuga, febrífuga e caminativa (MARONI; STACI; MACHADO, 2006).

Caracterização edáfica

As tradagens foram realizadas até 1 metro de profundidade em cada uma das 256 subparcelas (malha rígida) de cada parcela permanente, com coleta de amostras em três profundidades (0-5, 5-20 e 80-100 *cm* na EEC e no PEIC; 0-5, 0-25 e 80-100 *cm* na EEA e no PECB). As análises físico-químicas de cada amostra foram feitas de acordo com a EMBRAPA (2006).

Neste estudo, para cada subparcela, foram utilizadas as seguintes covariáveis ambientais:

- características topográficas: altitude (média de cada subparcela) e declividade (média de cada subparcela);
- características edáficas: profundidade do solo, densidade, ponto de inflexão da curva de retenção de água (índice S), propriedades granulométricas (percentual de argila, areia e silte), atributos químicos (pH, Al (cmolc/dm³), Ca (cmolc/dm³), Mg (cmolc/dm³), K (cmolc/dm³), P (cmolc/dm³), Na (cmolc/dm³), teor de matéria orgânica (MO(%)), e índices compostos saturação por base (V(%)), saturação por alumínio (m(%)) e capacidade de troca catiônica (cmolc/dm³)

As variáveis referentes a textura de solo (percentuais de areia, silte e argila) somam 100%, sendo portanto utilizada a transformação log-aditiva, em que as variáveis foram consideradas na seguinte forma: $X1 = \log(\text{areia}/\text{argila})$ e $X2 = \log(\text{silte}/\text{argila})$.

3.2 Método

Dentre as metodologias disponíveis para a análise de dados espaciais, a análise de processos pontuais é a que melhor se adéqua aos dados individuais. Grego¹ (em fase de elaboração), utilizando o mesmo banco de dados considerou as abordagens geoestatística e de dados de área, considerando como variável resposta a área basal das árvores.

¹Grego, S. (Escola Superior de Agricultura “Luiz de Queiroz”). **Modelos não espaciais e espaciais para relacionar variáveis de solos e número de espécies em uma área de vegetação natural**

Nesse estudo, além de considerar as covariáveis citadas anteriormente (originais), foram utilizadas variáveis expressas por fatores. Para a obtenção dessas novas covariáveis, foi utilizada a Análise Fatorial (AF), visando identificar um número menor de covariáveis alternativas, não correlacionadas e que, de algum modo sumarizem as informações principais das covariáveis originais encontrando fatores ou variáveis latentes (MINGOTI, 2005). Para verificar a adequação da análise fatorial foram utilizados os testes de esfericidade de Bartlett e Kaiser-Meyer-Olkin (KMO). Verificada a adequação dos dados à análise fatorial, aplicou-se o método das componentes principais para extração dos fatores via rotação ortogonal do tipo varimax, usando o Critério da Raiz Latente e Critério da Percentagem da Variância Explicada para escolha do número de fatores a extrair.

Também foi construída uma nova variável denominada “demais espécies”, que leva em consideração a presença de outras espécies na área de estudo para considerar o efeito dessas espécies no padrão de distribuição. Para a construção desta variável foi utilizado o estimador de intensidade de kernel descrito por Bailey e Gatrell (1995) e largura da banda (*bandwidth*), para estimar a intensidade das outras espécies no mesmo grid das variáveis disponibilizadas pelo programa Biota. Assim, na medida em que uma espécie foi analisada, a intensidade das outras espécies na região foram consideradas como uma nova variável. Os cálculos foram realizados no programa R versão 3.0.1 (R Development Core Team, 2008), sendo seus valores estimados pela função `kernel2d` a qual estima a intensidade do processo de pontos com função de interpolação quártico e largura de banda obtida por meio da função `mse2d`, ambas do pacote `Splancs` (ROWLINGSON; DIGGLE, 2013).

Realizou-se uma inspeção visual da homogeneidade espacial, através de mapas de superfície, utilizando as estimativas de intensidade de kernel dos padrões pontuais observados (DIGGLE, 2003; BAILEY; GATRELL, 1995) para cada uma das espécies, no intuito de verificar a ausência de homogeneidade na distribuição espacial. Para a análise da distribuição espacial de cada espécie, foi utilizado a função L , variação da função $K(\cdot)$ de Ripley para processos pontuais homogêneos univariada (RIPLEY, 1977), levando em consideração a correção de borda. A função L possibilita detectar o padrão espacial (aleatório, regular ou agrupado) em diferentes escalas simultaneamente (BATISTA; MAGUIRE, 1998). A função L versus a distância indica atração espacial caso assuma valores positivos e, caso assumam valores negativos, indica repulsão espacial ou regularidade dos eventos (CARAM et al., 2006). Para compara a estimativa da função L de um conjunto de dados observados, são

realizadas 1000 simulações Monte Carlo por meio da CSR, para obter os envelopes superior e inferior, para testar a significância dos desvios da função L com a CSR. A função estimada $\hat{L}$ e os envelopes de simulação são obtidos pelas funções **Lest** e **envelope**, respectivamente, ambas do pacote **spatstat** (BADDELEY; TURNER, 2005). Quando os valores estimados da função L estão dentro do envelope, aceita-se a hipótese de que os eventos observados têm padrão aleatório.

Modelos do processo Poisson homogêneo, processo Poisson não homogêneo e o processo log gaussiano de Cox foram utilizados para modelar a ocorrência de árvores das três espécies mais abundantes na Estação Ecológica de Assis. Ou seja, no primeiro modelo supomos que a intensidade da ocorrência das árvores é a mesma em toda a área. Para o segundo modelo assumimos que não existe interação com a ocorrência de uma árvore vizinha (ocorrer uma árvore em um local não afeta a probabilidade de ocorrência de outra), e que os padrões espaciais de ocorrência das árvores resultam somente dos efeitos espacial da heterogeneidade ambiental da região. Por fim, para o terceiro modelo consideramos que além dos efeitos das covariáveis ambientais existe uma fonte de variação extra, ou seja, algum fator espacialmente estruturado que não foi considerado entre as variáveis analisadas.

Para o ajuste dos modelos do processo Poisson não homogêneo e do modelo log gaussiano de Cox foi considerando como covariáveis, as variáveis da análise química e de perfil e variáveis obtidas pela análise fatorial, sempre ajustados com e sem a variável intensidade para modelar a intensidade do processo, objetivando verificar se de fato existe uma associação entre a localização das árvores e as variáveis analisadas. Os modelos propostos referentes à situação de não homogeneidade consideram a intensidade como função dependente das variáveis (originais, componentes principais ou fatores). Foram utilizadas funções log-lineares, log-quadráticas e log-cúbicas para verificar a relação das covariáveis com a intensidade da ocorrência das árvores.

Foi ainda realizada uma seleção de variáveis via método *stepwise* em dois estágios: no primeiro incluiu-se somente as covariáveis espaciais (originais ou transformadas); no segundo além de incluir as covariáveis espaciais também foi considerada a variável construída através do estimador de kernel. Dessa forma, foi considerada uma grande variedade de modelos do processo Poisson não homogêneo. Com o intuito de selecionar o modelo que apresenta o melhor ajuste aos dados, foi utilizado o Critério de Informação de Akaike (AIC) Akaike (1974) para a escolha dos modelos de regressão log-linear para cada espécie

analisada. O AIC para um determinado modelo é calculado por:

$$AIC = -2\ln(L) + 2K$$

em que L é a verossimilhança do modelo, que indica o ajuste, e K é o número de parâmetros.

Como o valor do AIC individual não é interpretável, uma alternativa proposta por Burnham e Anderson (2002) para avaliar a plausibilidade dos modelos foi adotada. Os autores recomendam o cálculo das diferenças de AIC entre os modelos candidatos e o modelo que apresenta o menor valor de AIC (AIC_{min}), dado por:

$$\Delta_i = AIC_i - AIC_{min}.$$

Os valores dos Δ_i são fáceis de interpretar e permitem uma rápida comparação entre os modelos candidatos. Quanto maior for a diferença entre os modelos (Δ_i) menos plausível é o modelo candidato e, se o valor do Δ_i estiver entre 0 e 2 podem ser considerados iguais.

Para os modelos do risco, o procedimento para a estimação de β e $g(\mathbf{x})$, baseia-se em métodos iterativos usuais de modelos aditivos generalizados, que segundo Shimakura et al. (2001), pode ser resumido no seguinte algoritmo:

Passo 1: Faz-se $g(\mathbf{x}) = 0$ e obtém-se as estimativas iniciais dos coeficientes de β ajustando um modelo de regressão logística

$$\log \left\{ \frac{p(x)}{1 - p(x)} \right\} = \beta d(\mathbf{x})$$

Passo 2: Calcula-se $\hat{\eta}_i = \hat{\beta}x_i + g(\mathbf{x})$, $\hat{p}_i = \exp(\hat{\eta}_i) / [1 + \exp(\hat{\eta}_i)]$ e $z_i = \hat{\eta}_i + (y_i - \hat{p}_i) / \hat{p}_i(1 - \hat{p}_i)$

Passo 3: Ajusta-se um modelo aditivo da forma $Z = \beta d(\mathbf{x}) + g(\mathbf{x}) + \varepsilon$:

- a) Estima-se $g(\mathbf{x})$ usando regressão kernel ponderada de $u_i = z_i - \hat{\beta}x_i$ em x_i com pesos $w_i = \hat{p}_i(1 - \hat{p}_i)$:

$$\hat{g}(x) = \frac{\sum_{i=1}^n w_i K\left(\frac{x-x_i}{h}\right) u_i}{\sum_{i=1}^n w_i K\left(\frac{x-x_i}{h}\right)},$$

em que $K(\cdot)$ é a função Kernel. O parâmetro h , representa o grau de suavização

da superfície de risco estimada;

- b) Estima-se β por métodos de mínimos quadrados ponderados de $z_i - g(\hat{\mathbf{x}})$ em u_i com pesos w_i ;
- c) Repete-se os passo (a) e (b) até a convergência das estimativas

Passo 4: Repete-se os passos (2) e (3) até convergência das estimativas

O critério para a escolha de h varia conforme a distribuição espacial dos dados e dos objetivos da análise.

A abordagem geral utilizada nesse trabalho é apresentada na Figura 7. As análises estatísticas deste trabalho foram realizadas no Ambiente R (R Development Core Team, 2008), bem como a utilização dos pacotes **spatstat** (BADDELEY; TURNER, 2005) para o ajuste dos modelos do processo pontual, **splancs** (ROWLINGSON; DIGGLE, 2013) para calcular a estimativa de kernel utilizado para a construção da variável “demais espécies” e o **maptools** (LEWIN-KOH et al., 2009) para transformar as variáveis no objeto *im*, necessários na utilização das funções *ppm* e *kppm* do pacote **spatstat** citado anteriormente.

Figura 7 – Fluxograma da abordagem geral do trabalho

4 RESULTADOS E DISCUSSÃO

Na Estação Ecológica de Assis foram encontrados 3977 indivíduos da espécie *Copaifera langsdorffii*, 2107 indivíduos da espécie *Vochysia tucanorum* e 1770 indivíduos da espécie *Xylopia aromatica* com circunferência na altura do peito igual ou superior a 15 cm. A representação gráfica da localização das árvores está representada na Figura 8.

Figura 8 – Distribuição espacial das três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

Na Figura 8 observa-se que algumas regiões da área em estudo podem apresentar aglomeração de pontos para todas as espécies, indicando a falta de homogeneidade no padrão da distribuição das espécies. Este fato pode ser verificado através de uma avaliação da propriedade de primeira ordem, analisando a intensidade pontual dos indivíduos

por meio do estimador de Kernel apresentado na seção 2.5.1.1.

A intensidade suavizada de Kernel para a espécie *Copaifera langsdorffii* (Figura 9(a)), calculada a partir das posições das árvores mostrou “áreas quentes” (onde a intensidade de ocorrência das árvores dessa espécie é mais elevada) na região nordeste, leste e noroeste da região e “áreas frias” (onde a intensidade de ocorrência da espécie é baixa) na região sul e central, indicando que o processo de ocorrência da espécie não é completamente aleatório. Para as espécies *Vochysia tucanorum* (Figura 9(b)) e *Xylopia aromatica* (Figura 9(c)) o mesmo fenômeno foi observado, indicando que o padrão de distribuição dessas

Figura 9 – Distribuição espacial das três espécies mais abundantes na Estação Ecológica de Assis por meio da suavização de Kernel, em que a escala (lado direito) representa a intensidade de pontos por unidade de área. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

espécies também não apresenta a completa aleatoriedade espacial.

Constatou-se que a distribuição das três espécies mais abundantes na Estação Ecológica de Assis ocorreu segundo um processo não homogêneo, ou seja, as espécies

apresentaram-se com alta intensidade em algumas áreas, o que é comum para a maioria das espécies arbóreas tropicais (CONDIT et al., 2000; HE; LEGENDRE; LAFRAKIE, 1997; PLOTKIN; CHAVE; ASHTON, 2002). Para a análise de segunda ordem foi utilizada a função L com correção de borda e envelope de simulação baseado na técnica de Monte Carlo para as três espécies em estudo.

Na Figura 10 para todas as escalas analisadas, a função de distribuição estimada ($\hat{L}$) manteve-se acima do envelope de simulação superior para as três espécies indicando um nível de agregação, ou seja, o número de indivíduos observados na vizinhança

Figura 10 – Estimativas da função L (linha cheia) e envelopes de simulação, contra a distribuição teórica sob hipótese de CAE (linha tracejada) referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

de um indivíduo qualquer foi maior do que o esperado para um padrão aleatório. Para a

espécie *Copaifera langsdorffii* a curva da função L apresenta um pico evidente por volta de 64m, ou seja, para uma vizinhança nessa distância, o número esperado de árvores da espécie *Copaifera langsdorffii* é muito superior ao que se esperaria em um processo de CAE. Para as espécies *Vochysia tucanorum* (Figura 10(b)) e *Xylopia aromatica* (Figura 10(c)) não foi observado um pico evidente para a distância analisada.

Para explicar esse padrão de aglomeração foram utilizadas as covariáveis originais e as covariáveis submetidas a análise multivariada pelo método da Análise Fatorial. Para cada uma delas foi utilizada uma sequência de modelos, considerando ou não a adição de uma covariável construída denominada de “demais espécies”.

4.1 Efeitos das covariáveis espaciais na falta de homogeneidade das três espécies mais abundantes na Estação Ecológica de Assis

4.1.1 Variáveis originais

Nesta seção, foi avaliado o efeito das covariáveis espaciais na variação espacial das espécies mais abundantes na Estação Ecológica de Assis, ajustando-se modelos do processo de Poisson homogêneo, processo Poisson não homogêneo e o modelo log gaussino de Cox.

Para o modelo do PPNH foram considerados como covariáveis o pH H₂O, MO, P, Na, K, Ca, Mg, Al, CTC, V, m, S, Declividade, Altitude, X1(log(areia/argila)) e X2(log(silte/argila)) nas profundidades 0–5, 5–25 e 80–100 e como resposta a ocorrência da espécie em análise. Além das covariáveis citadas foi construída uma nova covariável através do estimador de Kernel. Esta covariável é a intensidade de todas as outras espécies na região de estudo, objetivando levar em consideração a presença das outras espécies na área. A ideia é explicar as altas intensidades encontrados anteriormente pelas covariáveis espaciais e a covariável construída. Ao ajustar um modelo pelo método da máxima verossimilhança da forma $\lambda(x) = \exp\{\alpha + \beta Z(x)\}$, em que α e β são o intercepto e coeficiente das covariáveis do modelo, respectivamente, temos como resposta a função intensidade da ocorrência das espécies em estudo, explicada pelas covariáveis espaciais e a covariável construída.

Para a seleção de variáveis foi utilizado o método *Stepwise*, o qual foi realizado em duas etapas: na primeira o modelo inicial utilizado levou em consideração apenas as

covariáveis espaciais; para a segunda etapa, além de se considerar as covariáveis espaciais, o modelo inicial também levou em consideração a utilização da covariável construída. Para todas as covariáveis selecionadas, foi verificada a necessidade de transformação (quadrática ou cúbica).

Na Tabela 4 são apresentados os valores do AIC para cada um dos modelos testados, bem como os coeficientes das covariáveis espaciais selecionadas. Observando os valores de AIC dos modelos avaliados para a espécie *Copaifera langsdorffii*, revela-se como o modelo com melhor ajuste o modelo $M2_C$, por apresentar o menor AIC entre todos os modelos ajustados, de modo que o mesmo foi denominado modelo suporte (AIC_{min}) para o cálculo da diferença de AIC conforme Burnham e Anderson (2002). A diferença de AIC (Δ_i) entre os modelos $M0_C$ e $M2_C$ e os modelos $M1_C$ e $M2_C$ foram 563,24 e 38,55, respectivamente. O modelo $M0_C$, que considerou uma média constante em toda a região foi o modelo com menor plausibilidade quando comparado com os modelos $M1_C$ e $M2_C$, confirmando a não homogeneidade espacial na região. O modelo $M2_C$ que considerou tanto as covariáveis espaciais quanto a covariável construída foi o modelo mais plausível para estimar a intensidade de ocorrência das espécie *Copaifera langsdorffii*, indicando que as variáveis do solo são importantes para explicar o padrão espacial dessa espécie.

Tabela 4 – Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Copaifera langsdorffii* e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>			(continua)
	$M0_C$	$M1_C$	$M2_C$	
Intercepto	-3,35(0,02)	-20,08 (2,65)	-17,68 (2,68)	
MO_{05}		0,17 (0,04)	0,17 (0,04)	
MO_{80100}		10,98 (2,75)	7,86 (2,82)	
MO_{80100}^2		-11,06 (2,17)	-8,76 (2,24)	
P_{05}		-13,21 (3,15)	-16,56 (3,31)	
P_{05}^2		-8,53 (2,80)	-10,57 (2,90)	
P_{05}^3		6,27 (2,48)	5,79 (2,50)	
K_{525}		10,15 (3,32)	8,85 (3,38)	
K_{525}^2		6,57 (2,44)		
K_{80100}		-7,97 (2,56)	-8,54 (2,62)	
Mg_{05}		-15,95 (2,69)	-14,32 (2,75)	
Mg_{05}^2		-10,30 (2,53)	-11,31 (2,62)	
Al_{05}		18,82 (2,76)	19,47 (2,83)	
Al_{05}^2		6,57 (2,21)	7,14 (2,23)	

Tabela 4: Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “de mais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Copaifera langsdorffii* e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>			(conclusão)
	M0 _C	M1 _C	M2 _C	
Al ₈₀₁₀₀	8,12 (2,83)		8,75 (2,90)	
Al ₈₀₁₀₀ ²	-14,20 (2,61)		-13,89 (2,63)	
Al ₈₀₁₀₀ ³	-5,45 (2,54)		-5,13 (2,57)	
CTC ₀₅	-22,69 (4,17)		-23,07 (4,42)	
CTC ₀₅ ²	9,81 (2,62)		8,50 (2,63)	
CTC ₅₂₅	-1,46 (3,65)			
CTC ₅₂₅ ³			-8,21 (2,50)	
V ₈₀₁₀₀	4,00 (3,15)			
V ₈₀₁₀₀ ²	-5,46 (2,50)		-6,59 (2,55)	
m ₈₀₁₀₀	-0,02 (0,01)		-0,02 (0,01)	
Declividade	-0,18 (0,06)		-0,18 (0,06)	
Altitude	0,03 (0,00)		0,03 (0,00)	
X1 ₀₅			-10,11 (2,84)	
X1 ₀₅ ²			-6,73 (2,48)	
Outras _{S.C.}			13,68 (2,30)	
Outras _{S.C.} ²			-5,28 (2,25)	
AIC	34597,84	34073,15	34034,6	

Por meio da Tabela 5 referente aos coeficientes e valor do AIC para a espécie *Vochysia tucanorum*, é possível observar que o modelo com melhor ajuste para esta espécie também apontou como o melhor modelo o M2_V, por apresentar o menor AIC entre todos os modelos ajustados. Portanto, o modelo M2_V foi denominado modelo suporte (AIC_{min}) para o cálculo da diferença de AIC. A diferença de AIC (Δ_i) entre os modelos M0_V e M2_V e os modelos M1_V e M2_V foram 768,06 e 139,85, respectivamente. O modelo M0_V, que considerou uma média constante em toda a região foi o modelo com menor plausibilidade quando comparado com os modelos M1_V e M2_V, confirmando a não homogeneidade espacial da espécie na região. O modelo M2_V que considerou tanto as variáveis espaciais quanto a variável interpolada através do estimador de kernel foi o mais plausível.

Tabela 5 – Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Vochysia tucanorum* e os coeficientes das variáveis preditoras dos modelos selecionados

(continua)

Covariável	<i>Modelos</i>		
	M0 _C	M1 _C	M2 _C
Intercepto	-3,98(0,02)	-3,68(0,24)	-5,20(0,60)
pH H2O ₅₂₅		5,61(2,83)	
pH H2O ₅₂₅ ³		6,80(2,74)	
pH H2O ₈₀₁₀₀			7,05(2,88)
pH H2O ₈₀₁₀₀ ²			-9,89(2,76)
MO ₀₅		-17,38(4,60)	-17,50(4,55)
MO ₀₅ ²			10,34(3,10)
MO ₅₂₅ ²		-10,00(2,72)	-12,05(2,90)
MO ₈₀₁₀₀		11,12(3,20)	14,08(3,18)
MO ₈₀₁₀₀ ³		8,28(2,42)	12,31(2,60)
P ₀₅		20,34(3,74)	25,56(3,84)
P ₀₅ ²			-11,36(3,38)
P ₀₅ ³		-7,85(2,84)	
Na ₀₅ ²			-5,81(2,71)
Na ₀₅ ³		-15,36(2,60)	-14,20(2,56)
K ₀₅			-11,28(4,75)
K ₀₅ ²			12,77(2,67)
K ₅₂₅ ³		13,98(2,66)	
K ₈₀₁₀₀ ²			10,30(3,73)
K ₈₀₁₀₀ ³			-19,00(3,26)
Al ₀₅			-10,49(5,10)
Al ₀₅ ²		-15,77(2,99)	-10,95(3,26)
Al ₅₂₅			-6,36(2,90)
Al ₈₀₁₀₀			0,18(0,05)
CTC ₀₅ ³		7,20(2,78)	11,95(2,93)
CTC ₈₀₁₀₀			-0,04(0,01)
CTC ₈₀₁₀₀ ²		-7,26(2,93)	
V ₅₂₅		-11,20(3,97)	-19,28(3,70)
V ₅₂₅ ²			7,57(3,08)
V ₅₂₅ ³		-8,83(2,59)	-7,81(2,72)
V ₈₀₁₀₀			-0,07(0,03)
m ₀₅			0,01(0,01)
m ₈₀₁₀₀		-23,67(4,26)	-57,78(9,44)
m ₈₀₁₀₀ ²		-7,00(2,63)	-17,11(3,81)
m ₈₀₁₀₀ ³			6,06(2,94)
s ₅₂₅ ²			-5,08(2,41)
s ₈₀₁₀₀ ³		-8,62(2,82)	
Declividade			10,97(3,14)
Declividade ³		-17,57(3,82)	-20,36(3,66)
Altitude		54,23(4,65)	62,79(4,32)
Altitude ²		-45,39(4,31)	-42,40(3,59)

Tabela 5: Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Vochysia tu canorum* e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>			(conclusão)
	M0 _C	M1 _C	M2 _C	
Altitude ³		15,67(3,75)		
X1 ₀₅ ²			18,07(2,55)	
X1 ₈₀₁₀₀ ²		8,83(2,66)		
X2 ₀₅			-0,43(0,08)	
Outras _{S,V} ²			11,40(2,38)	
AIC	21007,75	20379,54	20239,69	

Assim como nas outras espécies foram ajustados modelos considerando a intensidade constante na região, variáveis espaciais como covariáveis e variáveis espaciais com adição da intensidade de outras espécies na região para a espécie *Xylopia aromatica* (Tabela 6). O valor do AIC dos modelos testados apontou como o melhor modelo o M2_X, por apresentar o menor AIC entre todos os modelos ajustados. A diferença de AIC (Δ_i) entre os entre os modelos M0_X e M2_X e os modelos M1_X e M2_X foram 719,89 e 49,32, respectivamente. Assim, foi possível observar que a presença de outras espécies na região influencia no padrão de agregação, já que a mesma foi significativa em todos os modelos selecionados.

Tabela 6 – Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Copaifera langsdorffii* e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>			(continua)
	M0 _C	M1 _C	M2 _C	
Intercepto	-4.16(0,02)	-0,99(1,46)	1,04(0,88)	
pH H2O ₅₂₅		-0,46(0,13)	-0,47(0,12)	
pH H2O ₈₀₁₀₀		8,60(3,01)		
MO ₀₅			-0,34(0,06)	
MO ₈₀₁₀₀			0,04(0,01)	
P ₀₅		0,12(0,02)	0,19(0,02)	
P ₅₂₅ ²			-10,81(2,60)	
P ₅₂₅ ³			13,19(2,61)	
Na ₀₅ ²			-7,00(2,48)	
Na ₀₅ ³		-8,22(2,22)		
K ₀₅		-1,45(0,23)	-1,49(0,23)	
K ₅₂₅		11,41(5,00)		

Tabela 6: Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “de mais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Copaifera langsdorffii* e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>			(conclusão)
	M0 _C	M1 _C	M2 _C	
K ₅₂₅ ²		-7,43(2,86)		
K ₈₀₁₀₀		-11,16(4,71)		
K ₈₀₁₀₀ ²		-15,80(5,29)		
K ₈₀₁₀₀ ³		-12,55(4,77)		
Ca ₀₅		-25,18(7,15)	-25,10(7,40)	
Ca ₀₅ ²		12,71(3,58)	12,44(4,19)	
Ca ₀₅ ³		-5,98(2,66)		
Al ₀₅		9,34(3,47)		
Al ₀₅ ²		-10,63(2,72)	-6,39(2,95)	
CTC ₀₅		0,01(0,00)	0,02(0,00)	
CTC ₅₂₅		-19,94(5,53)		
CTC ₅₂₅ ²		8,89(2,70)	7,36(2,78)	
CTC ₅₂₅ ³		-9,09(2,66)	-14,63(2,81)	
CTC ₈₀₁₀₀		13,17(4,12)		
CTC ₈₀₁₀₀ ²		-14,32(2,90)		
V ₀₅		0,58(0,23)	19,52(7,95)	
V ₀₅ ³			-5,72(2,80)	
V ₅₂₅		-24,10(6,83)	-0,71(0,16)	
V ₅₂₅ ²		6,87(2,85)		
m ₅₂₅		-0,04(0,01)	-33,48(7,32)	
m ₅₂₅ ²			8,00(2,60)	
m ₈₀₁₀₀		-13,57(4,46)	-12,03(4,12)	
m ₈₀₁₀₀ ²		-6,96(2,58)	-6,82(2,56)	
m ₈₀₁₀₀ ³			5,61(2,60)	
s ₀₅		-10,48(2,48)	-9,79(2,65)	
s ₀₅ ²		-5,79(2,53)		
s ₀₅ ³		-9,94(2,47)		
s ₈₀₁₀₀ ²			-8,49(3,29)	
Declividade			10,66(3,96)	
Declividade ³			-8,12(4,02)	
Altitude		-31,96(3,69)	-22,34(4,23)	
Altitude ²		-46,99(3,81)	-45,32(3,85)	
Altitude ³		-20,97(4,05)	-19,33(4,12)	
X1 ₀₅ ³			4,83(2,04)	
X1 ₈₀₁₀₀		0,49(0,22)		
X2 ₈₀₁₀₀			13,26(2,74)	
X2 ₈₀₁₀₀ ²			-6,40(2,72)	
Outras _{S,X} ,			-0,89(0,33)	
AIC	18264,99	17594,42	17545,1	

Os coeficientes dos modelos escolhidos a partir dos dados de ocorrência das

espécies forneceram informações a respeito de como cada covariável espacial afetou na intensidade de ocorrência das espécies. As Tabelas 4, 5 e 6 fornecem as estimativas dos parâmetros associados as covariáveis espaciais para os modelos que apresentaram o melhor ajuste. As covariáveis com coeficiente positivo tiveram contribuição positiva na intensidade de ocorrência das espécies, enquanto aquelas que apresentaram coeficientes negativos tiveram contribuição negativa. Os efeitos da maioria das covariáveis foram efeitos-curvilíneos (quadráticas ou cúbicas). Pode-se utilizar como exemplo a log-intensidade agregada, que é modelada como uma função quadrática em relação à saturação por alumínio (m).

Os modelos ajustados apontaram como características químicas significativas para a estimativa da intensidade as covariáveis MO, P, K, Al, CTC, V e m para as três espécies. As características físicas significativas em todas as espécies foram a declividade, altitude e $X1(\log(\text{areia/argila}))$. Esse resultado confirma que o padrão das árvores referente as três espécies não foi gerado por um processo de Poisson, já que pelo menos um dos β foi diferente de zero, ou seja, as covariáveis explicam de forma significativa a variação espacial da intensidade.

Outro resultado importante é que após ser considerado a influência de outras espécies nos modelos propostos, algumas variáveis que antes eram significativa para o modelo deixaram de ser e, outra que não eram significativas passaram a ser, indicando que a intensidade de espécies nesse caso é fator limitante para o desenvolvimento das espécies em estudo. Isso pode ser explicado a partir da ecologia de cada espécie em estudo, se tratando de espécies pioneiras como *Vochysia tucanorum* e *Xylopia aromatica*, as mesmas são espécies pioneiras, ou seja, espécies que se estabelecem, crescem e se reproduzem somente em clareiras. Por isso a interação com outras espécies foi significativa, considerando que as espécies estudadas podem interagir com um grupo de espécies secundárias que estão diretamente relacionadas a arquitetura do dossel da floresta. Já a espécie *Copaifera langsdorffii*, por ser de secundária tardia a clímax é uma espécie muito mais exigente em seu microábitat e a explicação que pode ser dada é que as espécies pioneiras formam um ambiente mais sombreado pela pioneiras e secundárias iniciais, favorecendo então a aglomeração desses indivíduos em uma determinada região. Esse estudo pode facilitar a escolha de espécies para a restauração ecológica.

Após a escolha dos modelos, foi possível calcular a estimativa da função L não homogênea para as espécies *Copaifera langsdorffii* (Figura 11(a)), *Vochysia tucanorum*

(Figura 11(b)) e *Xylopia aromatica* (Figura 11(c)). Os gráficos mostram que mesmo quando é considerada a dependência das covariáveis espaciais e a intensidade de outras espécies na região, a ocorrência dos mesmos ainda resulta em aglomeração, ou seja, rejeita-se a hipótese de CAE uma vez que a função L não está totalmente dentro dos envelopes de simulação. Esse resultado aponta que as variáveis ajudam a explicar a variação espacial das espécies, mas que não são suficientes para explicar toda a variação, indicando que existe mais algum fator que causa essa variabilidade.

Figura 11 – Estimativas da função L não homogênea (linha cheia) e envelopes de simulação dos modelos com menor valor de AIC referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

Como os modelos ajustados não foram suficientes para explicar todo o padrão de agregação das espécies, foi utilizado o modelo log gaussiano de Cox para explicar o padrão

de aglomeração não explicada pelas covariáveis espaciais incorporando um efeito espacial. Os valores estimados através da aplicação LGCP para os dados de ocorrência de árvores na Estação Ecológica de Assis considerando os modelos com melhor ajuste para cada espécie, podem ser vistos na Tabela 7.

Tabela 7 – Parâmetros estimados a partir do LGCP aplicado aos dados de ocorrência de árvores na Estação Ecológica de Assis

Parâmetro	Espécies		
	<i>Copaifera langsdorffii</i>	<i>Vochysia tucanorum</i>	<i>Xylopia aromatica</i>
σ^2	0,347	0,559	0,462
θ	27,352	22,271	35,218

Após ajustado o LGCP, foram novamente calculadas as estimativas da função L não homogênea para as três espécies. Os gráficos das Figuras 12(a), 12(b) e 12(c) mostram que acrescentando um efeito latente espacial, a função L apresenta-se completamente dentro dos envelopes de simulação indicando um bom ajuste do modelo. Com isso, pode-se concluir que a ocorrência de árvores das espécies em análise são espacialmente agrupadas e que existe uma relação entre os fatores ambientais que é capaz de explicar esse fenômeno. Apesar disso, apenas o uso dessas covariáveis não foi suficiente para explicar todo a variabilidade do processo, sendo necessário incluir um efeito espacial latente, indicando existir uma variação ambiental que é espacialmente estruturada não mensurada no estudo, como por exemplo a luminosidade ou abertura no dossel da floresta.

Figura 12 – Estimativas da função L não homogênea (linha cheia) e envelopes de simulação dos modelos com menor valor de AIC considerando um componente espacial estimado por meio do LGCP referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

4.1.2 Análise Fatorial

A adequação da análise fatorial foi determinada pelos Testes de Esfericidade de Bartlett e Kaiser-Meyer-Olkin (KMO), presentes na Tabela 8. O teste de KMO (0,77) indicou que as variáveis estão correlacionadas e a análise fatorial apresentou um bom nível de adequação aos dados. No teste de Bartlett, o valor p foi aproximadamente zero, indicando que as matrizes de correlação foram significativamente diferentes da matriz identidade. Todas as medidas indicaram que o conjunto de variáveis foi adequado à análise fatorial. Desta forma, foi aplicada a análise fatorial aos dados, utilizando o método das componentes

principais para extração dos fatores via rotação varimax para obter melhores combinações.

Tabela 8 – Medidas de avaliação da Adequação da Análise Fatorial

Estatísticas		Resultado
Teste de Kaiser-Meyer-Olkin (KMO)		0,77
Teste de Esfericidade de Bartlett	Aproximação χ^2	6850,869
	Graus de Liberdade	703,000
	Significância	≈ 0

Fonte: O autor (2013)

Foram extraídos oito fatores principais, que explicam, respectivamente 28,12%, 12,65%, 11,35%, 7,12%, 5,97%, 5,48%, 4,03% e 3,70%, totalizando 78,40% de explicação dos dados originais. Na Tabela 9 estão presentes as communalidades, os carregamentos após a rotação varimax e as variáveis associadas aos seus respectivos fatores. A associação das variáveis com os fatores foi feita por meio do valor absoluto dos carregamentos em cada fator. As variáveis que se correlacionam com cada fator estão destacadas em negrito.

Na Tabela 9 pode-se observar que o primeiro fator possui pesos mais altos nas covariáveis K_{5-25} , K_{80-100} , Ca_{5-25} , CTC_{5-25} , CTC_{80-100} , V_{5-25} , m_{5-25} e m_{80-100} . O segundo fator possui pesos maiores nas covariáveis Ca_{0-5} , Mg_{0-5} , Al_{0-5} , V_{0-5} e m_{0-5} . As covariáveis com maior peso no terceiro fator são areia (0 – 5, 5 – 25 e 80 – 100) e argila (0 – 5, 5 – 25 e 80 – 100). O quarto fator extraído é composto pelas covariáveis MO_{0-5} , P_{0-5} , K_{0-5} e CTC_{0-5} . No quinto fator a covariável s nas profundidades 0 – 5 e 5 – 25 apresentou pesos mais altos. As covariáveis pH H_2O_{0-5} e Al_{5-25} compõem o sexto fator. O sétimo fator é composto apenas pela covariável silt na profundidade 80 – 100. Por fim, o oitavo fator é caracterizado pelas covariáveis pH H_2O_{80-100} e V_{80-100} .

Tabela 9 – Resultado da Análise Fatorial (AF)

Variáveis	Comunalidade	Carreg. 1	Carreg. 2	Carreg. 3	Carreg. 4	Carreg. 5	Carreg. 6	Carreg. 7	Carreg. 8	Fator
1 2 3 4 5 6 7 8 pH H ₂ O 5-25	0,70	-0,03	-0,03	-0,16	0,12	0,05	0,72	-0,03	-0,37	6
pH H ₂ O 80-100	0,72	-0,11	0,00	-0,14	0,03	0,01	0,10	-0,03	-0,82	8
MO 0-5	0,76	-0,00	0,09	-0,30	0,81	-0,03	0,08	-0,04	-0,00	4
P 0-5	0,75	0,00	0,36	0,04	0,75	-0,04	-0,19	-0,07	0,14	4
K 0-5	0,69	-0,36	0,47	0,05	0,58	0,01	0,05	0,01	0,00	4
K 5-25	0,71	-0,76	0,22	-0,05	0,27	0,06	-0,06	-0,01	0,04	1
K 80-100	0,75	-0,81	-0,02	-0,16	0,11	-0,01	-0,12	-0,06	-0,21	1
Ca 0-5	0,85	-0,32	0,69	-0,27	0,35	-0,04	-0,14	0,01	-0,24	2
Ca 5-25	0,85	-0,83	0,03	-0,22	0,08	-0,09	-0,09	-0,00	-0,30	1
Mg 0-5	0,85	0,09	0,82	-0,07	0,39	0,02	-0,01	0,01	0,10	2
Al 0-5	0,71	0,02	-0,73	-0,04	0,34	0,07	-0,17	0,08	-0,10	2
Al 5-25	0,80	-0,05	-0,05	-0,27	0,18	0,08	-0,81	0,02	-0,17	6
CTC 0-5	0,81	-0,29	0,01	-0,21	0,81	-0,05	-0,05	-0,00	-0,11	4
CTC 5-25	0,77	-0,51	-0,05	-0,38	0,40	-0,03	-0,40	-0,08	-0,16	1
CTC 80-100	0,71	-0,72	-0,14	-0,28	0,14	-0,07	-0,14	-0,07	0,19	1
V 0-5	0,91	-0,09	0,94	-0,08	0,05	-0,01	-0,05	0,04	-0,08	2
V 5-25	0,77	-0,79	0,29	0,07	-0,11	-0,05	0,20	0,03	-0,02	1
V 80-100	0,73	-0,43	0,08	0,03	-0,06	0,09	-0,05	0,07	-0,72	8
m 0-5	0,95	0,18	-0,91	0,12	-0,25	0,05	-0,04	0,01	0,06	2
m 5-25	0,90	0,87	-0,23	0,07	-0,05	0,09	-0,28	0,04	0,06	1
m 80-100	0,81	0,79	0,06	0,13	0,02	0,00	0,09	0,03	0,39	1
Sand 0-5	0,77	-0,06	-0,04	0,72	-0,39	-0,00	-0,26	-0,12	0,11	3
Sand 5-25	0,69	0,03	-0,07	0,81	-0,01	-0,06	0,15	-0,07	0,02	3
Sand 80-100	0,79	0,19	-0,05	0,76	0,06	0,12	0,08	-0,38	0,05	3
Silt 80-100	0,97	0,10	-0,02	-0,02	-0,06	0,04	-0,03	0,97	-0,01	7
Clay 0-5	0,76	-0,11	0,08	-0,76	0,34	-0,05	0,15	0,04	-0,15	3
Clay 5-25	0,76	-0,21	0,09	-0,80	0,09	-0,00	-0,15	-0,19	0,03	3
Clay 80-100	0,76	-0,26	0,03	-0,77	0,01	-0,17	-0,06	-0,25	-0,03	3
s 0-5	0,77	0,06	-0,02	0,09	-0,11	0,86	0,07	0,01	0,00	5
s 5-25	0,77	0,05	-0,07	0,04	0,03	0,86	-0,09	0,02	-0,07	5

Fonte: O autor (2013)

Os fatores extraídos pela análise fatorial foram utilizados como covariáveis no modelo log-linear. Assim como na análise com as covariáveis originais, foi aplicada a metodologia *stepwise* para selecionar os fatores mais significativos na regressão log-linear, com e sem o acréscimo da covariável construída por meio do estimador de kernel.

Os coeficientes e valores de AIC dos modelos com melhor ajuste para a espécie *Copaifera langsdorffii* serão apresentados na Tabela 10. Observando os valores de AIC dos modelos avaliados, verificou-se como sendo o modelo com melhor ajuste o modelo $M2_C$, por apresentar o menor AIC entre todos os modelos ajustados. O modelo $M0_C$, que considerou uma média constante em toda a região foi o modelo com menor plausibilidade quando comparado com os modelos $M1_C$ e $M2_C$, confirmando a não homogeneidade espacial na região. O modelo $M2_C$ que considerou tanto os fatores extraídos dos dados originais quanto a covariável construída, foi o modelo mais plausível para estimar a intensidade de ocorrência das espécie *Copaifera langsdorffii*, indicando que fatores expressos pelas covariáveis espaciais do solo são importantes para explicar o padrão espacial dessa espécie.

Tabela 10 – Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Copaifera langsdorffii* considerando como covariáveis os fatores extraídos dos dados originais e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	Modelos		
	$M0_C$	$M1_C$	$M2_C$
Intercepto	-3,35(0,02)	-3,38(0,02)	-3,38(0,02)
Fator1		-5,51(2,13)	-5,34(2,12)
Fator1 ³		7,64(2,36)	7,58(2,34)
Fator2		-19,08(2,48)	-17,51(2,47)
Fator2 ²		-14,70(2,66)	-14,16(2,68)
Fator2 ³		7,20(2,57)	6,21(2,59)
Fator3		-17,44(2,36)	-17,78(2,36)
Fator3 ²		-18,09(2,43)	-17,84(2,41)
Fator5		4,01(2,01)	0,04(0,02)
Fator5 ²		4,79(1,99)	
Fator7		5,95(2,08)	7,16(2,08)
Fator7 ³		-8,65(2,09)	-7,78(2,09)
Fator8 ²		-9,33(2,37)	-9,79(2,37)
Outras _{S.C.}			12,97(2,15)
Outras _{S.C.} ²			-8,80(2,17)
AIC	34597,84	34269,42	34225,24

Por meio da Tabela 11 referente aos coeficientes e valor do AIC para a espécie *Vochysia tucanorum*, é possível observar que o efeito da covariável construída através do

estimador de kernel não foi significativa para o ajuste do modelo, com isso, os modelos $M1_V$ e $M2_V$ são os mesmos. Portanto, deve-se considerar apenas os modelos $M0_V$ e $M1_V$. O modelo denominado suporte (AIC_{min}) para o cálculo da diferença de AIC foi o modelo $M1_V$. A diferença de AIC (Δ_i) entre os entre os modelos $M0_V$ e $M1_V$ foi de 200,11. Assim, o modelo $M1_V$, que considerou o efeito dos fatores principais como covariáveis obteve mais plausibilidade em relação ao modelo $M0_V$ que considerou uma média constante em toda a região.

Tabela 11 – Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Vochysia tucanorum* considerando como covariáveis os fatores extraídos dos dados originais e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>		
	$M0_V$	$M1_V$	$M2_V$
Intercepto	-3,98(0,02)	-4,02(0,02)	-4,02(0,02)
Fator1		8,41(2,38)	8,41(2,38)
Fator1 ²		8,42(2,52)	8,42(2,52)
Fator2		-18,62(2,69)	-18,62(2,69)
Fator2 ²		-7,04(3,13)	-7,04(3,13)
Fator2 ³		-6,16(2,92)	-6,16(2,92)
Fator3		12,58(2,20)	12,58(2,20)
Fator3 ²		13,95(1,97)	13,95(1,97)
Fator4		-0,11(0,02)	-0,11(0,02)
Fator8		-7,77(2,59)	-7,77(2,59)
Fator8 ²		-5,75(2,46)	-5,75(2,46)
AIC	21007,75	20807,64	20807,64

Assim como nas outras espécies foram ajustados modelos considerando a intensidade constante na região, fatores extraídos dos dados originais com e sem adição da intensidade de outras espécies na região para a espécie *Vochysia tucanorum* (Tabela 6). O valor do AIC dos modelos testados apontou como o melhor modelo o $M2_X$, por apresentar o menor AIC entre todos os modelos ajustados. Assim, pode-se concluir que considerando os fatores extraídos das variáveis originais, a variável construída não influencia no padrão de agregação da espécie *Vochysia tucanorum*, diferentemente do que acontece quando consideramos as variáveis originais.

Os coeficientes dos modelos escolhidos a partir dos dados de ocorrência das espécies forneceram informações a respeito de como cada covariável espacial afeta na intensidade de ocorrência das espécies. As Tabelas 10, 11 e 12 fornecem as estimativas dos parâmetros associados aos fatores principais dos modelos que apresentaram o melhor ajuste. Em

Tabela 12 – Comparação dos modelos do processos de Poisson (homogêneo (M0), não homogêneo sem adição da variável “demais espécies” (M1) e não homogêneo com adição da covariável “demais espécies” (M2)) ajustada aos dados de ocorrência de árvores da espécie *Xylopia aromatica* considerando como covariáveis os fatores extraídos dos dados originais e os coeficientes das variáveis preditoras dos modelos selecionados

Covariável	<i>Modelos</i>		
	M0 _X	M1 _X	M2 _X
Intercepto	-4,16(0,02)	-4,22(0,03)	-4,00(0,06)
Fator1		-0,17(0,02)	-0,18(0,02)
Fator2		-20,10(3,13)	-20,70(3,10)
Fator2 ²		-11,26(4,22)	-13,09(4,19)
Fator2 ³		-15,36(3,52)	-15,67(3,53)
Fator3		0,16(0,02)	0,17(0,02)
Fator5		-7,31(2,27)	-7,09(2,24)
Fator5 ²		-5,76(2,31)	-4,53(2,30)
Fator5 ³		-7,55(2,28)	-6,90(2,28)
Fator6		-7,57(2,30)	-0,08(0,03)
Fator6 ²		-5,45(2,49)	
Fator8 ²		5,62(2,11)	
Int _{S,X}			-1,14(0,29)
AIC	18264,99	18081,34	18074,75

todos os modelos selecionados, os fatores 1, 2, 3 e 8 considerados como covariáveis independentes, apresentaram-se como significativos. Como estes fatores são combinações lineares das variáveis originais, pode-se considerar que variáveis K, Al, CTC, V, m, areia e argila são os fatores ambientais que influenciam no padrão espacial das três espécies analisadas.

Após a escolha dos modelos, foi possível calcular a estimativa da função L não homogênea para as espécies *Copaifera langsdorffii* (Figura 13(a)), *Vochysia tucanorum* (Figura 13(b)) e *Xylopia aromatica* (Figura 13(c)). Os gráficos mostram que os modelos ajustados utilizando os fatores principais como covariáveis não foram suficientes para explicar todo o padrão de agregação das espécies, sendo utilizado o modelo Log Gaussiano de Cox para explicar o padrão de aglomeração não explicada por estes fatores incorporando um efeito espacial. Os valores estimados através da aplicação LGCP para os dados de ocorrência de árvores na Estação Ecológica de Assis considerando os modelos com melhor ajuste para cada espécie, podem ser vistos na Tabela 13.

Após ajustado o LGCP, foram novamente calculadas as estimativas da função L não homogênea para as três espécies. Os gráficos das Figuras 14(a), 14(b) e 14(c) mostram que acrescentando um efeito latente espacial, a função L apresenta-se completamente dentro dos envelopes de simulação indicando um bom ajuste do modelo. Com isso, pode-se inferir

Figura 13 – Estimativas da função L não homogênea (linha cheia) e envelopes de simulação utilizando os fatores extraídos das variáveis originais, referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

Tabela 13 – Parâmetros estimados a partir do LGCP aplicado aos dados de ocorrência de árvores na Estação Ecológica de Assis, utilizando os fatores principais extraídos dos dados originais como covariáveis

Parâmetro	Espécies		
	<i>Copaifera langsdorffii</i>	<i>Vochysia tucanorum</i>	<i>Xylopia aromatica</i>
σ^2	0,372	0,563	0,615
θ	35,656	39,589	64,224

que a ocorrência de árvores das espécies em análise são espacialmente agrupadas e que existe uma relação entre os fatores extraídos das variáveis ambientais originais que é capaz de explicar esse fenômeno. No entanto, apenas o uso desses fatores não foi suficiente para explicar toda a variabilidade do processo, sendo necessário incluir um efeito espacial latente,

indicando existir uma variação ambiental que é espacialmente estruturada não mensurada no estudo.

Figura 14 – Estimativas da função L não homogênea (linha cheia) e envelopes de simulação dos modelos com menor valor de AIC utilizando os fatores extraídos das variáveis originais considerando um componente espacial estimado por meio do LGCP referente as três espécies mais abundantes na Estação Ecológica de Assis. (a) *Copaifera langsdorffii*; (b) *Vochysia tucanorum* e (c) *Xylopia aromatica*

4.2 Distribuição espacial do risco

Para o estudo da variação espacial no risco das espécies mais abundantes na região, foram realizadas análises separadas para cada uma das espécies consideradas. Para a seleção dos fatores significativos para o risco de incidência das árvores das espécies, foi

utilizado o método de seleção *stepwise*. Foram ajustado diversos modelos GAM e escolhido o melhor para cada uma das três espécies.

Na Tabela 14 são apresentadas às estimativas dos parâmetros do modelo selecionado para a espécie *Copaifera langsdorffii* controlando-se pelo fator espacial. A estimativa do termo não-paramétrico $g(\mathbf{x})$, que interpretada como superfície de risco estimada controlando-se pelos fatores individuais, está representada na Figura 15. A análise apresentou uma variação espacial significativa (p-valor $< 2e-16$) no risco da presença de árvores da espécie *Copaifera langsdorffii*, onde pode-se verificar regiões com menor propensão de presença de árvores da espécie, condicionado aos fatores ambientais significativos.

Tabela 14 – Estimativas dos efeitos dos fatores ambientais para a *Copaifera langsdorffii*

Variáveis	Estimativa	Erro Padrão	p-valor
Intercepto	-206,83	10,44	0,00
Al ₀₅	-0,86	0,04	0,00
Altitude	0,44	0,02	0,00
MO ₀₅	3,96	0,13	0,00
Na ₀₅	-7,02	0,30	0,00
Al ₈₀₁₀₀	1,52	0,07	0,00
CTC ₀₅	-0,26	0,01	0,00
Silt ₈₀₁₀₀	-2,58	0,18	0,00
K ₅₂₅	2,94	0,12	0,00
T ₈₀₁₀₀	-0,32	0,02	0,00
V ₀₅	-3,60	0,55	0,00
P ₀₅	-0,31	0,05	0,00
K ₈₀₁₀₀	-7,04	0,45	0,00
pH H ₂ O ₅₂₅	-2,60	0,31	0,00
MO ₈₀₁₀₀	-0,18	0,02	0,00
Declividade	-3,19	0,18	0,00
T ₅₂₅	0,16	0,02	0,00
MO ₅₂₅	-1,18	0,11	0,00
m ₈₀₁₀₀	-0,14	0,03	0,00
Ca ₀₅	0,28	0,05	0,00
pH H ₂ O ₈₀₁₀₀	2,43	0,32	0,00
P ₅₂₅	-1,21	0,30	0,00
Al ₅₂₅	-5,26	0,50	0,00
Mg ₀₅	4,08	0,38	0,00
V ₈₀₁₀₀	-0,48	0,05	0,00

Assim como para a espécie *Copaifera langsdorffii*, foram calculadas às estimativas dos parâmetros do modelo selecionado para a espécie *Vochysia tucanorum* (Tabela 15). A estimativa do termo não-paramétrico $g(\mathbf{x})$ para esta espécie (Figura 16), indicou uma variação espacial significativa (p-valor $< 2e-16$) no risco da presença de árvores da espécie *Vochysia tucanorum*, em que também pode-se verificar regiões com menor propensão de presença de árvores da espécie, condicionado aos fatores ambientais significativos. Esta

Figura 15 – Mapa de risco para a propensão de presença das espécie *Copaifera langsdorffii*, controlando para fatores ambientais

regiões apresentam-se parte centro-oeste e noroeste, áreas também consideradas de baixo risco de propensão para a espécie *Copaifera langsdorffii*.

Figura 16 – Mapa de risco para a propensão de presença das espécie *Vochysia tucanorum*, controlando para fatores ambientais

Tabela 15 – Estimativas dos efeitos dos fatores ambientais para a *Vochysia tucanorum*

Variáveis	Estimativa	Erro Padrão	p-valor
Intercepto	-360,76	39,89	0,00
MO ₀₅	7,53	0,39	0,00
pH H ₂ O ₈₀₁₀₀	-14,83	1,25	0,00
Altitude	1,20	0,08	0,00
pH H ₂ O ₀₅	-15,61	0,93	0,00
Declividade	-8,33	0,61	0,00
CTC ₀₅	-0,49	0,03	0,00
K ₅₂₅	7,74	0,48	0,00
Al ₈₀₁₀₀	2,70	0,21	0,00
MO ₈₀₁₀₀	-0,57	0,06	0,00
CTC ₅₂₅	0,16	0,03	0,00
CTC ₈₀₁₀₀	-0,59	0,07	0,00
V ₀₅	-7,14	1,40	0,00
m ₀₅	-1,20	0,06	0,00
Ca ₀₅	-2,97	0,24	0,00
MO ₅₂₅	-0,91	0,31	0,00
V ₅₂₅	-9,87	0,67	0,00
V ₈₀₁₀₀	-2,49	0,17	0,00
Na ₀₅	-6,90	0,87	0,00
pH H ₂ O ₅₂₅	4,36	0,81	0,00
Silt ₀₅	2,63	0,51	0,00
m ₈₀₁₀₀	-0,92	0,09	0,00
Sand ₈₀₁₀₀	11,17	1,36	0,00
K ₀₅	-2,78	1,30	0,03

Na Tabela 16 são apresentadas às estimativas dos parâmetros do modelo selecionado para a espécie *Xylopia aromatica* controlando-se pelo fator espacial. A estimativa do termo não-paramétrico $g(\mathbf{x})$, que interpretada como superfície de risco estimada controlando-se pelos fatores individuais, está representada na Figura 17. A análise também apresentou uma variação espacial significativa ($p\text{-valor} < 2e-16$) no risco da presença de árvores da espécie *Copaifera langsdorffii*. Para esta espécie, pode-se verificar um baixa propensão para a presença de árvores, condicionado aos fatores ambientais significativos na região extremo noroeste.

Tabela 16 – Estimativas dos efeitos dos fatores ambientais para a *Vochysia tucanorum*

Variáveis	Estimativa	Erro Padrão	p-valor
Intercepto	-818,26	50,88	0,00
MO ₀₅	5,80	0,36	0,00
Altitude	1,79	0,10	0,00
pH H ₂ O ₀₅	-11,40	0,74	0,00
Declividade	-7,17	0,56	0,00
MO ₅₂₅	-1,58	0,23	0,00
CTC ₀₅	-0,42	0,03	0,00
m ₀₅	-1,11	0,06	0,00
V ₀₅	-13,52	1,97	0,00
Ca ₀₅	-2,84	0,29	0,00
Al ₈₀₁₀₀	0,27	0,11	0,01
CTC ₈₀₁₀₀	0,16	0,04	0,00
K ₅₂₅	7,11	0,53	0,00
V ₅₂₅	-3,15	0,45	0,00
Mg ₀₅	4,79	0,97	0,00
Na ₀₅	-9,36	0,80	0,00
pH H ₂ O ₅₂₅	-5,15	0,66	0,00

Figura 17 – Mapa de risco para a propensão de presença das espécie *Xylpia aromatica*, controlando para fatores ambientais

5 CONCLUSÕES

O presente trabalho avaliou dados obtidos no projeto “Diversidade, dinâmicas e conservação em florestas do Estado de São Paulo: 40 ha de parcelas permanentes”, do programa Biota da FAPESP, na Estação Ecológica de Assis.

Foram ajustados diferentes modelos (Poisson Homogêneo, Poisson não Homogêneo e o Log Gaussiano), com a finalidade de identificar as variáveis que influenciavam no padrão espacial. Pelo critério de Akaike, foi possível observar que os modelos do Processo Poisson não homogêneo, que consideram os fatores ambientais como covariáveis, apresentaram um melhor ajuste, quando comparados com os modelos do Processo Poisson homogêneo, que por sua vez, considera que a intensidade com a qual as espécies ocorrem na área é constante. Esse resultado indica que os fatores ambientais selecionados apresentam efeito de interação, no padrão espacial das espécies. Porém, há evidências de que os ajustes não foram bons, pois a função L não homogênea para os modelos selecionados, apresenta valores fora do envelope de simulação.

Por meio do ajuste do modelo log gaussiano de Cox e realização do diagnóstico por meio da função L, em que o mesmo apresentou-se dentro dos envelopes de simulação, pode-se inferir que os fatores ambientais incluídos no modelo explicam o padrão espacial das espécies. Porém, como o modelo log gaussiano tem melhor ajuste, há evidências de que existe um processo espacialmente estruturado que não foi considerado na análise mas que é importante para explicar toda a variabilidade dos dados. Assim, neste estudo, foi possível fornecer evidência estatística de forte associação entre as localizações das árvores e os fatores ambientais.

É necessário aprofundar as análises de processos pontuais na área de estudo, pois ainda é preciso considerar outras variáveis ambientais não analisadas, bem como interação entre elas. Além disso, deve-se melhorar a coleta dos dados, pois isso pode evidenciar agrupamentos e a intensidade das relações ecológicas que não foram mostrados com os dados disponíveis.

A variação espacial do risco na região para a presença das espécies mais abundantes evidenciou-se significativa no espaço controlando por fatores ambientais, assim, identificando regiões com menor propensão para a presença das árvores.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on automatic control**, Piscataway, v. 19, n. 6, p. 716–723, 1974.
- ANJOS, A.; MAZZA, M. C. M.; SANTOS, A. C. M. C.; DELFINI, L. T. Análise do padrão de distribuição espacial da araucária (*Araucaria angustifolia*) em algumas áreas no estado do paraná, utilizando a função k de ripley. **Scientia Forestalis**, Piracicaba, n. 66, p. 38–45, 2004.
- BADDELEY, A.; TURNER, R. Spatstat: an R package for analyzing spatial point patterns. **Journal of Statistical Software**, Los Angeles, v. 12, n. 6, p. 1–42, 2005.
- BADDELEY, A. J.; MØLLER, J.; WAAGEPETERSEN, R. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. **Statistica Neerlandica**, Hoboken, v. 54, n. 3, p. 329–350, 2000.
- BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. New York: Longman Scientific & Technical, 1995. 413 p.
- BATISTA, J. L. F.; MAGUIRE, D. A. Modeling the spatial structure of tropical forests. **Forest Ecology and Management**, Amsterdam, v. 110, n. 1, p. 293–314, 1998.
- BERG, E. van den; SANTOS, F. A. M. dos. Aspectos da variação ambiental em uma floresta de galeria em itutinga, mg, brasil. **Ciência Florestal**, Santa Maria, v. 13, n. 2, p. 83–98, 2003.
- BERMAN, M.; DIGGLE, P. J. Estimating weighted integrals of the second-order intensity of a spatial point process. **Journal of the Royal Statistical Society. Series B (Methodological)**, Hoboken, v. 51, n. 1, p. 91–92, 1989.
- BESAG, J.; DIGGLE, P. J. Simple monte carlo tests for spatial pattern. **Journal of the Royal Statistical Society Series C - Applied Statistics**, Hoboken, p. 327–333, 1977.
- BIVAND, R. S.; PEBESMA, E. J.; GÓMES-RUBIO, V. **Applied Spatial Data Analysis with R**. New York: Springer, 2008.
- BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multi-model inference: a practical information-theoretic approach**. 2nd ed. New York: Springer, 2002. 350 p.
- CÂMARA, G.; DRUCK, S.; CARVALHO, M. S.; MONTEIRO, A. V. M. **Análise Espacial de Dados Geográficos**. Planaltina: EMBRAPA, 2004. 208 p.
- CÂMARA, G.; MONTEIRO, A. M.; DRUCK, S.; CARVALHO, M. S. **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2002. 186 p.
- CAPRETZ, R. L.; BATISTA, J. L. F.; SOTOMAYOR, J. F. M.; CUNHA, C. R.; NICOLETTI, M. F.; RODRIGUES, R. R. Padrão espacial de quatro formações florestais do estado de São Paulo, através da função K. **Ciência Florestal**, Santa Maria, v. 22, n. 3, p. 551–565, 2012.

CARAM, C.; Monteiro-de-Castro, M. S.; CAIAFFA, W. T.; OLIVEIRA, C. L.; PROIETTI, A. B. F. C.; ALMEIDA, M. C. M.; BRENER, S.; PROIETTI, F. A. Distribuição espaço-temporal dos candidatos à doação de sangue da fundação hemominas, belo horizonte, minas gerais, brasil, nos anos de 1994 e 2004. **Cad Saúde Pública** [online]., Rio de Janeiro, v. 26, n. 2, p. 229–239, 2006.

CLEMENTE, M. S. **Análise da variação genética de populações de *Vochysia pyramidalis* e *V. tucanorum* por AFLP e da composição de ácidos graxos de sementes**. 2010. 98 p. Dissertação (Mestrado em Botânica) — Universidade de São Paulo, São Paulo, 2010.

CONDIT, R.; ASHTON, P. S.; BAKER, P.; BUNYAVEJCHEWIN, S.; GUNATILLEKE, S.; GUNATILLEKE, N.; HUBBELL, S. P.; FOSTER, R. B.; ITOH, A.; LAFRANKIE, J. V.; LEE, H. S.; LOSOS, E.; MANOKARAN, N.; SUKUMAR, R.; YAMAKURA, T. Spatial patterns in the distribution of tropical tree species. **Science**, Washington, v. 288, n. 5470, p. 1414–1418, 2000.

CRESSIE, N. A. C. **Statistics for spatial data**. New York: Wiley, John & Sons, Incorporated, 1993. 900 p.

CUNHA, U. S. **Análise da estrutura espacial horizontal de uma floresta de terra firme da Amazônia**. 2003. 126 p. Tese (Doutorado em Engenharia Florestal) — Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2003.

DIGGLE, P. J. A kernel method for smoothing point process data. **Journal of the Royal Statistical Society Series C- Applied Statistics**, Hoboken, v. 34, n. 2, p. 138–174, 1985.

DIGGLE, P. J. **Statistical Analysis of Spatial Point Patterns**. 2nd ed. London: Hodder Arnold Publishers, 2003. 159 p.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 2. ed. Rio de Janeiro: Embrapa, 2006. 306 p.

GANDOLFI, S.; LEITÃO, H. F.; BEZERRA, C. L. F. Levantamento florístico e caráter sucessional das espécies arbustivo-arbóreas de uma floresta mesófila semidecídua no município de guarulhos, SP. **Revista brasileira de biologia**, São Carlos, v. 55, n. 4, p. 753–767, 1995.

HAY, J. D.; BIZERRIL, M. X.; CALOURO, A. M.; COSTA, E. M. N.; FERREIRA, A. A.; GASTAL, M. L. A.; JR, C. D. Goes; MANZAN, D. J.; MARTINS, C. R.; MONTEIRO, J. M. G.; OLIVEIRA, S. A.; RODRIGUES, M. C. M.; SEYFFARTH, J. A. S.; WALTER, B. M. T. Comparação do padrão da distribuição espacial em escalas diferentes de espécies nativas do cerrado, em Brasília, DF. **Revista Brasileira de Botânica**, São Paulo, v. 23, n. 3, p. 341–347, 2000.

HE, F. L.; LEGENDRE, P.; LAFRANKIE, J. V. Distribution patterns of tree species in a Malaysian tropical rain forest. **Journal of Vegetation Science**, Hoboken, v. 8, n. 1, p. 105–114, 1997.

ILLIAN, J.; PENTTINEN, A.; STOYAN, H.; STOYAN, D. **Statistical analysis and modelling of spatial point patterns**. New York: John Wiley & Sons, 2008. 534 p.

- KAWAMOTO, M. T. **Análise de técnicas de distribuição espacial com padrões pontuais e aplicação a dados de acidentes de trânsito e a dados de dengue de Rio Claro - SP**. 2012. 53 p. Dissertação (Mestrado em Biometria) — Universidade Estadual Paulista, Botucatu, 2012.
- KELSALL, J. E.; DIGGLE, P. J. Kernel estimation of relative risk. **Bernoulli**, New York, v. 1, p. 3–16, 1995.
- KELSALL, J. E.; DIGGLE, P. J. Spatial variation in risk of disease: A nonparametric binary regression approach. **Journal of the Royal Statistical Society Series C - Applied Statistics**, Hoboken, v. 47, p. 559–573, 1998.
- KORMAN, V. **Fatores abióticos definidores da distribuição de diferentes tipos florestais (florestas paludíca, floresta estacional semidecídua ribeirinha e cerradão), nos municípios de Batatais e Restinga, SP**. 2008. 136 p. Tese (Doutorado Ecologia Aplicada) — Escola Superior de Agricultura “Luiz de Queiroz”. Universidade de São Paulo, Piracicaba, 2008.
- LAWSON, A. B.; WILLIAMS, F. L. R. Applications of extraction mapping in environmental epidemiology. **Statistics in Medicine**, Hoboken, v. 12, p. 1249–1258, 1993.
- LEVIN, S. A. The problem of pattern and scale in ecology. **Ecology**, Washington, v. 73, n. 6, p. 406–413, 1992.
- LEWIN-KOH, N. J.; BIVAND, R.; PEBESMA, E. J.; ARCHER, E.; A., Baddeley. **maptools: tools for reading and handling spatial objects**. [S.l.], 2009. R package version 0.7-26. Disponível em: <<http://CRAN.R-project.org/package=maptools>>. Acesso em: 05 mar. 2013.
- LOPES, S. B. **Efeitos da dependência espacial em modelos de previsão de demanda por transporte**. 2005. 137 p. Dissertação (Mestrado em Engenharia Civil) — Escola de Engenharia de São Carlos. Universidade de São Paulo, São Carlos, 2005.
- MARONI, B. C.; STACI, L. C.; MACHADO, S. R. **Plantas medicinais do cerrado de Botucatu**. São Paulo: UNESP, 2006. 312 p.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005. 295 p.
- MOEUR, M. Characterizing spatial patterns of trees using stem-mapped data. **Forest science**, Bethesda, v. 39, n. 4, p. 756–775, 1993.
- MØLLER, J.; SYVERSVEEN, A. R.; WAAGEPETERSEN, R. P. Log gaussian cox processes. **Scandinavian Journal of Statistics**, Hoboken, v. 25, n. 3, p. 451–482, 1998.
- NEWBERY, D. McC.; RENSHAW, E.; BRÜNIG, E. F. Spatial pattern of trees in kerangas forest, sarawak. **Plant Ecology**, Dordrecht, v. 65, n. 2, p. 77–89, 1986.
- NUNES, A. R. D. T. **Modelação espacial de acidentes rodoviários na cidade de Lisboa**. 2011. 158 p. Dissertação (Mestrado em Matemática e Aplicações) — Universidade Nova de Lisboa, Lisboa, 2011.

- OLINDA, R. A.; SCALON, J. D. Monte Carlo methods for analysis of marked point processes. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 1, p. 39–56, 2010.
- PIERI, F.A.; MUSSI, M. C. M.; MOREIRA, M. A. S. Óleo de copaiba (*Copaifera sp.*): histórico, extração, aplicações industriais e propriedades medicinais. **Revista Brasileira de Plantas Mediciniais**, Botucatu, v. 11, n. 4, p. 465–472, 2009.
- PLOTKIN, J. B.; CHAVE, J.; ASHTON, P. S. Cluster analysis of spatial patterns in Malaysian tree species. **The American Naturalist**, Chicago, v. 160, n. 5, p. 629–644, 2002.
- R Development Core Team. **R**: A language and environment for statistical computing. Vienna, Austria, 2008. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 18 ago. 2013.
- REIS, A. **Dispersão de sementes de Euterpe edulis Martius. (Plamae) em uma floresta ombrófila densa montana da encosta atlântica em Blumenau, SC.** 1995. 154 p. Tese (Doutorado em Biologia Vegetal) — Universidade de Campinas, Campinas, 1995.
- REIS, E. A. **Modelos Dinâmicos Bayesianos para Processos Pontuais Espaço-Temporais.** 2008. 127 p. Tese (Doutorado em Estatística) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- REIS, M.S. **Distribuição e dinâmica da variabilidade genética em populações naturais de palmitreiro (Euterpe edulis).** 1996. 210 p. Tese (Doutorado em Agronomia/Genética e Melhoramento de Plantas) — Escola Superior de Agricultura “Luiz de Queiroz”. Universidade de São Paulo., Piracicaba, 1996.
- RIPLEY, B. D. Modelling spatial patterns. **Journal of the Royal Statistical Society. Series B (Methodological)**, Hoboken, v. 39, n. 2, p. 172–212, 1977.
- RIPLEY, B. D. **Spatial Statistics.** New York: John Wiley & Sons, 1981. 183 p.
- ROWLINGSON, B.; DIGGLE, P. **splancs: Spatial and space-time point pattern analysis.** Lancaster, 2013. R package version 2.01-34. Disponível em: <<http://CRAN.R-project.org/package=splancs>>. Acesso em: 05 mar. 2013.
- SALIS, S M.; SHEPHERD, G. J.; JOLY, C. A. Floristic comparison of mesophytic semideciduous forests of the interior of the state of são paulo, southeast brazil. **Plant Ecology**, Dordrecht, v. 119, n. 2, p. 155–164, 1995.
- SHIMAKURA, S. E.; CARVALHO, M. S.; AERTS, D. R. G. C.; FLORES, R. Distribuição espacial do risco: modelagem da mortalidade infantil em porto alegre, rio grande do sul. **Cadernos de Saude Pública**, Rio de Janeiro, v. 17, n. 5, p. 1251–1261, 2001.
- SILVERMAN, B. W. **Density Estimation for Statistics and Data Analysis.** London: Chapman & Hall, 1986. 176 p.
- SOUSA, J. P. B.; NANAYAKKARA, D.; SILVA, A. A. B.; BASTOS, J. K. Leishmanicidal and antimalarial activities of crude extracts from aerial parts of copaifera langsdorffii and isolation of secondary metabolites. **Journal of Pharmacy Research**, Mohali, v. 5, n. 8, p. 4103–4107, 2012.

TORRES, R. B.; MARTINS, F. R.; KINOSHITA, L. S. Climate, soil and tree flora relationships in forests in the state of são paulo, southeastern brasil. **Revista brasileira de botânica**, São Paulo, v. 20, n. 1, p. 41–49, 1997.

VEIGA, V. F; PINTO, A. C. O gênero *Copaifera l.* **Química Nova**, São Paulo, v. 25, n. 2, p. 273–286, 2002.

APÊNDICE

APÊNDICE A - Script do *software* R

```
##### Pacotes Necessários para análise #####
require(MASS);require(rpart);require(splancs);require(akima)
require(spatstat);require(sp);require(maptools);require(vegan)
require(splines);require(xtable)

##### Selecionando as covariável #####

dados      <- read.table("solo.txt", h=T)
coordenadas <- dados[,2:3]
tradagem    <- data.frame(dados[,5:ncol(dados)])+0.0000001

#===== Transformação razão log-aditiva

X105      <- log(trad["Sand_0_5_Porc"]/trad["Clay_0_5_Porc"])
X205      <- log(trad["Silt_0_5_Porc"]/trad["Clay_0_5_Porc"])
X1525     <- log(trad["Sand_5_25_Porc"]/trad["Clay_5_25_Porc"])
X2525     <- log(trad["Silt_5_25_Porc"]/trad["Clay_5_25_Porc"])
X180100   <- log(trad["Sand_80_100_Porc"]/trad["Clay_80_100_Porc"])
X280100   <- log(trad["Silt_80_100_Porc"]/trad["Clay_80_100_Porc"])
trad      <- trad[,-(30:38)]
trad[35:40] <- c(X105,X1525,X180100,X205,X2525,X280100)

#===== Interpolando nova covariável

sem_cop    <- read.table("especies_sem_Copaifera.txt", h=T)
sem_voc    <- read.table("especies_sem_vochysia.txt", h=T)
sem_xyl    <- read.table("especies_sem_xylophia.txt", h=T)

w1         <- owin(c(range(sem_cop$x)[1]-15,range(sem_cop$x)[2]+15),
                  c(range(sem_cop$y)[1]-15,range(sem_cop$y)[2]+15))
w2         <- owin(c(range(sem_voc$x)[1]-15,range(sem_voc$x)[2]+15),
                  c(range(sem_voc$y)[1]-15,range(sem_voc$y)[2]+15))
w3         <- owin(c(range(sem_xyl$x)[1]-15,range(sem_xyl$x)[2]+15),
                  c(range(sem_xyl$y)[1]-15,range(sem_xyl$y)[2]+15))

pts_sem_cop <- ppp(sem_cop$x, sem_cop$y, window=w1)
pts_sem_voc <- ppp(sem_voc$x, sem_voc$y, window=w2)
pts_sem_xyl <- ppp(sem_xyl$x, sem_xyl$y, window=w3)

summary(pts_sem_cop);summary(pts_sem_voc);summary(pts_sem_xyl)

caixa      <- rbind(c(564001.89-10,7500797.80-10),
                  c(564313.11+10,7500850.77-10),
                  c(564257.44+10,7501162.24+10),
                  c(563948.11-15,7501108.99+10),
                  c(564001.89-10,7500797.80-10))

semcop.spp <- as.points(sem_cop)
semvoc.spp <- as.points(sem_voc)
semxyl.spp <- as.points(sem_xyl)

bandaest1  <- mse2d(semcop.spp,caixa,nsmse=50,range=80)
bandaest2  <- mse2d(semvoc.spp,caixa,nsmse=50,range=80)
```

```

bandaest3      <- mse2d(semXyl.spp,caixa,nsmse=50,range=80)

range(bandaest1$mse); bandaestimada1 # h = 3.20
range(bandaest2$mse); bandaestimada2 # h = 3.20
range(bandaest3$mse); bandaestimada3 # h = 3.20

kernel1      <- kernel2d(semcop.spp, caixa, h0=3.2, nx=16, ny=16)
kernel2      <- kernel2d(semvoc.spp, caixa, h0=3.2, nx=16, ny=16)
kernel3      <- kernel2d(semxyl.spp, caixa, h0=3.2, nx=16, ny=16)

int_sem_cop   <- as.vector(kernel1$z)
int_sem_voc   <- as.vector(kernel2$z)
int_sem_xyl   <- as.vector(kernel3$z)

#===== Interpolando para as coordenadas dos dados

cov           <- as.vector(trad[,2:ncol(trad)])
data1         <- read.table("Copaifera.txt", h=T)
data2         <- read.table("Vochysia.txt", h=T)
data3         <- read.table("Xylopia.txt", h=T)

w             <- owin(poly=list(x=c(564001.89-10,564313.11+10,
                                   564257.44+10,563948.11-15),
                                y=c(7500797.80-10,7500850.77-10,
                                   7501162.24+10,7501108.99+10)))

pts_cop       <- ppp(data1$x, data1$y, window=w)
pts_voc       <- ppp(data2$x, data2$y, window=w)
pts_xyl       <- ppp(data3$x, data3$y, window=w)

summary(pts_cop); summary(pts_voc); summary(pts_xyl)

plot(pts_cop,main='Pontos Copaifera langsdorffii',pch=16,cex=0.6)
plot(pts_voc,main='Pontos Vochysia ',pch=16,cex=0.6)
plot(pts_xyl,main='Pontos Xylopia aromatica',pch=16,cex=0.6)

grid          <- expand.grid(seq(min(range(sem_cop$x)[1]-1),
                                max(range(sem_cop$x)[1]+1),l=16),
                             seq(min(range(sem_cop$x)[1]-1),
                                max(range(sem_cop$x)[1]+1),l=16))
dimnames(grid)[2] <- list(c("x","y"))

covs          <- cbind(grid,cov,int_sem_cop,int_sem_voc,int_sem_xyl)
grid.covs     <- covs
coordinates(grid.covs) <- c("x","y")
class(grid.covs)

grid.covs1    <- as(grid.covs, "SpatialPixelsDataFrame")
class(grid.covs1)

pH_H2O_0_5    <- as.im.SpatialGridDataFrame(grid.covs1[,1])
pH_H2O_5_25   <- as.im.SpatialGridDataFrame(grid.covs1[,2])
pH_H2O_80_100 <- as.im.SpatialGridDataFrame(grid.covs1[,3])
MO_0_5        <- as.im.SpatialGridDataFrame(grid.covs1[,4])
MO_5_25       <- as.im.SpatialGridDataFrame(grid.covs1[,5])
MO_80_100     <- as.im.SpatialGridDataFrame(grid.covs1[,6])
P_0_5        <- as.im.SpatialGridDataFrame(grid.covs1[,7])
P_5_25       <- as.im.SpatialGridDataFrame(grid.covs1[,8])

```

```

P_80_100      <- as.im.SpatialGridDataFrame(grid.covs1[,9])
Na_0_5        <- as.im.SpatialGridDataFrame(grid.covs1[,10])
Na_5_25       <- as.im.SpatialGridDataFrame(grid.covs1[,11])
Na_80_100     <- as.im.SpatialGridDataFrame(grid.covs1[,12])
K_0_5         <- as.im.SpatialGridDataFrame(grid.covs1[,13])
K_5_25        <- as.im.SpatialGridDataFrame(grid.covs1[,14])
K_80_100      <- as.im.SpatialGridDataFrame(grid.covs1[,15])
Ca_0_5        <- as.im.SpatialGridDataFrame(grid.covs1[,16])
Ca_5_25       <- as.im.SpatialGridDataFrame(grid.covs1[,17])
Ca_80_100     <- as.im.SpatialGridDataFrame(grid.covs1[,18])
Mg_0_5        <- as.im.SpatialGridDataFrame(grid.covs1[,19])
Mg_5_25       <- as.im.SpatialGridDataFrame(grid.covs1[,20])
Mg_80_100     <- as.im.SpatialGridDataFrame(grid.covs1[,21])
Al_0_5        <- as.im.SpatialGridDataFrame(grid.covs1[,22])
Al_5_25       <- as.im.SpatialGridDataFrame(grid.covs1[,23])
Al_80_100     <- as.im.SpatialGridDataFrame(grid.covs1[,24])
CTC_0_5       <- as.im.SpatialGridDataFrame(grid.covs1[,25])
CTC_5_25      <- as.im.SpatialGridDataFrame(grid.covs1[,26])
CTC_80_100    <- as.im.SpatialGridDataFrame(grid.covs1[,27])
V_0_5_Porc    <- as.im.SpatialGridDataFrame(grid.covs1[,28])
V_5_25_Porc   <- as.im.SpatialGridDataFrame(grid.covs1[,29])
V_80_100_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,30])
m_0_5_Porc    <- as.im.SpatialGridDataFrame(grid.covs1[,31])
m_5_25_Porc   <- as.im.SpatialGridDataFrame(grid.covs1[,32])
m_80_100_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,33])
S_0_5         <- as.im.SpatialGridDataFrame(grid.covs1[,34])
S_5_25        <- as.im.SpatialGridDataFrame(grid.covs1[,35])
S_80_100      <- as.im.SpatialGridDataFrame(grid.covs1[,36])
Declividade   <- as.im.SpatialGridDataFrame(grid.covs1[,37])
Altitude      <- as.im.SpatialGridDataFrame(grid.covs1[,38])
Sand_0_5_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,39])
Sand_5_25_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,40])
Sand_80_100_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,41])
Silt_0_5_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,42])
Silt_5_25_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,43])
Silt_80_100_Porc <- as.im.SpatialGridDataFrame(grid.covs1[,44])
Int_sem_cop   <- as.im.SpatialGridDataFrame(grid.covs1[,45])
Int_sem_voc   <- as.im.SpatialGridDataFrame(grid.covs1[,46])
Int_sem_xyl   <- as.im.SpatialGridDataFrame(grid.covs1[,47])

## Análise exploratória para verificar a heterogeneidade espacial

plot(density(pts_cop, sigma=bw.diggle(pts_cop)),main="")
points(pts_cop,cex=0.3)
plot(density(pts_voc, sigma=bw.diggle(pts_voc)),main="")
points(pts_voc,cex=0.3)
plot(density(pts_xyl, sigma=bw.diggle(pts_xyl)),main="")
points(pts_xyl,cex=0.3)

plot(envelope(pts_cop, Lest, nsim = 1000,correction="border"),.r ~ r,
      legend="",main="",cex.axis=1.3,cex.lab=1.8,lwd=2)
plot(envelope(pts_voc, Lest, nsim = 1000,correction="border"),.r ~ r,
      legend="",main="",cex.axis=1.3,cex.lab=1.8,lwd=2)
plot(envelope(pts_xyl, Lest, nsim = 1000,correction="border"),.r ~ r,
      legend="",main="",cex.axis=1.3,cex.lab=1.8,lwd=2)

# modelos nulos

```

```
fit.null_cop      <- ppm(pts_cop, ~1, Poisson())
fit.null_voc      <- ppm(pts_voc, ~1, Poisson())
fit.null_xyl      <- ppm(pts_xyl, ~1, Poisson())
```

Seleção de variáveis Stepwise sem variável construída

```
fit.com_cop <- ppm(pts_cop, ~ pH_H2O_0_5 + pH_H2O_5_25 + pH_H2O_80_100 +
MO_0_5 + MO_5_25 + MO_80_100 + P_0_5 + P_5_25 + P_80_100 + Na_0_5 +
Na_5_25 + Na_80_100 + K_0_5 + K_5_25 + K_80_100 + Ca_0_5 + Ca_5_25 +
Ca_80_100 + Mg_0_5 + Mg_5_25 + Mg_80_100 + Al_0_5 + Al_5_25 + Al_80_100 +
CTC_0_5 + CTC_5_25 + CTC_80_100 + V_0_5_Porc + V_5_25_Porc + V_80_100_Porc +
m_0_5_Porc + m_5_25_Porc + m_80_100_Porc + S_0_5 + S_5_25 + S_80_100 +
Declividade + Altitude + Sand_0_5_Porc + Sand_5_25_Porc + Sand_80_100_Porc +
Silt_0_5_Porc + Silt_5_25_Porc + Silt_80_100_Porc, covariates = list(
pH_H2O_0_5 = pH_H2O_0_5, pH_H2O_5_25 = pH_H2O_5_25, pH_H2O_80_100 =
pH_H2O_80_100, MO_0_5 = MO_0_5, MO_5_25 = MO_5_25, MO_80_100 = MO_80_100,
P_0_5 = P_0_5, P_5_25 = P_5_25, P_80_100 = P_80_100, Na_0_5 = Na_0_5,
Na_5_25 = Na_5_25, Na_80_100 = Na_80_100, K_0_5 = K_0_5, K_5_25 = K_5_25,
K_80_100 = K_80_100, Ca_0_5 = Ca_0_5, Ca_5_25 = Ca_5_25, Ca_80_100 =
Ca_80_100, Mg_0_5 = Mg_0_5, Mg_5_25 = Mg_5_25, Al_0_5 = Al_0_5, Al_5_25 =
Al_5_25, Al_80_100 = Al_80_100, CTC_0_5 = CTC_0_5, CTC_5_25 = CTC_5_25,
CTC_80_100 = CTC_80_100, V_0_5_Porc = V_0_5_Porc, V_5_25_Porc = V_5_25_Porc,
V_80_100_Porc = V_80_100_Porc, m_0_5_Porc = m_0_5_Porc, m_5_25_Porc =
m_5_25_Porc, m_80_100_Porc = m_80_100_Porc, S_0_5 = S_0_5, S_5_25 =
S_5_25, S_80_100 = S_80_100, Declividade = Declividade, Altitude = Altitude,
Sand_0_5_Porc = Sand_0_5_Porc, Sand_5_25_Porc = Sand_5_25_Porc,
Sand_80_100_Porc = Sand_80_100_Porc, Silt_0_5_Porc = Silt_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Silt_80_100_Porc = Silt_80_100_Porc),
Poisson())
```

```
fit.com_voc <- ppm(pts_voc, ~ pH_H2O_0_5 + pH_H2O_5_25 + pH_H2O_80_100 +
MO_0_5 + MO_5_25 + MO_80_100 + P_0_5 + P_5_25 + P_80_100 +
Na_0_5 + Na_5_25 + Na_80_100 + K_0_5 + K_5_25 + K_80_100 +
Ca_0_5 + Ca_5_25 + Ca_80_100 + Mg_0_5 + Mg_5_25 + Mg_80_100 +
Al_0_5 + Al_5_25 + Al_80_100 + CTC_0_5 + CTC_5_25 +
CTC_80_100 + V_0_5_Porc + V_5_25_Porc + V_80_100_Porc +
m_0_5_Porc + m_5_25_Porc + m_80_100_Porc + S_0_5 + S_5_25 +
S_80_100 + Declividade + Altitude + Sand_0_5_Porc +
Sand_5_25_Porc + Sand_80_100_Porc + Silt_0_5_Porc +
Silt_5_25_Porc + Silt_80_100_Porc, covariates = list(
pH_H2O_0_5 = pH_H2O_0_5, pH_H2O_5_25 = pH_H2O_5_25,
pH_H2O_80_100 = pH_H2O_80_100, MO_0_5 = MO_0_5, MO_5_25 =
MO_5_25, MO_80_100 = MO_80_100, P_0_5 = P_0_5, P_5_25 =
P_5_25, P_80_100 = P_80_100, Na_0_5 = Na_0_5, Na_5_25 =
Na_5_25, Na_80_100 = Na_80_100, K_0_5 = K_0_5, K_5_25 =
K_5_25, K_80_100 = K_80_100, Ca_0_5 = Ca_0_5, Ca_5_25 =
Ca_5_25, Ca_80_100 = Ca_80_100, Mg_0_5 = Mg_0_5, Mg_5_25 =
Mg_5_25, Al_0_5 = Al_0_5, Al_5_25 = Al_5_25, Al_80_100 =
Al_80_100, CTC_0_5 = CTC_0_5, CTC_5_25 = CTC_5_25,
CTC_80_100 = CTC_80_100, V_0_5_Porc = V_0_5_Porc,
V_5_25_Porc = V_5_25_Porc, V_80_100_Porc = V_80_100_Porc,
m_0_5_Porc = m_0_5_Porc, m_5_25_Porc = m_5_25_Porc,
m_80_100_Porc = m_80_100_Porc, S_0_5 = S_0_5, S_5_25 =
S_5_25, S_80_100 = S_80_100, Declividade = Declividade,
Altitude = Altitude, Sand_0_5_Porc = Sand_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Sand_80_100_Porc =
Sand_80_100_Porc, Silt_0_5_Porc = Silt_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Silt_80_100_Porc =
```

```

Silt_80_100_Porc),Poisson())

fit.com_xyl <- ppm(pts_xyl, ~ pH_H2O_0_5 + pH_H2O_5_25 + pH_H2O_80_100 +
  MO_0_5 + MO_5_25 + MO_80_100 + P_0_5 + P_5_25 + P_80_100 +
  Na_0_5 + Na_5_25 + Na_80_100 + K_0_5 + K_5_25 + K_80_100 +
  Ca_0_5 + Ca_5_25 + Ca_80_100 + Mg_0_5 + Mg_5_25 + Mg_80_100 +
  Al_0_5 + Al_5_25 + Al_80_100 + CTC_0_5 + CTC_5_25 +
  CTC_80_100 + V_0_5_Porc + V_5_25_Porc + V_80_100_Porc +
  m_0_5_Porc + m_5_25_Porc + m_80_100_Porc + S_0_5 + S_5_25 +
  S_80_100 + Declividade + Altitude + Sand_0_5_Porc +
  Sand_5_25_Porc + Sand_80_100_Porc + Silt_0_5_Porc +
  Silt_5_25_Porc + Silt_80_100_Porc, covariates = list(
  pH_H2O_0_5 = pH_H2O_0_5, pH_H2O_5_25 = pH_H2O_5_25,
  pH_H2O_80_100 = pH_H2O_80_100, MO_0_5 = MO_0_5, MO_5_25 =
  MO_5_25, MO_80_100 = MO_80_100, P_0_5 = P_0_5, P_5_25 =
  P_5_25, P_80_100 = P_80_100, Na_0_5 = Na_0_5, Na_5_25 =
  Na_5_25, Na_80_100 = Na_80_100, K_0_5 = K_0_5, K_5_25 =
  K_5_25, K_80_100 = K_80_100, Ca_0_5 = Ca_0_5, Ca_5_25 =
  Ca_5_25, Ca_80_100 = Ca_80_100, Mg_0_5 = Mg_0_5, Mg_5_25 =
  Mg_5_25, Al_0_5 = Al_0_5, Al_5_25 = Al_5_25, Al_80_100 =
  Al_80_100, CTC_0_5 = CTC_0_5, CTC_5_25 = CTC_5_25,
  CTC_80_100 = CTC_80_100, V_0_5_Porc = V_0_5_Porc,
  V_5_25_Porc = V_5_25_Porc, V_80_100_Porc = V_80_100_Porc,
  m_0_5_Porc = m_0_5_Porc, m_5_25_Porc = m_5_25_Porc,
  m_80_100_Porc = m_80_100_Porc, S_0_5 = S_0_5, S_5_25 =
  S_5_25, S_80_100 = S_80_100, Declividade = Declividade,
  Altitude = Altitude, Sand_0_5_Porc = Sand_0_5_Porc,
  Sand_5_25_Porc = Sand_5_25_Porc, Sand_80_100_Porc =
  Sand_80_100_Porc, Silt_0_5_Porc = Silt_0_5_Porc,
  Sand_5_25_Porc = Sand_5_25_Porc, Silt_80_100_Porc =
  Silt_80_100_Porc),Poisson())

stepAIC(fit.com_cop)
stepAIC(fit.com_voc)
stepAIC(fit.com_xyl)

plot(fit.com_cop,ngrid=c(16,16),superimpose=FALSE,se=FALSE)
points(pts_cop,cex=0.5,pch=16)

plot(fit.com_voc,ngrid=c(16,16),superimpose=FALSE,se=TRUE,trend=FALSE)
points(pts_voc,cex=0.5,pch=16)

plot(fit.com_xyl,ngrid=c(16,16),superimpose=FALSE,se=TRUE,trend=FALSE)
points(pts_xyl,cex=0.5,pch=16)

# Seleção de variáveis Stepwise com variável construída

fit.com2_cop <- ppm(pts_cop, ~ pH_H2O_0_5 + pH_H2O_5_25 + pH_H2O_80_100 +
  MO_0_5 + MO_5_25 + MO_80_100 + P_0_5 + P_5_25 + P_80_100 +
  Na_0_5 + Na_5_25 + Na_80_100 + K_0_5 + K_5_25 + K_80_100 +
  Ca_0_5 + Ca_5_25 + Ca_80_100 + Mg_0_5 + Mg_5_25 + Mg_80_100 +
  Al_0_5 + Al_5_25 + Al_80_100 + CTC_0_5 + CTC_5_25 +
  CTC_80_100 + V_0_5_Porc + V_5_25_Porc + V_80_100_Porc +
  m_0_5_Porc + m_5_25_Porc + m_80_100_Porc + S_0_5 + S_5_25 +
  S_80_100 + Declividade + Altitude + Sand_0_5_Porc +
  Sand_5_25_Porc + Sand_80_100_Porc + Silt_0_5_Porc +
  Silt_5_25_Porc + Silt_80_100_Porc + Int_sem_cop,
  covariates = list(

```



```

pH_H2O_0_5 = pH_H2O_0_5, pH_H2O_5_25 = pH_H2O_5_25,
pH_H2O_80_100 = pH_H2O_80_100, MO_0_5 = MO_0_5, MO_5_25 =
MO_5_25, MO_80_100 = MO_80_100, P_0_5 = P_0_5, P_5_25 =
P_5_25, P_80_100 = P_80_100, Na_0_5 = Na_0_5, Na_5_25 =
Na_5_25, Na_80_100 = Na_80_100, K_0_5 = K_0_5, K_5_25 =
K_5_25, K_80_100 = K_80_100, Ca_0_5 = Ca_0_5, Ca_5_25 =
Ca_5_25, Ca_80_100 = Ca_80_100, Mg_0_5 = Mg_0_5, Mg_5_25 =
Mg_5_25, Al_0_5 = Al_0_5, Al_5_25 = Al_5_25, Al_80_100 =
Al_80_100, CTC_0_5 = CTC_0_5, CTC_5_25 = CTC_5_25,
CTC_80_100 = CTC_80_100, V_0_5_Porc = V_0_5_Porc,
V_5_25_Porc = V_5_25_Porc, V_80_100_Porc = V_80_100_Porc,
m_0_5_Porc = m_0_5_Porc, m_5_25_Porc = m_5_25_Porc,
m_80_100_Porc = m_80_100_Porc, S_0_5 = S_0_5, S_5_25 =
S_5_25, S_80_100 = S_80_100, Declividade = Declividade,
Altitude = Altitude, Sand_0_5_Porc = Sand_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Sand_80_100_Porc =
Sand_80_100_Porc, Silt_0_5_Porc = Silt_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Silt_80_100_Porc =
Silt_80_100_Porc, Int_sem_cop = Int_sem_cop), Poisson())

fit.com2_voc <- ppm(pts_voc, ~ pH_H2O_0_5 + pH_H2O_5_25 + pH_H2O_80_100 +
MO_0_5 + MO_5_25 + MO_80_100 + P_0_5 + P_5_25 + P_80_100 +
Na_0_5 + Na_5_25 + Na_80_100 + K_0_5 + K_5_25 + K_80_100 +
Ca_0_5 + Ca_5_25 + Ca_80_100 + Mg_0_5 + Mg_5_25 + Mg_80_100 +
Al_0_5 + Al_5_25 + Al_80_100 + CTC_0_5 + CTC_5_25 +
CTC_80_100 + V_0_5_Porc + V_5_25_Porc + V_80_100_Porc +
m_0_5_Porc + m_5_25_Porc + m_80_100_Porc + S_0_5 + S_5_25 +
S_80_100 + Declividade + Altitude + Sand_0_5_Porc +
Sand_5_25_Porc + Sand_80_100_Porc + Silt_0_5_Porc +
Silt_5_25_Porc + Silt_80_100_Porc + Int_sem_voc,
covariates = list(
pH_H2O_0_5 = pH_H2O_0_5, pH_H2O_5_25 = pH_H2O_5_25,
pH_H2O_80_100 = pH_H2O_80_100, MO_0_5 = MO_0_5, MO_5_25 =
MO_5_25, MO_80_100 = MO_80_100, P_0_5 = P_0_5, P_5_25 =
P_5_25, P_80_100 = P_80_100, Na_0_5 = Na_0_5, Na_5_25 =
Na_5_25, Na_80_100 = Na_80_100, K_0_5 = K_0_5, K_5_25 =
K_5_25, K_80_100 = K_80_100, Ca_0_5 = Ca_0_5, Ca_5_25 =
Ca_5_25, Ca_80_100 = Ca_80_100, Mg_0_5 = Mg_0_5, Mg_5_25 =
Mg_5_25, Al_0_5 = Al_0_5, Al_5_25 = Al_5_25, Al_80_100 =
Al_80_100, CTC_0_5 = CTC_0_5, CTC_5_25 = CTC_5_25,
CTC_80_100 = CTC_80_100, V_0_5_Porc = V_0_5_Porc,
V_5_25_Porc = V_5_25_Porc, V_80_100_Porc = V_80_100_Porc,
m_0_5_Porc = m_0_5_Porc, m_5_25_Porc = m_5_25_Porc,
m_80_100_Porc = m_80_100_Porc, S_0_5 = S_0_5, S_5_25 =
S_5_25, S_80_100 = S_80_100, Declividade = Declividade,
Altitude = Altitude, Sand_0_5_Porc = Sand_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Sand_80_100_Porc =
Sand_80_100_Porc, Silt_0_5_Porc = Silt_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Silt_80_100_Porc =
Silt_80_100_Porc, Int_sem_voc = Int_sem_voc), Poisson())

fit.com_xyl <- ppm(pts_xyl, ~ pH_H2O_0_5 + pH_H2O_5_25 + pH_H2O_80_100 +
MO_0_5 + MO_5_25 + MO_80_100 + P_0_5 + P_5_25 + P_80_100 +
Na_0_5 + Na_5_25 + Na_80_100 + K_0_5 + K_5_25 + K_80_100 +
Ca_0_5 + Ca_5_25 + Ca_80_100 + Mg_0_5 + Mg_5_25 + Mg_80_100 +
Al_0_5 + Al_5_25 + Al_80_100 + CTC_0_5 + CTC_5_25 +
CTC_80_100 + V_0_5_Porc + V_5_25_Porc + V_80_100_Porc +
m_0_5_Porc + m_5_25_Porc + m_80_100_Porc + S_0_5 + S_5_25 +

```

```

S_80_100 + Declividade + Altitude + Sand_0_5_Porc +
Sand_5_25_Porc + Sand_80_100_Porc + Silt_0_5_Porc +
Silt_5_25_Porc + Silt_80_100_Porc + Int_sem_xyl,
covariates = list(
pH_H2O_0_5 = pH_H2O_0_5, pH_H2O_5_25 = pH_H2O_5_25,
pH_H2O_80_100 = pH_H2O_80_100, MO_0_5 = MO_0_5, MO_5_25 =
MO_5_25, MO_80_100 = MO_80_100, P_0_5 = P_0_5, P_5_25 =
P_5_25, P_80_100 = P_80_100, Na_0_5 = Na_0_5, Na_5_25 =
Na_5_25, Na_80_100 = Na_80_100, K_0_5 = K_0_5, K_5_25 =
K_5_25, K_80_100 = K_80_100, Ca_0_5 = Ca_0_5, Ca_5_25 =
Ca_5_25, Ca_80_100 = Ca_80_100, Mg_0_5 = Mg_0_5, Mg_5_25 =
Mg_5_25, Al_0_5 = Al_0_5, Al_5_25 = Al_5_25, Al_80_100 =
Al_80_100, CTC_0_5 = CTC_0_5, CTC_5_25 = CTC_5_25,
CTC_80_100 = CTC_80_100, V_0_5_Porc = V_0_5_Porc,
V_5_25_Porc = V_5_25_Porc, V_80_100_Porc = V_80_100_Porc,
m_0_5_Porc = m_0_5_Porc, m_5_25_Porc = m_5_25_Porc,
m_80_100_Porc = m_80_100_Porc, S_0_5 = S_0_5, S_5_25 =
S_5_25, S_80_100 = S_80_100, Declividade = Declividade,
Altitude = Altitude, Sand_0_5_Porc = Sand_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Sand_80_100_Porc =
Sand_80_100_Porc, Silt_0_5_Porc = Silt_0_5_Porc,
Sand_5_25_Porc = Sand_5_25_Porc, Silt_80_100_Porc =
Silt_80_100_Porc, Int_sem_xyl = Int_sem_xyl), Poisson())

stepAIC(fit.com2_cop)
stepAIC(fit.com2_voc)
stepAIC(fit.com2_xyl)

AIC(fit.com_cop, fit.com_voc, fit.com_xyl)
AIC(fit.com_cop, fit.com_voc, fit.com_xyl)

E1_cop = envelope(fit.com2_cop, Linhom, nsim = 500, global = FALSE,
correction = "border")
E1_voc = envelope(fit.com2_voc, Linhom, nsim = 500, global = FALSE,
correction = "border")
E1_xyl = envelope(fit.com2_xyl, Linhom, nsim = 500, global = FALSE,
correction = "border")

plot(E1_cop, main = "", legend = FALSE, .r~r)
plot(E1_voc, main = "", legend = FALSE, .r~r)
plot(E1_xyl, main = "", legend = FALSE, .r~r)

# Ajuste de modelos LGCP utilizando os melhores modelos

fit2_cop_lgcp <- kppm(pts_cop, ~ MO_0_5 + poly(MO_80_100, 3)+
poly(P_0_5, 3)+poly(K_5_25, 2)+poly(K_80_100, 2)+
poly(Mg_0_5, 3)+poly(Al_0_5, 3)+poly(Al_80_100, 3)+
poly(CTC_0_5, 3)+poly(CTC_5_25, 2)+poly(V_80_100_Porc, 2)+
m_80_100_Porc+Declividade+Altitude+poly(Sand_0_5_Porc, 3)+
poly(Int_sem_cop, 2),
covariates = list(MO_0_5=MO_0_5, MO_80_100=MO_80_100,
P_0_5=P_0_5, K_5_25=K_5_25, K_80_100=K_80_100, Mg_0_5=Mg_0_5,
Al_0_5=Al_0_5, Al_80_100=Al_80_100, CTC_0_5=CTC_0_5,
CTC_5_25=CTC_5_25, V_80_100_Porc=V_80_100_Porc,
m_80_100_Porc= m_80_100_Porc, Declividade=Declividade,
Altitude=Altitude, Sand_0_5_Porc=Sand_0_5_Porc,
Silt_80_100_Porc=Silt_80_100_Porc, Int_sem_cop=Int_sem_cop),
clusters = "LGCP")

```

```

fit2_voc_lgcp <- kppm(pts_voc, ~poly(pH_H2O_80_100, 2)+poly(MO_0_5, 2)+
  poly(MO_5_25, 2)+poly(MO_80_100, 3)+poly(P_0_5, 2)+
  poly(Na_0_5, 3)+poly(K_0_5, 2)+poly(K_80_100, 3)+
  poly(Al_0_5, 2)+poly(Al_5_25, 2)+Al_80_100+poly(CTC_0_5, 3)+
  CTC_80_100+poly(V_5_25_Porc, 3)+V_80_100_Porc+m_0_5_Porc+
  poly(m_80_100_Porc, 3)+poly(S_5_25, 2)+poly(Declividade, 3)+
  poly(Altitude, 2)+poly(Sand_0_5_Porc, 2)+Silt_0_5_Porc+
  poly(Int_sem_voc, 2),
  covariates = list(
    pH_H2O_80_100=pH_H2O_80_100,MO_0_5=MO_0_5,MO_5_25=MO_5_25,
    MO_80_100=MO_80_100,P_0_5=P_0_5,Na_0_5=Na_0_5,K_0_5=K_0_5,
    K_80_100=K_80_100,Al_0_5=Al_0_5,Al_5_25=Al_5_25,Al_80_100=
    Al_80_100,CTC_0_5=CTC_0_5,CTC_80_100=CTC_80_100,
    V_5_25_Porc=V_5_25_Porc,V_80_100_Porc=V_80_100_Porc,
    m_0_5_Porc=m_0_5_Porc,m_80_100_Porc=m_80_100_Porc,
    S_5_25=S_5_25,Declividade=Declividade,Altitude=Altitude,
    Sand_0_5_Porc=Sand_0_5_Porc,Silt_0_5_Porc=Silt_0_5_Porc,
    Int_sem_voc=Int_sem_voc),clusters = "LGCP")

fit2_xyl_lgcp <- kppm(pts_xyl, ~pH_H2O_5_25+MO_0_5+MO_80_100 +P_0_5 +
  poly(P_5_25, 3)+poly(Na_0_5, 2)+K_0_5+poly(Ca_0_5, 2)+
  poly(Al_0_5, 2)+CTC_0_5+poly(CTC_5_25, 3)+
  poly(V_0_5_Porc, 3)+V_5_25_Porc+poly(m_5_25_Porc, 2) +
  poly(m_80_100_Porc, 3)+poly(S_0_5, 3)+poly(S_80_100, 2) +
  poly(Declividade, 3)+poly(Altitude, 3)+
  poly(Sand_0_5_Porc, 3)+poly(Silt_80_100_Porc, 2)+
  Int_sem_xyl,
  covariates = list(
    pH_H2O_5_25=pH_H2O_5_25,MO_0_5=MO_0_5,MO_80_100=MO_80_100,
    P_0_5=P_0_5,P_5_25=P_5_25,Na_0_5=Na_0_5,K_0_5=K_0_5,Ca_0_5=
    Ca_0_5,Al_0_5=Al_0_5,CTC_0_5=CTC_0_5,CTC_5_25=CTC_5_25,
    V_0_5_Porc=V_0_5_Porc,V_5_25_Porc=V_5_25_Porc,m_5_25_Porc=
    m_5_25_Porc,m_80_100_Porc=m_80_100_Porc,S_0_5=S_0_5,S_80_100=
    S_80_100,Declividade=Declividade,Altitude=Altitude,
    Sand_0_5_Porc=Sand_0_5_Porc,Silt_80_100_Porc=Silt_80_100_Porc,
    Int_sem_xyl=Int_sem_xyl),
  clusters = "LGCP")

plot(envelope(fit2_cop_lgcp, Linhom, nsim = 1000, global = FALSE,
  correction = "border"), main = "", legend = FALSE,.r~r)
plot(envelope(fit2_voc_lgcp, Linhom, nsim = 1000, global = FALSE,
  correction = "border"), main = "", legend = FALSE,.r~r)
plot(envelope(fit2_xyl_lgcp, Linhom, nsim = 1000, global = FALSE,
  correction = "border"), main = "", legend = FALSE,.r~r)

# Utilizando a análise multivariada Análise Fatorial

load("~/Pasta Final/fatorial.RData")
fatorial <- fatorial
cov <- as.vector(fatorial)
covs <- cbind(grid1,cov,int_sem_cop,int_sem_voc,
  int_sem_xyl)

grid.covs <- covs
coordinates(grid.covs) <-c("x","y")
class(grid.covs)

```

```

grid.covs1      <- as(grid.covs, "SpatialPixelsDataFrame")
class(grid.covs1)
names(grid.covs1)

Fator1      <- as.im.SpatialGridDataFrame(grid.covs1[,1])
Fator2      <- as.im.SpatialGridDataFrame(grid.covs1[,2])
Fator3      <- as.im.SpatialGridDataFrame(grid.covs1[,3])
Fator4      <- as.im.SpatialGridDataFrame(grid.covs1[,4])
Fator5      <- as.im.SpatialGridDataFrame(grid.covs1[,5])
Fator6      <- as.im.SpatialGridDataFrame(grid.covs1[,6])
Fator7      <- as.im.SpatialGridDataFrame(grid.covs1[,7])
Fator8      <- as.im.SpatialGridDataFrame(grid.covs1[,8])
Int_sem_cop <- as.im.SpatialGridDataFrame(grid.covs1[,9])
Int_sem_voc <- as.im.SpatialGridDataFrame(grid.covs1[,10])
Int_sem_xyl <- as.im.SpatialGridDataFrame(grid.covs1[,11])

Fator1      <- rotate(Fator1,ang=-ang)
Fator2      <- rotate(Fator2,ang=-ang)
Fator3      <- rotate(Fator3,ang=-ang)
Fator4      <- rotate(Fator4,ang=-ang)
Fator5      <- rotate(Fator5,ang=-ang)
Fator6      <- rotate(Fator6,ang=-ang)
Fator7      <- rotate(Fator7,ang=-ang)
Fator8      <- rotate(Fator8,ang=-ang)
Int_sem_cop <- rotate(Int_sem_cop,ang=-ang)
Int_sem_voc <- rotate(Int_sem_voc,ang=-ang)
Int_sem_xyl <- rotate(Int_sem_xyl,ang=-ang)

# Seleção de variáveis Stepwise

# Copaifera

fit.completo_cop <- ppm(pts_cop, ~Fator1+Fator2+Fator3+Fator4+
Fator5+Fator6+Fator7+Fator8,covariates=list(Fator1=Fator1,
Fator2=Fator2,Fator3=Fator3,Fator4=Fator4,Fator5=Fator5,
Fator6=Fator6,Fator7=Fator7,Fator8=Fator8),Poisson())

# Vochysia

fit.completo_voc <- ppm(pts_voc, ~Fator1+Fator2+Fator3+Fator4+Fator5+
Fator6+Fator7+Fator8,covariates=list(Fator1=Fator1,Fator2=Fator2,
Fator3=Fator3, Fator4=Fator4,Fator5=Fator5,Fator6=Fator6,Fator7=
Fator7, Fator8=Fator8),Poisson())

# Xylopia

fit.completo_xyl <- ppm(pts_xyl, ~Fator1+Fator2+Fator3+Fator4+Fator5 +
Fator6+Fator7+Fator8,covariates=list(Fator1=Fator1,Fator2=Fator2,Fator3=
Fator3,Fator4=Fator4,Fator5=Fator5,Fator6=Fator6,Fator7=Fator7,
Fator8=Fator8),Poisson())

stepAIC(fit.completo_cop);stepAIC(fit.completo_voc);stepAIC(fit.completo_xyl)

# 1 Análise (Considerar apenas o efeito das covariáveis)

# Copaifera

fit.completo_cop <- ppm( pts_cop, ~ poly(Fator1,3)+poly(Fator2,3)+

```

```
poly(Fator3,2)+poly(Fator5,2)+poly(Fator7,3)+ poly(Fator8,2),covariates =
list(Fator1=Fator1, Fator2=Fator2,Fator3=Fator3,Fator5=Fator5,
Fator7=Fator7,Fator8=Fator8),Poisson())
summary(fit.completo_cop)[33]
AIC(fit.completo_cop)
```

```
# Vochysia
```

```
fit.completo_voc <- ppm(pts_voc, ~poly(Fator1,3)+poly(Fator2,3)+
poly(Fator3,2)+poly(Fator4)+poly(Fator8,2),covariates = list(
Fator1=Fator1,Fator2=Fator2,Fator3=Fator3,Fator4=Fator4,Fator8=Fator8),
Poisson())
summary(fit.completo_voc)[33]
AIC(fit.completo_voc)
```

```
# Xylopia
```

```
fit.completo_xyl <- ppm(pts_xyl, ~ Fator1+poly(Fator2,3)+Fator3+
poly(Fator5,3)+poly(Fator6,2)+poly(Fator8,2), covariates=list(Fator1=
Fator1,Fator2=Fator2, Fator3=Fator3,Fator5=Fator5,Fator6=Fator6,
Fator8=Fator8),Poisson())
summary(fit.completo_xyl)[33]
AIC(fit.completo_xyl)
```

```
# 2 Análise (Considerar além do efeito das covariáveis, a intensidade
# das outras espécies)
```

```
# Copaifera
```

```
fit.completo2_cop <- ppm(pts_cop, ~ Fator1 + Fator2 + Fator3 + Fator4 +
Fator5 + Fator6 + Fator7 + Fator8 + Int_sem_cop, covariates = list(Fator1=
Fator1,Fator2 = Fator2, Fator3 = Fator3, Fator4 = Fator4, Fator5 = Fator5,
Fator6 = Fator6, Fator7 = Fator7, Fator8 = Fator8,Int_sem_cop = Int_sem_cop),
Poisson())
stepAIC(fit.completo2_cop)
```

```
fit.completo2_cop <- ppm(pts_cop, ~ poly(Fator1, 3)+ poly(Fator2, 3)+
poly(Fator3, 2)+Fator5+poly(Fator7, 3)+poly(Fator8, 2)+ poly(Int_sem_cop, 2),
covariates = list( Fator1=Fator1,Fator2=Fator2, Fator3 = Fator3,Fator5 = Fator5,
Fator7 = Fator7, Fator8 =Fator8, Int_sem_cop = Int_sem_cop),Poisson())
summary(fit.completo2_cop)[33]
AIC(fit.completo2_cop)
```

```
# Vochysia
```

```
fit.completo2_voc <- ppm(pts_voc, ~ Fator1 + Fator2 +Fator3 + Fator4 +
Fator5 +Fator6 +Fator7 + Fator8+ Int_sem_voc, covariates = list(Fator1=Fator1,
Fator2=Fator2,Fator3 = Fator3,Fator4=Fator4,Fator5=Fator5, Fator6=Fator6,
Fator7 = Fator7, Fator8 = Fator8, Int_sem_voc = Int_sem_voc),
Poisson())
stepAIC(fit.completo2_voc)
```

```
fit.completo2_voc <- ppm(pts_voc, ~ poly(Fator1, 3)+ poly(Fator2, 3)+
poly(Fator3, 2)+ Fator4+ poly(Fator8, 2), covariates = list( Fator1=Fator1,
Fator2=Fator2, Fator3 = Fator3,Fator4 = Fator4, Fator8 = Fator8),Poisson())
summary(fit.completo2_voc)[33]
AIC(fit.completo2_voc)
```

```
# Xylophia
```

```
fit.completo2_xyl <- ppm(pts_xyl, ~ Fator1 + Fator2 + Fator3 +Fator4 +
Fator5+Fator6+Fator7+Fator8+Int_sem_xyl,covariates = list( Fator1 = Fator1,
Fator2 = Fator2,Fator3 = Fator3,Fator4 = Fator4, Fator5 = Fator5,
Fator6 = Fator6,Fator7 = Fator7,Fator8 = Fator8,Int_sem_xyl = Int_sem_xyl),
Poisson())
stepAIC(fit.completo2_xyl)
```

```
fit.completo2_xyl <- ppm(pts_xyl, ~Fator1+poly(Fator2, 3)+Fator3+
poly(Fator5,3)+Fator6+Int_sem_xyl,covariates=list(Fator1=Fator1,
Fator2=Fator2,Fator3=Fator3, Fator5=Fator5, Fator6 = Fator6,Int_sem_xyl =
Int_sem_xyl),Poisson())
summary(fit.completo2_xyl)[33]
AIC(fit.completo2_xyl)
```

```
E1_cop = envelope(fit.completo2_cop, Linhom, nsim = 50, global = FALSE,
correction = "border",rank=2)
E1_voc = envelope(fit.completo2_voc, Linhom, nsim = 50, global = FALSE,
correction = "border",rank=2)
E1_xyl = envelope(fit.completo2_xyl, Linhom, nsim = 50,global = FALSE,
correction = "border",rank=2)
```

```
plot(E1_cop, main = "", legend = FALSE,.r~r)
plot(E1_voc, main = "", legend = FALSE,.r~r)
plot(E1_xyl, main = "", legend = FALSE,.r~r)
```

```
# Copaifera
```

```
fit2_cop <- kppm(pts_cop, ~ poly(Fator1,3)+poly(Fator2,3)+
poly(Fator3,2)+Fator5+poly(Fator7, 3)+poly(Fator8, 2)+poly(Int_sem_cop, 2),
covariates = list(Fator1 = Fator1,Fator2 = Fator2, Fator3 =Fator3,
Fator5 = Fator5,Fator7 = Fator7,Fator8 =Fator8,Int_sem_cop = Int_sem_cop)
,clusters = "LGCP")
```

```
# Vochysia
```

```
fit2_voc <- kppm(pts_voc, ~poly(Fator1, 3)+poly(Fator2, 3)+poly(Fator3, 2) +
Fator4 +poly(Fator8, 2),covariates = list(Fator1 = Fator1,Fator2 = Fator2,
Fator3 = Fator3,Fator4 = Fator4,Fator8 = Fator8, Int_sem_voc = Int_sem_voc),
clusters = "LGCP")
```

```
# Vochysia
```

```
fit2_xyl <- kppm(pts_xyl, ~Fator1+poly(Fator2, 3)+Fator3+poly(Fator5, 3)+
Fator6+Int_sem_xyl,covariates = list(Fator1=Fator1,Fator2=Fator2,Fator3=Fator3,
Fator5=Fator5,Fator6=Fator6,Int_sem_xyl = Int_sem_xyl),clusters = "LGCP")
```

```
E2_xyl = envelope(fit2_xyl, Linhom, nsim = 50, global = FALSE,
correction = "border",rank=9)
E2_voc = envelope(fit2_voc, Linhom, nsim = 50, global = FALSE,
correction = "border",rank=2)
E2_xyl = envelope(fit2_xyl, Linhom, nsim = 50, global = FALSE,
correction = "border",rank=2)
plot(E2_xyl, main = "", legend = FALSE,.r~r)
```

```
plot(E2_cop, main = "", legend = FALSE,.r~r)
```

```
plot(E2_voc, main = "", legend = FALSE, .-r~r)  
plot(E2_xyl, main = "", legend = FALSE, .-r~r)
```