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

Modelos para relacionar variáveis de solos e área basal de
espécies florestais em uma área de vegetação natural

Simone Grego

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Esta-
tística e Experimentação Agronômica

Piracicaba

2014

Simone Grego
Bacharel em Estatística

**Modelos para relacionar variáveis de solos e área basal de
espécies florestais em uma área de vegetação natural**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **PAULO JUSTINIANO RIBEIRO Jr**

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Esta-
tística e Experimentação Agronômica

**Piracicaba
2014**

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

Grego, Simone

Modelos para relacionar variáveis de solos e área basal de espécies florestais em uma área de vegetação natural / Simone Grego. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.
67 p: il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Modelo condicional autoregressivo simultâneo 2. Modelo condicional autoregressivo
3. Geoestatística 4. Regressão múltipla 5. Stepwise 6. Modelo aditivo generalizado 7. Árvore de regressão I. Título

CDD 634.9285
G819m

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

Dedicatória

Meus pais ,
por tudo.

AGRADECIMENTOS

Em primeiro lugar queria agradecer a Deus, pois mesmo quando eu não acreditava mais em mim, Ele continuou me dando forças, obrigada Deus.

Agradeço em primeiro a minha família, meus pais Natal Grego Neto e Zilda Roque Grego, irmãos Célia Regina Grego e Sérgio Roberto Grego, sobrinhos Caio Grego, Eduardo Grego e Henrique Grego Rodrigues de Lima e cunhados Eduardo Rodrigues de Lima e Mônica Lopes Grego. Vocês são as coisas mais importante na minha vida, minha base e o que me dá força para continuar na luta.

A meus pais, obrigada pelo carinho e por me educar e fazer de mim o que sou hoje, vocês são o maior tesouro, amo vocês!

Célia uma irmã quase mãe, sem você eu não estaria onde estou, sua ajuda e força desde o começo foi fundamental. Obrigada por estar do meu lado sempre, amo você.

Ao meus amigos queridos, Talita Tanaka, Luiza Tanaka, Vanessa Rufino, George Lucas, Felipe Granado, Aline Tanaka, Noeli Sasaki, Elisangela Soares, Veronica Paukstys, Naimara Padro, Luiz Nakamura, Everton Batista, Erasnilson Vieira Camilo, Daniel Miquelluti, Reginaldo Hilário, Glaucio Parini, Maria Joseane, Marcello Neiva, Douglas Toledo, que estiveram presentes nessa etapa da minha vida, obrigada pela amizade e por dividir comigo as angústias e alegrias, vocês sempre estarão no meu coração, adoro vocês!

Ao meu orientador Paulo Justiniano Ribeiro Júnior, muito obrigada pela orientação no decorrer deste trabalho.

À todos os professores do Departamento de Ciências Exatas da ESALQ: o Prof. Dr. Carlos Tadeu dos Santos Dias, a Profa. Dra. Clarice Garcia Borges Demétrio, o Prof. Dr. Cristian Marcelo Villegas Lobos, o Prof. Dr. Décio Barbin, o Prof. Dr. Edwin Moises Marcos Ortega, o Prof. Dr. Idemauro Antonio Rodrigues de Lara, a Profa. Dra. Renata Alcarde, a Profa. Dra. Roseli Aparecida Leandro, o Prof. Dr. Silvio Sandoval Zocchi, a Profa. Dra. Sonia Maria De Stefano Piedade e a Profa. Dra. Taciana Villela Savian; pelo ensinamento, sabedoria e dedicação nesses anos.

Aos professores do Departamento de Estatística da Faculdade de Ciências e Tecnologia de Presidente Prudente da “Universidade Estadual Paulista Júlio de Mesquita Filho”. Muito Obrigada.

Às secretárias Luciane Brajão, Solange de Assis Paes Sabadin e Mayara

Segatto, obrigada pela grande ajuda ao longo desses anos.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP Eduardo Bonilha, Jorge Alexandre Wiendl e Rosni Honofre Aparecido Pinto.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo apoio financeiro em forma de bolsa de estudos.

A Simone Magalhães e toda a equipe do Biota Parcelas Permanentes por fornecerem os dados que foram utilizados neste trabalho.

"Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível."

Charles Chaplin

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
LISTA DE ABREVIATURAS	18
1 INTRODUÇÃO	21
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Caracterização de Florestas Restauradas	23
2.2 Métodos não-espaciais	24
2.2.1 Regressão Linear Múltipla	25
2.2.1.1 Seleção de variáveis por <i>Stepwise</i>	26
2.2.2 Modelo Aditivo Generalizado - GAM	26
2.2.3 Árvore de Classificação e Regressão	27
2.3 Análise Espacial	29
2.3.1 Dados de Área	29
2.3.2 Geoestatística	32
3 METODOLOGIA	35
3.1 Material	35
3.2 Método	38
4 RESULTADOS E DISCUSSÃO	41
4.1 Seleção de Covariáveis	42
4.2 Análise Fatorial	53
5 CONSIDERAÇÕES FINAIS	61
REFERÊNCIAS	63

RESUMO

Modelos para relacionar variáveis de solos e área basal de espécies florestais em uma área de vegetação natural

O padrão espacial de ocorrência de atributos de espécies florestais, tal como a área basal das árvores, pode fornecer informações para o entendimento da estrutura da comunidade vegetal. Uma vez que fatores ambientais podem influenciar tanto o padrão espacial de ocorrência quanto os atributos das espécies em florestas nativas. Desse modo, investigar a relação entre as características ambientais e o padrão espacial de espécies florestais pode ajudar a entender a dinâmica das florestas. Especificamente, neste trabalho, o objetivo é avaliar métodos estatísticos que permitam identificar quais atributos do solo são capazes de explicar a variação da área basal de cada espécie de árvore. A área basal foi considerada como variável resposta e como covariáveis, um grande número de atributos físicos e químicos do solo, medidos em uma malha de localizações cobrindo a área de estudo. Foram revisados e utilizados os métodos de regressão linear múltipla com método de seleção *stepwise*, modelos aditivos generalizados e árvores de regressão. Em uma segunda fase das análises, adicionou-se um efeito espacial aos modelos, com o intuito de verificar se havia ainda padrões na variabilidade, não capturados pelos modelos. Com isso, foram considerados os modelos autoregressivo simultâneo, condicional autoregressivo e geoestatístico. Dado o grande número de atributos do solo, as análises foram também conduzidas utilizando-se as covariáveis originais, fatores identificados em uma análise fatorial prévia dos atributos de solo. A seleção de modelos com melhor ajuste foi utilizada para identificar os atributos de solo relevantes, bem como a presença e melhor descrição de padrões espaciais. A área de estudo foi a Estação Ecológica de Assis, 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 Biota da FAPESP. As análises reportadas aqui se referem à área basal das espécies *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica*. Com os atributos de solo reduzidos e consistentemente associados à área basal, a declividade, altitude, saturação por alumínio e potássio mostraram-se relevantes para duas das espécies. Resultados obtidos mostraram a presença de um padrão na variabilidade, mesmo levando-se em consideração os efeitos das covariáveis, ou seja, os atributos do solo explicam parcialmente a variabilidade da área basal, mas existe um padrão que ocorre no espaço que não é capturado por essas covariáveis.

Palavras-chave: Modelo condicional autoregressivo simultâneo; Modelo condicional autoregressivo; Geoestatística; Regressão múltipla; *Stepwise*; Modelo aditivo generalizado; Árvore de regressão

ABSTRACT

Models to relate variable soil and basal area of forest species in an area of natural vegetation

The spatial pattern of occurrenceis of forest species and their attributes, such as the basal area of trees, can provide information for understanding the structure of the vegetable community. Considering the environmental factors can influence the spatial pattern of occurrences of species in native forests and related attributes, describing relationship between environmental characteristics and spatial pattern of forest species can be associated with the dynamics of forests. The objective of the present study is to assess different statistical methods used to identify which soil attributes are associated with the basal area of each tree selected species. The basal area was considered as the response variable and the covariates are given by a large number of physical and chemical attributes of the soil, measured at a grid of locations covering the study area. The methods considered are the multiple linear regression with stepwise model selection, generalized additive models and regression trees. Spatial effects were added to the models, in order to ascertain whether there is residual spatial patterns not captured by the covariates. Thus, simultaneous autoregressive model, autoregressive conditional and geostatistical were considered. Considering the large number of soil attributes, analysis were were conducted both ways, using the original covariates, and using factors identified in a preliminar factor analysis of the soil attributes. Model selection was used to identify the relevant attributes of soil as well as the presence and better description of spatial patterns. The study area was the Ecological Station of Assis, the Conservation Unit of the State of São Paulo in permanent plots within the "Diversity Dynamics and Conservation Forests in the State of São Paulo: 40 ha of permanent plots" project, under the research project FAPESP biota. The analyzes reported here refer to the basal area of the species *Copaifera langsdorffii*, *Vochysia tucanorum* and *Xylopia aromatica*. Results differ among the considered methods reinforcing the reccomendation of considering differing modeling strategies. Covariates consistently associated with basal area are slope, altitude and aluminum saturation, potassium, relevant to at least two of the species. Results obtained showed the presence of patterns in residual variability, even taking into account the effects of covariates. The soil characteristics only partially explain the variability of the basal area and there are spatial patterns not captured by these covariates.

Keywords: Simultaneous autoregressive conditional model; Conditional autoregressive model; Geostatistics; Multiple regression; stepwise; Generalized additive model; Regression tree

LISTA DE FIGURAS

Figura 1 - Ilustração de uma árvore de regressão/classificação	28
Figura 2 - Unidades de Conservação no Estado de São Paulo: Assis, Caetetus, Carlos Botelho e Ilha do Cardoso	35
Figura 3 - Representação da Parcela permanente subdividida em 256 subparcelas contínuas de 20 x 20m	36
Figura 4 - Configuração espacial das coordenadas do centro das subparcelas das quais foram avaliadas as espécies na Estação Ecológica de Assis.	41
Figura 5 - Gráfico mostrando os 256 centroides das subparcelas com pontos proporcionais a área basal de: <i>Copaifera langsdorffii</i> (a), <i>Vochysia tucanorum</i> (b), <i>Xylopia aromatica</i> (c)	42
Figura 6 - Variograma empírico e envelope simulado para verificação de dependência espacial das espécies: <i>Copaifera langsdorffii</i> (a), <i>Vochysia tucanorum</i> (b), <i>Xylopia aromatica</i> (c)	47
Figura 7 - Valores preditos do modelo M2 com o termo espacial (CAR, SAR e Geoestatístico) para a espécie <i>Copaifera langsdorffii</i>	48
Figura 8 - Valores preditos do modelo M5 com o termo espacial (CAR, SAR e Geoestatístico) para a espécie <i>Vochysia tucanorum</i>	48
Figura 9 - Valores preditos do modelo M6 com o termo espacial (CAR, SAR e Geoestatístico) para a espécie <i>Xylopia aromatica</i>	48
Figura 10 - Variograma empírico e envelope simulado para verificação de dependência espacial das espécies: <i>Copaifera langsdorffii</i> (a), <i>Vochysia tucanorum</i> (b), <i>Xylopia aromatica</i> (c)	56
Figura 11 - Comparação dos valores preditos dos modelos CAR, SAR e Geoestatístico com termo espacial para a espécie <i>Copaifera langsdorffii</i>	56
Figura 12 - Comparação dos valores preditos dos modelos CAR, SAR, e Geoestatístico com o termo espacial para a espécie <i>Vochysia tucanorum</i>	57
Figura 13 - Comparação dos valores preditos dos modelos CAR, SAR e Geoestatístico com termo espacial para a espécie <i>Xylopia aromatica</i>	57

LISTA DE TABELAS

Tabela 1 - Espécies mais abundantes na Estação Ecológica de Assis. As espécies estudadas neste trabalho estão em negrito	36
Tabela 2 - Estimativas dos parâmetros, erro padrão e valor p das covariáveis para a espécie <i>Copaifera langsdorffii</i> pelo método regressão linear múltipla . . .	43
Tabela 3 - Estimativas do parâmetros, erro padrão e valor p das covariáveis para a espécie <i>Vochysia tucanorum</i> pelo método regressão linear múltipla . . .	43
Tabela 4 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis para a espécie <i>Xylopia aromatica</i> pelo método de regressão linear múltipla . . .	44
Tabela 5 - Covariáveis selecionadas em cada método (RM, GAM e TREE), para os dois conjuntos de modelos	45
Tabela 6 - AIC dos modelos de regressão linear múltipla antes e depois do <i>stepwise</i>	46
Tabela 7 - Valor p do Índice de Moran para cada espécie	47
Tabela 8 - Valores do AIC para os modelos RM, CAR, SAR e geoestatístico, utilizando as covariáveis dos modelos M2, M5 e M6	49
Tabela 9 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis do modelo geoestatístico para a espécie <i>Copaifera langsdorffii</i>	50
Tabela 10 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis significativas do modelo geoestatístico para a espécie <i>Vochysia tucanorum</i> . .	51
Tabela 11 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis do modelo CAR para a espécie <i>Xylopia aromatica</i>	52
Tabela 12 - Valores dos parâmetros dos modelos para as espécies <i>Copaifera langsdorffii</i> , <i>Vochysia tucanorum</i> e <i>Xylopia aromatica</i>	52
Tabela 13 - Resultados da Análise Fatorial Análise (AF)	54
Tabela 15 - Fatores significativos a 22% pelos métodos: STEP, RM, GAM e TREE .	54
Tabela 16 - Covariáveis pertencentes ao modelo de regressão linear múltipla para as espécies: <i>Copaifera langsdorffii</i> , <i>Vochysia tucanorum</i> e <i>Xylopia aromatica</i>	55
Tabela 17 - O valor p do Índice de Moran para cada espécie	55
Tabela 18 - Comparando os modelos RM, SAR, CAR e geoestatístico pelo AIC . . .	58
Tabela 19 - Estimativas dos parâmetros, erro padrão e valor p dos fatores significativas a 5% dos modelos geoestatístico e o CAR	58
Tabela 20 - Valores dos parâmetros para os modelos geoestatístico e CAR	59

LISTA DE ABREVIATURAS

ESALQ – Escola Superior de Agricultura “Luiz de Queiroz”

EEA – Estação Ecológica de Assis

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

Silt –Silte

Sand –Areia

Clay –Argila

SAB-Restante –Soma da área basal das espécies restante

AF – Análise Fatorial

GAM –Modelo Aditivo Generalizado

TREE –Árvore de Regressão

RM –Regressão Linear Múltipla

CAR –Modelo Condicional Autoregressivo

SAR –Modelo Simultaneo Autoregressivo

1 INTRODUÇÃO

O padrão espacial de ocorrência das espécies de plantas e seus atributos em determinada área reflete diferentes influências nas limitações de variáveis ambientais existentes em cada área (CAMARERO; GUTIÉRREZ; FORTIN, 2000). O padrão espacial de medidas características das árvores, como o número de indivíduos, área basal, altura, volume e da biomassa, entre outros, depende de diversos processos ecológicos e características próprias de cada ambiente, de modo que a melhor compreensão deste quadro auxilia no conhecimento sobre formações florestais (CAPRETZ, 2004).

Segundo Mello (2004), a dinâmica de vida das florestas sofre mudanças rapidamente devido a fatores, que variam constantemente no espaço e no tempo. Fatores relacionados ao solo, ao clima e à própria planta influenciam acentuadamente nesta dinâmica, além das ações antrópicas, como adubação e melhoramento genético, dentre outros.

Um dos principais objetivos de estudo na ecologia vegetal de florestas tropicais é investigar as relações existentes entre a distribuição das populações de espécies arbóreas e as variáveis ambientais que podem afetar tal distribuição. Para Berg e Santos (2003) as propriedades físicas e químicas do solo e a topografia, em conjunto com os níveis de umidade e distribuição das chuvas, podem exercer importante papel na distribuição e composição de espécies.

Korman (2008) afirma que os fatores bióticos e abióticos atuam muitas vezes de forma simultânea e fortemente correlacionadas, influenciando na distribuição das espécies dentro de uma mesma formação florestal. Pesquisas relacionadas em formações florestais podem explicar de que modo uma grande quantidade de indivíduos de várias espécies estão aparentemente aptos a coexistirem num mesmo local e como estas espécies distribuem (NEWBERY; RENSHAW; BRUNING, 1986). Essas informações são relevantes, uma vez que, possibilita melhorar manejos tradicionais, orientando pesquisadores na elaboração de modelos de manejo sustentável (MOEUR, 1993).

Neste trabalho serão abordados métodos espaciais e não espaciais que visam explicar o comportamento de uma variável resposta a partir de um conjunto relativamente grande de covariáveis. Considera-se a área basal das espécies *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica* como variável resposta. As covariáveis contemplam atributos físicos e químicos do solo, coletados em uma malha de 256 pontos, sobreposta à área de estudo. Foram considerados um método paramétrico (regressão linear múltipla com seleção de variáveis *stepwise*), um método semi-paramétrico (modelo aditivo generali-

zado) e um não paramétrico (árvore de regressão), para reduzir o conjunto de covariáveis.

O objetivo é reduzir o conjunto de covariáveis, identificando os atributos do solo que contribuem para um ajuste de modelos, ou seja, investigar relações entre os fatores ambientais e a área basal das espécies arbóreas mais abundantes. Os atributos do solo são considerados nos modelos de duas formas, uma, com seus valores originais e outra, por meio de fatores construídos a partir de uma análise fatorial prévia. Os resultados dos modelos ajustados de uma e outra forma são apresentados e comparados.

Considera-se, ainda, a adição ao modelo de regressão de um efeito espacial, ajustando-se modelos como o modelo condicional autoregressivo (CAR), modelo autoregressivo simultâneo (SAR) e o geoestatístico. Buscou-se, desta forma, verificar se mesmo com a inclusão das covariáveis, há um padrão espacial remanescente, ou seja, se existe um padrão que ocorre no espaço que não é explicado pela variação dos atributos do solo.

A necessidade de inclusão da estrutura de autocorrelação espacial na análise dos dados ecológicos já havia sido sugerida por ecologistas por perceberem que a maioria se não todas as espécies de plantas, possui padrão espacial (GOSLEE, 2006). Identificar a estrutura espacial das espécies é de grande relevância para o entendimento dos processos ecológicos que estruturam a distribuição das plantas na natureza, os quais contribuem para a manutenção da alta diversidade biológica (ILLIAN et al., 2008). Conhecer o padrão espacial das espécies é fundamental para entender como dada espécie explora os recursos disponíveis, a importância de tais recursos, e suas funções no sucesso do estabelecimento e reprodução da espécie (CONDIT; ASHTON; BAKER, 2000).

Para a análise estatística, considerou-se a Estação de Assis, do projeto "Diversidade, dinâmica e conservação em florestas do Estado de São Paulo: 40 ha de parcelas permanentes", do programa Biotá da FAPESP. Este projeto contempla os quatro principais tipos de vegetação do Estado de São Paulo. O presente trabalho utiliza diferentes métodos estatísticos, espaciais e não espaciais, com o intuito de verificar se existe associação entre fatores ambientais e a distribuição espacial das espécies e, com isso, identificar as covariáveis relevantes e verificar a sua concordância.

2 REVISÃO BIBLIOGRÁFICA

2.1 Caracterização de Florestas Restauradas

No espaço geográfico do Estado de São Paulo restam, hoje, remanescentes florestais pertencentes a pelo menos quatro tipos de vegetação: a Floresta Restinga, a Floresta da Encosta da Serra do Mar, a Floresta Semidecídua do Planalto e o Cerradão. Estes tipos de vegetação apresentam flora e fauna próprias, e particularidades no processo de degradação e fragmentação. Cada uma destas unidades ocorre em ambientes condicionados a fatores abióticos particulares.

Nas últimas décadas, muita informação vem sendo acumulada sobre a composição e a estrutura dos remanescentes florestais paulistas, com destaque para os estudos de florística e aspectos da estrutura da vegetação. Uma característica em área reflorestada é a sua aparente homogeneidade, entretanto, é frequente a observação de variações ao longo do plantio, estas variações que ocorrem devido a vários fatores, como a topografia e a prática de plantio. O conhecimento dessas variações é importante para o manejo florestal.

Em síntese, as formações florestais paulistas parecem ser determinadas por fatores abióticos, tais como o clima, especialmente a temperatura e a precipitação e a água, através da umidade atmosférica disponível no solo. Estes determinantes, por sua vez, são influenciados por fatores como altitude e relevo.

Embora determinadas espécies possam estabelecer-se em solos pobres, diversos estudos relatam que o manejo dos solos auxilia o desenvolvimento da vegetação. A fertilização (BARROS et al., 1984), o tratamento da acidez (COSTA, 2005), a descompactação (MOREIRA et al., 2001), o plantio direto (SECCO et al., 2004) e a rotação de culturas (CORREIA; SHARMA, 2004) são práticas constantemente aplicadas que melhoram as características físico-químicas dos solos e, por consequência, aumentam a produção agrícola e/ou florestal.

As condições edáficas podem influenciar a riqueza e a composição de espécies na área, além de favorecer o desenvolvimento da vegetação. No entanto, a relação entre diversidade e fertilidade nem sempre é positiva. Na restauração de determinados ecossistemas a fertilização pode acarretar em perda de diversidade, o que pode estar relacionado com as características do ecossistema ou a facilitação do estabelecimento de espécies invasoras (REZENDE, 2010).

Devido à grande variabilidade de comportamento das espécies vegetais em

relação às condições físicas, químicas e biológicas dos solos, bem como a heterogeneidade edáfica dos ambientes a serem restaurados, torna-se indispensável o conhecimento do comportamento das espécies florestais frente às características físicas dos ecossistemas, tais como as exigências nutricionais (FURTINI NETO et al., 2000).

Os ecossistemas possuem fluxos de energia, de nutrientes e de água associados a sua capacidade suporte (ODUM, 1988). Assim, as intervenções que visam à restauração ecológica devem considerar tais características para garantir o sucesso destas atividades. Neste sentido, a fertilização, ao inserir nutrientes no sistema, pode favorecer e prejudicar o desenvolvimento do ecossistema.

O conhecimento acumulado das condições abióticas, que se relacionam com as espécies, ainda é insuficiente para direcionar os projetos de restauração. As variáveis geralmente possuem autocorrelação espacial, além disso, tão importante quanto observar a ocorrência do indivíduo, ou seja, a abundância, é a observação da capacidade de desenvolvimento ou acúmulo de biomassa, o que pode ser refletido pela sua área basal. A abundância da espécie fornece informações quanto à capacidade de dispersão e superação dos filtros ecológicos do local. Por outro lado, a área basal expressa a capacidade do indivíduo se desenvolver na área.

Desta maneira, a variável resposta área basal pode ser relacionada com as condições abióticas para estabelecer a real influência destas no sucesso de desenvolvimento da espécie. O papel dos fatores abióticos nas espécies florestais tem sido investigado em diversos estudos, como neste trabalho que utiliza métodos espaciais e não espacial, com o intuito de verificar se existe associação entre os fatores ambientais e a distribuição das espécies.

2.2 Métodos não-espaciais

Tendo-se um conjunto grande de variáveis, estas podem definir um modelo de várias formas diferentes, tanto do ponto de vista de quais as variáveis irão compor o modelo ou de como serão incorporadas. O ajuste de todos os possíveis modelos pode ser bastante trabalhoso e diferentes estratégias podem ser adotadas para a seleção de modelos.

2.2.1 Regressão Linear Múltipla

A análise de regressão é uma das técnicas mais usadas de estimação, havendo uma ampla literatura sobre o assunto, o livro de Weisberg (1985) é uma referência. Esta técnica estatística é usada para investigar e modelar a relação entre variáveis, sendo uma das mais utilizadas na análise de dados, com o intuito de estimar os parâmetros desconhecidos do modelo e verificar quais covariáveis contribuem significativamente para a variável resposta.

No ajuste de modelos de regressão, deve-se considerar a possível falta de ortogonalidade entre as covariáveis, ou seja, estas podem estar correlacionadas entre si. Se isto ocorre, ao se ajustar um modelo de regressão, a influência de uma covariável é alterada diante da presença ou ausência da outra variável explanatória no modelo. Além disso, com o aumento do número de variáveis regressoras diminui-se os graus de liberdade dos resíduos, até o ponto de não ser possível estimar os efeitos das variáveis (AZEVEDO, 1997).

Em situações em que as covariáveis são fortemente correlacionadas podem ocorrer problemas na regressão múltipla. Variáveis colineares não fornecem novas informações, dificultando a interpretação dos coeficientes obtidos na regressão. Por isso, recomenda-se a matriz de coeficientes de correlação simples entre as covariáveis para verificar a existência de uma possível colinearidade entre essas variáveis.

Gama et al. (2005), utilizaram regressão múltipla com o objetivo de desenvolver um modelo para explicar a relação entre as variáveis biofísicas de biomassa com o volume e a medida da altura das árvores e o seu diâmetro à altura do peito (DAP).

Existem alguns procedimentos que facilitam na seleção de covariáveis nos modelos de regressão múltipla. Dentre os vários métodos, pode ser destacado o da regressão passo a passo.

A regressão passo a passo consiste na incorporação ao modelo de uma variável a cada vez, com a finalidade de explicar a maior parte da variância que ainda não foi explicada pelo modelo. Métodos foram desenvolvidos para selecionar o melhor subconjunto de covariáveis sequencialmente, adicionando ou removendo variáveis em cada passo, como, por exemplo, o método *Stepwise* que foi utilizado neste trabalho.

2.2.1.1 Seleção de variáveis por *Stepwise*

Stepwise é um método de regressão para seleção de variáveis que consiste na combinação de dois outros métodos, o *forward* e *backward* (CHONG; JUN, 2005). Esse algoritmo adiciona, progressivamente, novas variáveis ao modelo, iniciando daquela com maior correlação com a resposta, como no método *forward*, e incorpora um mecanismo de eliminação de variáveis igual ao método de *backward*, ou seja, ele adiciona e elimina gradativamente novas variáveis ao modelo, iniciando daquela com maior correlação com a variável resposta.

Segundo Abbad e Torres (2002), *stepwise* é utilizado quando o interesse é em descrever a relação entre as variáveis, sem um modelo teórico consistente a ser seguido. Este método é utilizado como estudo exploratório, em que é elaborado um modelo teórico, e assim, investigado uma relação entre as variáveis.

O processo é iniciado quando duas variáveis são incluídas no modelo, verifica-se se a primeira não sai do modelo, e desta maneira, o processo continua até que nenhuma variável seja incluída ou retirada.

Alguns trabalhos relacionados com dados de floresta, que utilizaram *stepwise* como método de seleção de variáveis, são de Matos (2006), Verona (2003) e Maciel, Bastos e Watrin (2007).

2.2.2 Modelo Aditivo Generalizado - GAM

As referências iniciais sobre modelos aditivos generalizados o Wood (2006) descreve os fundamentos da GAM, com métodos de seleção automática de grau de liberdade, diversos exemplos e aplicações do pacote mgcv (WOOD, 2014) e Faraway (2006) ilustra o uso de modelos GAM. Hastie e Tibshirani (1990) e no artigo de Hastie e Tibshirani (1986) são consideradas as principais referências sobre modelo aditivo generalizado.

O modelo aditivo generalizado é uma extensão do modelo linear generalizado, em que o termo do modelo relacionado às covariáveis $\sum_i x_{ij}\beta_j$ é substituído por $\sum_j g_j(X_j)$, em que X_j são as covariáveis e $g_j(X_j)$ denota uma função semiparamétrica (isto é, cuja forma não é especificada), estimada por meio de funções de suavização (HASTIE; TIBSHIRANI, 1990). Com essa substituição, não é necessário assumir uma relação linear entre a variável resposta e as covariáveis. De fato, não é necessário nem mesmo conhecer previamente a forma dessa relação, mas é possível estimá-la a partir de um conjunto de dados. Essa função estimada $g_j(X_j)$, também é chamada de função de suavização ou curva

suavizada. Em algumas situações, é um tipo de média dos valores de Y_i na vizinhança de um dado valor x_i .

A função de suavização permite descrever a forma e também revelar uma possível ausência de linearidade nas relações estudadas, uma vez que não apresenta a estrutura rígida de uma função paramétrica. Portanto, GAM permite especificar tanto a relação entre a variável resposta e as covariáveis, de uma forma flexível como um modelo em termos de funções suaves, evitando modelos complexos. Os gráficos de curvas de suavização permitem uma visão da relação entre a variável resposta e as covariáveis, apresentando as possíveis não linearidades nas relações estudadas.

O modelo aditivo generalizado é considerado modelo semiparamétrico, porém, como o modelo linear generalizado, é paramétrico no que diz respeito a distribuição de probabilidade da variável resposta, no entanto, alguns preditores podem ser modelados de forma não-paramétrica por meio de termos lineares e polinômios.

Hastie e Tibshirani (1990) apresentam uma série de funções suavizadas. As mais comumente utilizadas e o caso mais simples de função suavizada são as *cubic*, *smooth*, *splines*, entre outras.

2.2.3 Árvore de Classificação e Regressão

O método de árvore de classificação e regressão, (*Classification And Regression Tree* - CART), permite a explicação de uma variável categórica ou numérica com base em um conjunto de covariáveis e das eventuais interações entre as mesmas. Essa técnica se destaca por ser flexível, não impondo qualquer restrição paramétrica nas variáveis sob estudo (BREIMAN; FRIEDMAN; OLSHEN, 1984). também fornece exemplos introdutórios dessa metodologia.

Este método pode ser utilizado como uma alternativa ao modelo clássico de regressão, uma vez que as variáveis respostas em CART podem ser numéricas ou não (categóricas). Se a variável resposta for numérico, como no presente trabalho, o modelo é chamado de árvore de regressão, caso contrário, o modelo é chamado de árvore de classificação.

Alguns termos são utilizados de forma recorrente na caracterização dos componentes de uma árvore de regressão e classificação. Denomina-se nó inicial à amostra original (equivale ao conjunto completos de dados), nós intermediários às subamostras que originam novas subamostras e os nós finais às subamostras não particionadas. Um

nó qualquer é denominado t e as partições são denominadas ramos. A representação gráfica de nós e ramos são conhecidos por árvore, em que cada nó corresponde uma divisão correspondente a uma covariável (BEEL, 1996).

Ou seja, uma árvore de classificação e regressão particiona os dados em diferentes sub-grupos ou regiões, que são tratados por modelos específicos localizados nas folhas destas estruturas, que são compostas por nós. A Figura 1 ilustra uma árvore de regressão e de classificação.

Figura 1 - Ilustração de uma árvore de regressão/classificação

A construção de uma árvore de regressão e classificação é dividida em quatro etapas: definir um critério de partição das amostras, aplicação do processo de poda, seleção do melhor modelo e classificação dos nós finais.

A árvore é o resultado final das partições sucessivas de uma amostra original, essas partições devem ser executadas com base nos resultados verificados para as covariáveis, buscando reduzir ao máximo a variação dos resultados nas subamostras constituídas em relação à variação avaliada no nó original. Deve-se estabelecer critérios para comparação e seleção de partições amostrais.

No entanto, algumas restrições devem ser consideradas quanto à partição de um nó. É adequado, por exemplo, determinar o número mínimo de elementos para que um nó possa ser partido. Subdividir nós com reduzido número de elementos pode aumentar a complexidade do modelo e afetar sua capacidade preditiva, dada a possível instabilidade dos nós formados. Também podem ocorrer problemas semelhantes caso nós com poucos

elementos sejam originados. Por isso, recomenda-se limitar o número mínimo de elementos nos nós formados.

Em árvores de regressão e classificação, como em qualquer outro procedimento de modelagem estatística, deve-se buscar um modelo parcimonioso, no caso, uma árvore de tamanho reduzido, com baixa heterogeneidade em seus nós finais. Construída a árvore, deve-se iniciar o procedimento de poda, que consiste em desfazer, uma a uma, aquelas partições que menos contribuem para a explicação da variável resposta.

Escolhida uma árvore, a caracterização dos nós finais é realizada pela classe que aparece com maior frequência dentre os elementos que constituem o nó, caso as variáveis sejam numéricas.

2.3 Análise Espacial

Dados espaciais são definidos por observações que são obtidas em localizações espaciais $s_1, s_2, \dots, s_n$, que são coordenadas no plano $\mathbb{R}^2$ ou espaço $\mathbb{R}^3$.

A análise de dados espaciais pode ser realizada sempre que as informações estiverem representadas no espaço. Seu objetivo é aprofundar a compreensão do processo, avaliar evidências de hipóteses ou ainda, tentar prever valores em áreas onde as observações não estão disponíveis (BAILEY; GATTREL, 1995).

Segundo Bailey e Gattrel (1995), a análise espacial pode ser distinguida entre os vários métodos a seguir:

- i) métodos voltados para a visualização de dados espaciais;
- ii) métodos exploratórios para investigar e resumir relações e padrões mapeados;
- iii) métodos para especificação de modelos estatísticos e para estimar parâmetros.

Para estes tipos de análise, conceitos como dependência espacial e autocorrelação espacial são fundamentais. Entende-se por dependência espacial o fato de que a maior parte das ocorrências naturais ou sociais apresenta, entre si, uma relação que depende da distância entre os pontos.

2.3.1 Dados de Área

Dados de área ocorrem em locais onde não se têm a localização exata dos eventos, mas de um valor por área. Exemplos relacionados a dados de área são a presença

de espécies de uma planta em um quadrado, o número de casos de dengue nos diversos bairros da cidade entre outros. A região sob estudo pode ser dividida em sub-regiões regulares (a divisão da região é uma grade) ou irregulares (municípios ou bairros de uma cidade).

Embora o valor da variável de interesse esteja associado a toda a área e não a um ponto particular, associa-se este valor a um ponto específico dentro desta área. Esse ponto é denominado centróide, que corresponde ao centro de massa do polígono ou malha regular que delimita a área (ASSUNÇÃO, 2001).

Os tipos de análise de área incluem:

- i) Análise exploratória, que procura uma descrição clara dos dados, a fim de ajudar o analista a desenvolver hipóteses sobre o assunto e modelos apropriados para tais dados (BAILEY; GATRELL, 1995);
- ii) Modelos de regressão espacial, que buscam estabelecer relações entre duas ou mais variáveis;
- iii) Modelos inferenciais, que buscam estabelecer a distribuição estatística subjacente aos dados coletados. Estes incluem modelos bayesianos empíricos e completos (CÂMARA et al., 2004).

Segundo Waller e Gotway (2004), na análise de dados de área é verificado o grau de similaridade ou a dependência espacial por meio da autocorrelação espacial, que pode ser quantificada pelo do índice de Moran (WALLER; GOTWAY, 2004). A aplicação desse índice depende da definição de uma matriz de vizinhança ou matriz de proximidade espacial.

Na regressão espacial é necessário verificar os resíduos em busca de sinais de estruturas espaciais. Inicialmente, para averiguar a presença de autocorrelação espacial utilizou-se o índice I de Moran que é calculado por (WALLER; GOTWAY, 2004):

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

em que n é o número de áreas ou de observações, Y_i é a variável aleatória na área i , Y_j é a variável aleatória na área j , $\bar{Y}$ é a média amostral da variável aleatória em toda região

e w_{ij} são elementos da matriz de proximidade espacial normalizada nas linhas, para que a soma dos pesos de cada linha seja igual a 1.

A matriz de proximidade $\mathbf{W}$ reflete a estrutura de vizinhança entre as áreas. Uma possível definição para $\mathbf{W}$ é feita a partir da matriz de adjacência $\mathbf{A}$, que considera $\mathbf{A}_{ij} = 1$ se as áreas i e j têm a borda comum e $\mathbf{A}_{ij} = 0$ quando não existe borda comum (KRAINSKI, 2008).

Como a função de autocorrelação é espacial, precisamos utilizar a matriz de vizinhança, que considera vizinhos das áreas os que partilham de fronteira comum. Nesse caso, a matriz recebe valor 1 quando são vizinhos e 0 caso contrário (SOUZA et al, 2008).

Há dois modelos mais utilizados na literatura (KRAINSKI, 2008), um é especificado por um conjunto de equações de regressão da observação de uma área nas observações das áreas vizinhas. Esse sistema de equações é resolvido de forma simultânea, conhecido como *Simultaneous Autoregressive* - SAR ou modelo espacial autoregressivo simultâneo. O outro modelo é especificado por um conjunto de distribuições condicionais é o *Conditional Autoregressive* - CAR ou modelo condicional autoregressivo, em que se propõe que a observação de uma área tem distribuição gaussiana, com parâmetro de média (média ponderada) das observações das áreas vizinhas e a variância sendo inversamente proporcional ao número de áreas vizinhas.

O modelo SAR tem parâmetro de ponderação da média condicional, esse parâmetro mede a força da dependência da observação de uma área com as observações das áreas vizinhas.

A expressão para o modelo SAR é dada por (CÂMARA et al, 2004):

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

em que:

- $\mathbf{W}$: é a matriz de proximidade espacial;
- $\mathbf{WY}$: induz a dependência espacial em $\mathbf{Y}$;
- $\mathbf{Y}$: vetor de observações ($n \times 1$) nas n áreas;
- $\mathbf{X}$: é uma matriz ($n \times p$) com $p - 1$ variáveis explicativas;
- $\boldsymbol{\beta}$: é o vetor de parâmetros $p \times 1$;

- ε : é um vetor ($n \times 1$) de erros aleatórios não correlacionados que segue uma distribuição normal com média zero e variância constante ($\varepsilon \sim N(0, I\sigma^2)$);
- ρ é o coeficiente espacial autoregressivo.

Neste caso, a hipótese nula para a não existência de autocorrelação é que $\rho = 0$. A idéia do modelo SAR é incorporar a autocorrelação espacial como componente. Observa-se que, na ausência de autocorrelação espacial ($\rho = 0$), o modelo SAR é exatamente o modelo de regressão linear.

O segundo modelo de regressão espacial mais utilizado na literatura é o CAR, que considera os efeitos espaciais como um ruído. Assim, os efeitos de autocorrelação espacial são associados ao termo de erro ξ e o modelo é expressado por:

$$\mathbf{Y} = \mathbf{X}\beta + \xi$$

em que $\mathbf{W}$ é a matriz de proximidade, considerando que o termo ξ é dado por:

$$\xi = \lambda \mathbf{W} + \varepsilon$$

em que λ é o coeficiente autoregressivo e ε é o componente do erro com uma variância constante e não correlacionada. A hipótese nula para a não existência de autocorrelação é $\lambda = 0$, ou seja, o termo não é espacialmente correlacionado. Pode-se observar que na ausência de autocorrelação espacial ($\lambda=0$), o modelo CAR é o modelo de regressão linear.

2.3.2 Geoestatística

A aplicação das técnicas da estatística tradicional, em alguns estudos, não envolve qualquer suposição a respeito da posição dos elementos amostrais entre eles. Na geoestatística assume-se que as diferenças entre dois pontos amostrados depende da distância entre eles e da orientação dos pontos (CLARK, 1979).

Considerando que em alguma área A exista um campo aleatório gaussiano Z latente, ou seja, o processo existe mas não é observável, é necessário fazer uma amostragem de n localizações espaciais dentro da área A e observar valores de atributos de interesse nas localizações amostradas. Desta forma, existe um vetor $\mathbf{Y}(\mathbf{x})$ de dimensão $n \times 1$, dos valores observados em $s = (s_1, s_2, \dots, s_n)$, que segundo Diggle e Ribeiro Jr 2007, sob o

pressuposto de gaussianidade, pode ser modelado como:

$$\mathbf{Y}(\mathbf{x}) = \mu + \mathbf{S}(\mathbf{x}) + \varepsilon$$

em que μ é um modelo de regressão, $S(\mathbf{x})$ é um processo espacial contínuo e ε é o ruído branco, um vetor de dimensão $n \times 1$, com distribuição normal de média zero e desvio padrão τ .

Com as técnicas geoestatísticas, é possível descrever a estrutura de dependência espacial dos dados e estabelecer a função de correlação espacial. Essa função representa o comportamento da variabilidade espacial dos dados, por meio de um modelo.

No estudo da dependência espacial de uma variável aleatória regionalizada, a teoria geoestatística baseia-se na função de semivariância ou covariância. A função semivariância ou covariância é importante para a definição da estrutura de variabilidade espacial e interpolação de valores.

A variável regionalizada é definida como uma função numérica com distribuição espacial, que varia de um local para o outro com continuidade aparente e cujas variações não podem ser representadas por uma função matemática simples. As características principais de uma variável regionalizada é a localização. Por isso, para analisar ou expressar a variabilidade espacial de alguma propriedade ou característica, é absolutamente necessário ter as coordenadas dos pontos de onde as amostras foram retiradas (VIEIRA, 2000).

Segundo Cressie (1993), o semivariograma, que é o gráfico da função semivariância pela distância, mostra a medida do grau de dependência espacial entre elementos amostrais ao longo da distância ou do espaço. Para sua construção, são usados valores da variável resposta, obtidos no campo com diferentes distâncias.

O variograma amostral é utilizado para verificar a presença de correlação espacial entre as unidades amostrais. Para a obtenção de um variograma é necessário supor que a variável regionalizada tenha um comportamento fracamente estacionário.

Segundo Camargo (1997), o variograma amostral é calculado por:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

em que, h é a distância em módulo de um ponto ao outro, N_h é o número de pares de

valores medidos que estão distantes por um vetor de módulo h , $z(x_i)$ é o valor da i -ésima observação coletada no ponto x_i e $z(x_i + h)$ é o valor da i -ésima observação coletada no ponto $x_i + h$.

No processo de modelagem do variograma são utilizados modelos básicos, os mais usuais são o Esférico, o Exponencial, o Gaussiano e o Matérn. Esses modelos têm formato característico como transitivo ou patamar.

A função de correlação de Matérn é dada por (MATÉRN, 1960):

$$\rho(\mu; \phi, \kappa) = 2^{k-1} \Gamma(\kappa)^{-1} (\mu/\phi)^k K_\kappa(\mu/\phi)$$

em que $\Gamma(\cdot)$ é a função gama, K_κ a função Bessel de ordem κ , μ é a distância euclidiana entre duas localizações quaisquer do campo aleatório, dado que $\phi > 0$ e $\kappa > 0$. O parâmetro ϕ está relacionado ao alcance das correlações e o parâmetro κ é relacionado à suavidade do processo, de forma que quando maior κ , maior a suavidade.

3 METODOLOGIA

3.1 Material

Os dados utilizados no presente trabalho foram coletados a partir de um estudo desenvolvido em quatro Unidades 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 Biota da FAPESP (RODRIGUES, 2002). As parcelas permanentes foram instaladas nas Unidades de Conservação do Parque Estadual Ilha do Cardoso, Parque Estadual de Carlos Botelho, Estação Ecológica de Caetetus e a Estação Ecológica de Assis. A Figura 2 apresenta o mapa do Estado de São Paulo com as quatro Unidades de Conservação.

Figura 2 - Unidades de Conservação no Estado de São Paulo: Assis, Caetetus, Carlos Botelho e Ilha do Cardoso

Em cada Unidade de Conservação está alocada uma parcela de 320 x 320 m, totalizando 10,24 ha, subdividida em 256 subparcelas de 400 m² (20 x 20 m), conforme representado pela Figura 3. Em cada subparcela foi fornecida a descrição dos fatores físicos e bióticos e também todos os indivíduos arbóreos no interior das parcelas, com circunferência na altura do peito igual ou superior a 15 cm, foram medidos, georreferenciados e identificados.

Figura 3 - Representação da Parcela permanente subdividida em 256 subparcelas contínuas de 20 x 20m

Para este trabalho, foram utilizados apenas as informações obtidas na Estação Ecológica de Assis, que possui 4.480 ha e localiza-se no município de Assis oeste paulista. Seu relevo é suave-ondulado e ocorrem dois tipos de solo, o Latossolo e Neossolo (EMBRAPA, 1999) e o clima local é Temperado Chuvoso e Quente, com chuvas de verão e verão quente (CLAUSET; SOARES, 1999).

Na Estação Ecológica de Assis foram levantadas 22576 árvores e as dez espécies mais abundantes nessa área são apresentadas na Tabela 1. As três espécies mais abundantes (*Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica*) foram utilizadas na análise estatística conduzida neste trabalho, uma vez que a soma de seus indivíduos correspondem a 34,8% das árvores presentes na parcela.

Tabela 1 - Espécies mais abundantes 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 pau-de-óleo, foi escolhida para este estudo por ser a mais abundante na região de estudo. Esta espécie produz um óleo que tem sido utilizado pela medicina popular como bactericida e antifalmatório (PIERI; MUSSI; MOREIRA, 2009).

A *Vochysia tucanorum*, popularmente conhecida como pau-de-tucano ou amarelinho devido as suas flores que são amarelas. Esta espécie tem uma substância com propriedade capaz de prevenir doenças como o câncer e ulcera gástrica (CLEMENTE 2010).

A espécie *Xylopia aromatica* apresenta altura de até 8m. Sua madeira utilizada na confecção de caixas leves e forros e florescem duas vezes ao ano.

Para este trabalho, utilizou-se as médias da área basal das espécies em cada subparcela (20 m x 20 m), que foram georreferenciadas pelas coordenadas centrais, que por sua vez foram utilizadas como variáveis resposta para o modelo.

A área basal expressa a capacidade do indivíduo se desenvolver na área, que pode ter relação com seu genótipo. Assim, variáveis resposta devem ser relacionadas com as condições abióticas para estabelecer a real influência destas no sucesso de desenvolvimento da espécie.

Os fatores físicos e químicos do solo também foram fornecidos. Essas informações foram coletadas a partir de uma grade de pontos cobrindo a área retangular em estudo (Figura 3). Para coleta dos fatores físicos e químicos, foram feitas as tradagens realizadas até um metro de profundidade em cada uma das 256 subparcelas das parcelas permanentes. As amostras foram coletadas em três profundidades (0-5, 5-25 e 80-100 cm na EEA). As análises físico-químicas de cada amostra foram feitas de acordo com a EMBRAPA (2006).

Para cada subparcela foram utilizadas as seguintes covariáveis ambientais:

- a) características topográficas: altitude e declividade (médias de cada subparcela);
- b) características edáficas: 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(%)), índices compostos de saturação por base (V(%)), saturação por alumínio (m(%)) e capacidade de troca de cation (cmolc/dm³).

Foi realizada primeiramente uma seleção de variáveis e depois verificado o melhor ajuste do modelo. Em seguida, acrescentou-se o efeito espacial, na forma de modelo para dados de área e geoestatístico e, compararam-se os modelos e resultados.

3.2 Método

Para as variáveis referentes a textura de solo (percentuais de areia, silte e argila) que somam 100%, utilizou-se a transformação Log-Aditiva. As covariáveis serão consideradas da seguinte maneira:

$$X1 = \log \left(\frac{\text{silte}}{\text{argila}} \right)$$

$$X2 = \log \left(\frac{\text{areia}}{\text{argila}} \right)$$

essa transformação será feita em todas as profundidades (0-5, 5-25 e 80-100).

Neste trabalho, 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 (MINGOTI, 2007). As variáveis alternativas são chamadas de fatores ou variáveis latentes. Primeiramente, deve-se verificar a adequação da análise fatorial, para tanto, foram utilizados os teste 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 escolher o número de fatores a extrair.

Também foi contruída uma nova variável denominada Soma das Áreas Basais Restantes, que leva em consideração a área basal de outras espécies na área de estudo. Então, neste trabalho, considerou-se dois conjuntos de modelos para cada espécie, um com as covariáveis ambientais em outro acrescentou-se mais a covariável soma das áreas basais restantes.

A transformação raiz quadrada foi utilizada na variável resposta para obter normalidade dos resíduos, verificada pelo BoxCox.

Realizou-se seleção de covariáveis com os métodos regressão linear múltipla com *stepwise*; análises de modelos aditivos generalizados; árvore de regressão e regressão

múltipla em dois estágios: no primeiro inclui-se somente as covariáveis ambientais, no segundo além das covariáveis ambientais, também foi considerada a variável soma de área basal das espécies restantes. Dessa forma, considerou-se uma grande variedade de modelos, com o intuito de selecionar o que apresenta o melhor ajuste aos dados. Utilizou-se o Critério de Informação de Akaike (AIC) para a escolha desses modelos.

Obtém-se o AIC através da função de máxima verossimilhança, a partir dos parâmetros ajustados para os modelos. O critério de AIC foi desenvolvido a partir da distância ou informação de Kullback e Leibler (1951). O AIC é dado por:

$$AIC = -2 \log(L) + 2k$$

em que L é a verossimilhança maximizada do modelo ajustado e k é o número de parâmetros. O AIC penaliza os modelos com número grande de parâmetros para evitar excesso de parametrização.

Em seguida, acrescenta-se o parâmetro espacial no modelo com as covariáveis reduzidas, ou seja, as covariáveis não são suficientes para explicar a variável resposta, mas existe um padrão que ocorre no espaço, que não é capturado por essas covariáveis.

Pela estrutura dos dados utilizados essa dissertação, as três abordagens mais utilizadas para caracterizar problemas de análise espacial, são:

- i) Superfícies contínuas: os dados distribuem-se continuamente em uma região e as localizações onde são feitas as medidas são fixas e podem estar regularmente ou irregularmente distribuídas;
- ii) Dados de área : utiliza dados cuja localização está associada a áreas delimitadas por polígonos ou grid regular, em que não se tem necessariamente a localização exata do evento, mas um valor que represente toda a área;
- iii) Padrões pontuais: são fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço.

Dentre as metodologias disponíveis para a análise de dados espaciais, a análise de dados de área e geoestatística são as mais adequadas para a área basal das espécies. Frade (2013), utilizou o mesmo banco de dados na abordagem de processos pontuais, considerando como variável resposta a abundância das árvores.

O modelo geoestatístico e os modelos para dados de área foram utilizados para incluir padrões espaciais em adição aos efeitos das cováriaveis ambientais.

Para o ajuste dos modelos de dados de área considerou as médias da área basal das espécies representando o valor da subparcela (20 m x 20 m). Este trabalho utilizou o modelo espacial autoregressivo simultâneo (SAR) e o modelo condicional autoregressivo (CAR), conhecido como regressão espacial.

Considerando a média da área basal como um ponto no espaço amostrado, utiliza-se a análise geoestatística, adotando-se como coordenada o centróide de cada área.

O Critério de Informação Akaike foi utilizado para comparar o ajuste dos modelos de regressão, geoestatístico, modelo condicional autoregressivo e o modelo condicional autoregressivo simultâneo. Ao verificar que as cováriaveis não são suficiente para explicar o modelo, existe um efeito espacial que o melhora.

As análises estatísticas desse trabalho foram realizadas no ambiente R (R Development Core team, 2008), bem como para a seleção de covariáveis, a utilização dos pacotes **rpart** (THERNEAU; ATKINSON; RIPLEY, 2014) para o método árvore de regressão e o **mgcv** (WOOD, 2014) para método modelo aditivo generalizado. Os pacotes **spdep** (BIVAND, 2014) para o ajuste do modelo SAR e CAR e o **geoR** (RIBEIRO JR; DIGGLE, 2013) para o ajuste do modelo geoestatístico.

4 RESULTADOS E DISCUSSÃO

Na Estação Ecológica de Assis foram encontradas 3977 indivíduos da espécie *Copaifera langsdorffii*, 2017 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 coordenada central, que representa a localização utilizada para as árvores, está representada na Figura 4.

Figura 4 - Configuração espacial das coordenadas do centro das subparcelas das quais foram avaliadas as espécies na Estação Ecológica de Assis.

Um gráfico que ajuda na inspeção de possíveis valores discrepantes com relação a sua vizinhança espacial está na Figura 5.

Figura 5 - Gráfico mostrando os 256 centroides das subparcelas com pontos proporcionais a área basal de: *Copaifera langsdorffii* (a), *Vochysia tucanorum* (b), *Xylopia aromatica* (c)

A Figura 5 apresenta círculos de tamanhos diferentes que indicam os valores dos quartis para os valores amostrados. Quanto maior o círculo maior é o valor observado das espécies *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica*.

4.1 Seleção de Covariáveis

Foram utilizados dois conjuntos para o ajuste de modelos, um composto pelas covariáveis topográficas e edáficas e ao outro foi acrescentada a soma da área basal das espécies restantes, como uma covariável adicional, com o intuito de verificar se, além das covariáveis do solo, as outras espécies restantes também influenciam nas espécies em estudo. Em seguida, para cada modelo ajustado, fez-se a seleção de covariáveis utilizando o método *stepwise* e prosseguiu-se com as análises restantes, utilizando-se as covariáveis pré selecionadas pelo *stepwise*.

Os resultados da análise de regressão múltipla para as espécies *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica* estão nas Tabelas 2, 3 e 4, respectivamente, utilizando um nível de significância de 20%. As covariáveis que estão nas Tabelas 2, 3 e 4 são as covariáveis selecionadas pelo método *stepwise*.

Tabela 2 - Estimativas dos parâmetros, erro padrão e valor p das covariáveis para a espécie *Copaifera langsdorffii* pelo método regressão linear múltipla

Covariáveis	Estimativa	Erro padrão	Valor p
Intercepto	0,57	0,23	0,01
P 5-25	-0,02	0,01	0,09
Na 0-5	-0,06	0,03	0,07
K 0-5	-0,04	0,02	0,03
K 80-100	0,11	0,04	<0,01
Ca 0-5	-0,02	0,01	0,07
Ca 5-25	-0,03	0,02	0,05
Mg 0-5	-0,03	0,01	0,01
Al 80-100	0,02	0,01	0,04
CTC 0-5	0,01	0,01	<0,01
CTC 5-25	-0,01	0,01	0,10
V 0-5	0,02	0,01	<0,01
m 5-25	-0,01	0,01	<0,01
S 80-100	0,09	0,07	0,17
Declividade	-0,05	0,02	0,05
Altitude	0,04	0,02	0,08

No caso da espécie *Copaifera langsdorffii* (Tabela 2), as covariáveis selecionadas pelo *stepwise*, para os dois conjuntos de modelos, foram idênticas.

Tabela 3 - Estimativas dos parâmetros, erro padrão e valor p das covariáveis para a espécie *Vochysia tucanorum* pelo método regressão linear múltipla

Sem a covariável SAB-Restante				Com a covariável SAB-Restante			
Covariáveis	Estimativa	Erro padrão	valor p	Covariáveis	Estimativa	Erro padrão	valor p
Intercepto	-0,47	0,24	0,05	Intercepto	0,53	0,34	0,12
pH H ₂ O 0-5	0,08	0,036	0,04	SAB-Restante	-0,30	0,05	<0,01
P 5 -25	-0,03	0,01	<0,01	pH H ₂ O 0-5	0,06	0,03	0,09
Ca 5-25	0,02	0,01	0,16	P 0-5	0,01	0,01	0,02
V 5-25	-0,01	0,01	<0,01	P 5-25	-0,05	0,01	<0,01
V 80-100	0,01	0,01	<0,01	Na 5-25	0,45	0,27	0,10
m 0-5	0,01	0,01	<0,01	V 5-25	-0,02	0,01	<0,01
S 5-25	0,14	0,07	0,05	V 80-100	0,02	0,01	<0,01
Declividade	0,06	0,03	0,03	m 0-5	0,01	0,01	<0,01
Altitude	-0,04	0,03	0,10	m 5-25	-0,01	0,01	0,03
				Declividade	0,05	0,02	0,02
				Altitude	-0,04	0,03	0,15

Tabela 4 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis para a espécie *Xylopia aromatica* pelo método de regressão linear múltipla

Sem a covariável SAB-Restante				Com a covariável SAB-Restante			
Covariáveis	Estimativa	Erro padrão	valor p	Covariáveis	Estimativa	Erro padrão	valor p
Intercepto	-0,03	0,15	0,86	Intercepto	1,32	0,35	<0,01
pH H ₂ O 5-25	-0,06	0,02	<0,01	SAB-Restante	-0,18	0,02	<0,01
pH H ₂ O 80-100	0,05	0,03	0,08	pH H ₂ O 5-25	-0,06	0,02	<0,01
MO 0-5	-0,01	0,01	0,02	MO 0-5	-0,01	0,01	<0,01
MO 5-25	-0,01	0,01	0,04	MO 80-100	0,01	0,01	0,06
MO 80-100	0,01	0,01	0,05	P 0-5	0,01	0,01	<0,01
P 0-5	0,01	0,01	0,02	P 5-25	-0,01	0,01	0,07
K 0-5	-0,02	0,01	0,03	K 0-5	-0,03	0,01	<0,01
Ca 0-5	-0,01	0,01	<0,01	Ca 0-5	-0,01	0,01	0,5
Ca 5-25	0,02	0,01	0,02	Ca 5-25	0,02	0,01	0,06
Al 0-5	-0,01	0,01	0,10	Al 5-25	0,01	0,01	0,13
CTC 0-5	0,01	0,01	<0,01	CTC 0-5	0,01	0,01	<0,01
CTC 80-100	0,01	0,01	<0,01	CTC 5-25	-0,01	0,01	0,01
S 5-25	0,09	0,04	0,02	CTC 80-100	0,01	0,01	0,08
				V 5-25	-0,02	0,01	<0,01
				m 5-25	-0,01	0,01	0,01
				m 80-100	-0,01	0,01	0,08
				S 5-25	0,07	0,04	0,06

Para as espécies *Vochysia tucanorum* e *Xylopia aromatica*, (Tabela 3 e 4), as covariáveis selecionadas pelo *stepwise*, para os dois conjuntos de modelos, foram distintas, e considerou significativa a 20%.

Os métodos GAM, RM e TREE têm o objetivo de verificar as covariáveis que são mais importantes para o modelo. Considera-se que as covariáveis significativas em todos os métodos são realmente relevantes para a escolha do modelo. Optou-se por prosseguir com as covariáveis que foram significativas em pelo menos dois métodos.

A Tabela 5 indica as covariáveis selecionadas pelos métodos de regressão linear múltipla (RM), modelos aditivos generalizados (GAM) e árvore de regressão (TREE). As variáveis indicadas com "X" foram incluídas com significância de 20% no caso da regressão linear múltipla e modelo aditivo generalizado.

Tabela 5 - Covariáveis selecionadas em cada método (RM, GAM e TREE), para os dois conjuntos de modelos

Espécies	Sem a covariável SAB-Restante				Com a covariável SAB-Restante			
	Covariáveis	RM	GAM	TREE	Covariáveis	RM	GAM	TREE
<i>Copaifera langsdorffii</i>	P 5-25	X						
	Na 0-5	X	X		Na 0-5	X	X	
	K 0-5	X	X	X	K 0-5	X	X	X
	K 80-100	X	X	X	K 80-100	X	X	X
	Ca 0-5	X	X	X	Ca 0-5	X	X	X
	Ca 5-25	X	X		Ca 5-25	X	X	
	Mg 0-5	X	X	X	Mg 0-5	X	X	X
	Al 80-100	X	X	X	Al 80-100	X	X	X
	CTC 0-5	X	X	X	CTC 0-5	X	X	X
	CTC 5-25	X	X	X	CTC 5-25	X	X	X
	V 0-5	X	X		V 0-5	X	X	
	m 5-25	X	X	X	m 5-25	X	X	X
	S 80-100	X	X	X	S 80-100	X	X	X
	Declividade	X	X	X	Declividade	X	X	X
	Altitude	X	X	X	Altitude	X	X	X
<i>Vochysia tucanorum</i>	pH H ₂ O 0-5	X	X		SAB-Restante	X	X	X
	P 5-25	X	X	X	pH H ₂ O 0-5	X	X	X
	Ca 5-25	X	X		P 0-5	X	X	X
	V 5-25	X	X	X	P 5-25	X	X	X
	V 80-100	X	X	X	Na 5-25	X	X	
	m 0-5	X	X	X	V 5-25	X	X	
	S 5-25	X	X	X	V 80-100	X	X	X
	Declividade	X	X	X	m 0-5	X	X	X
	Altitude	X	X		m 5-25	X	X	
					Declividade	X	X	X
<i>Xylopia aromatica</i>					Altitude	X	X	
	pH H ₂ O 5-25	X	X	X	SAB-Restante			X
	pH H ₂ O 80-100	X	X		pH H ₂ O 5-25	X	X	X
	MO 0-5	X	X	X	MO 0-5	X	X	X
	MO 5-25	X	X	X	MO 80-100	X	X	X
	MO 80-100	X	X	X	P 0-5	X	X	X
	P 0-5	X	X	X	P 5-25			X
	K 0-5	X	X		K 0-5	X	X	X
	Ca 0-5	X	X		Ca 0-5	X	X	X
	Al 0-5	X	X	X	Ca 5-25	X	X	X
	CTC 0-5	X	X	X	Al 5-25		X	X
	CTC 80-100	X	X		CTC 0-5	X	X	X
	S 5-25	X	X	X	CTC 5-25	X	X	X
					CTC 80-100	X		
					V 5-25	X	X	X
					m 5-25	X	X	X
					m 80-100	X		
					S 5-25	X	X	X

Observando a Tabela 5, a covariável P 5-25 foi retirada do modelo para a espécie *Copaifera langsdorffii*, no caso da espécie *Vochysia tucanorum* nenhuma covariável

foi removida e as covariáveis SAB-Restante, P 5-25, CTC 80-100 e m 80-100 foram retiradas do modelo para a espécie *Xylopia aromatica*, pois não foram significativas em pelo menos dois métodos.

Com o método modelo aditivo generalizado é possível verificar graficamente se as covariáveis tem algum termo (polinômio) de segundo ou terceiro grau em relação a variável resposta. Então, nas covariáveis que mostraram a existencia de efeitos não lineares, optou-se por acrescentar os polinômios de até terceiro grau.

Para a escolha do melhor modelo, primeiramente compara-se os dois conjuntos de modelos descritos no início desta seção. Em seguida, com o modelo escolhido, aplica-se o *stepwise*, verificando se é necessário remover mais alguma covariável. Se sim, compara-se os dois modelos pelo critério de Akaike, os valores estão na Tabela 6.

Tabela 6 - AIC dos modelos de regressão linear múltipla antes e depois do *stepwise*

Espécies	SAB _{Restante}	Covariáveis incluídas	AIC
<i>Copaifera</i>	Sem	Na ₀₋₅ , K ₀₋₅ , K ₈₀₋₁₀₀ , Ca ₀₋₅ ³ , Ca ₀₋₅ ² , Ca ₀₋₅ , Ca ₅₋₂₅ , Mg ₀₋₅ , Al ₈₀₋₁₀₀ , CTC ₀₋₅ , CTC ₅₋₂₅ , V ₀₋₅ , m ₅₋₂₅ , S₈₀₋₁₀₀ , Declividade, Altitude ² , Altitude	-321,48 (M1)
		Na ₀₋₅ , K ₀₋₅ , K ₈₀₋₁₀₀ , Ca ₀₋₅ ³ , Ca ₀₋₅ ² , Ca ₀₋₅ , Ca ₅₋₂₅ , Mg ₀₋₅ , Al ₈₀₋₁₀₀ , CTC ₀₋₅ , CTC ₅₋₂₅ , V ₀₋₅ , m ₅₋₂₅ , Declividade, Altitude ² , Altitude	-323,48 (M2)
	Com		
<i>Vochysia</i>	Sem	pH H ₂ O ₀₋₅ , P ₅₋₂₅ , K ₀₋₅ , Ca ₅₋₂₅ , V ₅₋₂₅ , V ₈₀₋₁₀₀ , m ₀₋₅ , S ₅₋₂₅ , Declividade, Altitude	-267,41 (M3)
	Com	SAB _{Restante} ³ , SAB _{Restante} ² , SAB _{Restante} , pH H ₂ O ₀₋₅ , P ₀₋₅ , P ₅₋₂₅ , Na₅₋₂₅ , V ₅₋₂₅ , V ₈₀₋₁₀₀ , m ₀₋₅ , m ₅₋₂₅ , Declividade, Altitude	-306,57 (M4)
		SAB _{Restante} ³ , SAB _{Restante} ² , SAB _{Restante} , P ₀₋₅ , P ₅₋₂₅ , V ₅₋₂₅ , V ₈₀₋₁₀₀ , m ₀₋₅ , m ₅₋₂₅ , Declividade, Altitude	-310,01 (M5)
<i>Xylopia</i>	Sem	pH H ₂ O ₅₋₂₅ , pH H ₂ O ₈₀₋₁₀₀ , MO ₀₋₅ , MO ₅₋₂₅ , MO ₈₀₋₁₀₀ , P ₀₋₅ , K ₀₋₅ , Ca ₀₋₅ , Ca ₅₋₂₅ , Al ₀₋₅ , Al ₀₋₅ ² , CTC ₀₋₅ , CTC ₈₀₋₁₀₀ , S ₅₋₂₅	-553,14 (M6)
	Com	pH H ₂ O ₅₋₂₅ , MO ₀₋₅ , MO ₈₀₋₁₀₀ , P ₀₋₅ ³ , P ₀₋₅ ² , P ₀₋₅ , K ₀₋₅ , Ca ₀₋₅ , Ca₅₋₂₅ , CTC ₀₋₅ , CTC₅₋₂₅ , CTC₈₀₋₁₀₀ , V₅₋₂₅² , V₅₋₂₅ , m₅₋₂₅ , m₈₀₋₁₀₀ , S₅₋₂₅	-537,69 (M7)
		pH H ₂ O ₅₋₂₅ , MO ₀₋₅ , MO ₈₀₋₁₀₀ , P ₀₋₅ ³ , K ₀₋₅ Ca ₀₋₅ , CTC ₀₋₅	-542,00 (M8)

Os melhores modelos para cada espécie são aqueles que tiveram o menor valor do AIC na Tabela 6. Para a espécie *Copaifera langsdorffii* o melhor modelo foi M2, para a *Vochysia tucanorum* foi o M5 e para a espécie *Xylopia aromatica* o M6. As covariáveis que estão em negrito são as que foram removidas após o *stepwise*. Com o melhor modelo, verifica-se a existência de autocorrelação espacial nos resíduos pelo índice de Moran. Na Tabela 7 estão os valores p do índice de Moran, para os modelos M2, M5 e

M6 de cada espécie.

Tabela 7 - Valor p do Índice de Moran para cada espécie

Espécies	Valor do Índice de Moran	Valor p
<i>Copaifera langsdorffii</i>	0,24	< 0,01
<i>Vochysia tucanorum</i>	0,21	< 0,01
<i>Xylopia aromatica</i>	0,49	< 0,01

Pela Tabela 7, o índice de Moran foi significativo para todas as espécies. Assim, todas têm autocorrelações espaciais, justificando a necessidade de inclusão de um termo no modelo para capturar essa dependência.

Figura 6 - Variograma empírico e envelope simulado para verificação de dependência espacial das espécies: *Copaifera langsdorffii* (a), *Vochysia tucanorum* (b), *Xylopia aromatica* (c)

A dependência espacial pode ser verificada por meio do gráfico utilizando envelopes simulados e variogramas. Observando a Figura 6, alguns pontos estão fora do envelope simulado, indicando assim, uma dependência espacial para as três espécies. Para que haja dependência, deve haver ao menos um ponto do variograma fora do envelope simulado. Se isso ocorrer, rejeita-se H_0 de que não há dependência espacial (DIGGLE, RIBEIRO JR, 2007) e as análises clássicas podem ser substituídas por análises espaciais, que levam em consideração as relações entre observações vizinhas, ou seja, dependência espacial.

Para cada espécie, comporou-se graficamente os valores preditos dos modelos M2, M5 e M6 com o acréscimo do parâmetro espacial. Os gráficos contendo valores preditos do modelo M2 com o acréscimo do parâmetro espacial (CAR, SAR e geoestatístico) para a espécie *Copaifera langsdorffii* estão na Figura 7.

Figura 7 - Valores preditos do modelo M2 com o termo espacial (CAR, SAR e Geoestatístico) para a espécie *Copaifera langsdorffii*

Para a espécie *Vochysia tucanorum*, os gráficos contento os valores preditos do modelo M5 com o acréscimo do parâmetro espacial, estão na Figura 8.

Figura 8 - Valores preditos do modelo M5 com o termo espacial (CAR, SAR e Geoestatístico) para a espécie *Vochysia tucanorum*

Os gráficos com os valores preditos no modelo M6 com acréscimo do parâmetro espacial (CAR, SAR e Geoestatístico) da espécie *Xylopiia aromatica*, estão na Figura 9.

Figura 9 - Valores preditos do modelo M6 com o termo espacial (CAR, SAR e Geoestatístico) para a espécie *Xylopia aromatica*

Observando as Figuras 7, 8 e 9, verifica-se que os modelos (CAR, SAR e geoestatístico) são diferentes, mas estão indo para a mesma direção, com isso, pode-se inferir que os resultados estão correlacionados, mas não são equivalentes, ou seja, os métodos têm diferenças. Por isso, compará-los e verificar qual deles se ajustou melhor é interessante, uma vez que diferentes modelos irão indicar diferentes resultados. Na Tabela 8 estão todos os modelos com e sem parâmetro espacial, para assim verificar o melhor modelo, comparando-os pelo AIC:

Tabela 8 - Valores do AIC para os modelos RM, CAR, SAR e geoestatístico, utilizando as covariáveis dos modelos M2, M5 e M6

Espécies	Método	Covariáveis incluídas	parâmetros	AIC
<i>Copaifera</i>	RM	$Na_{0-5}, K_{0-5}, K_{80-100}, Ca_{0-5}^3, Ca_{0-5}^2$	17	-323,48
	SAR	$Ca_{0-5}, Ca_{5-25}, Mg_{0-5}, Al_{80-100}$	19	-348,04
	CAR	$CTC_{0-5}, CTC_{5-25}, V_{0-5}, m_{5-25}$	19	-348,92
	GEO	Declividade, $Altitude^2$, Altitude	20	-349,23
<i>Vochysia</i>	RM	$SAB_{Restante}^3, SAB_{Restante}^2, SAB_{Restante}$	12	-310,01
	SAR	$Ca_{0-5}, Ca_{5-25}, CTC_{0-5}, CTC_{5-25}$	14	-332,23
	CAR	$CTC_{80-100}, m_{0-5}, m_{5-25}, Declividade$	14	-340,97
	GEO	Altitude	15	-621,15
<i>Xylopia</i>	RM	$pH\ H_2O_{5-25}, pH\ H_2O_{80-100}, MO_{5-25}$	14	-553,15
	SAR	$MO_{0-5}, MO_{80-100}, P_{0-5}, K_{0-5}$	16	-729,03
	CAR	$Ca_{0-5}, Ca_{5-25}, Al_{0-5}, Al_0 - 5^2$	16	-755,80
	GEO	$CTC_{0-5}, CTC_{80-100}, S_{5-25}$	17	-753,10

Verifica-se que o melhor modelo para as espécies *Copaifera langsdorffii* e *Vochysia tucanorum* é o geoestatístico e para a *Xylopia aromatica* é o autoregressivo (CAR). Observa-se que a diferença do valor AIC do modelo CAR e do modelo geoestatístico foi pequena.

Na Tabela 9 está o modelo geoestatístico com as estimativas dos parâmetros das covariáveis que foram significativas a 5%.

Tabela 9 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis do modelo geoestatístico para a espécie *Copaifera langsdorffii*

Espécies	Modelo	Covariáveis	Estimativa	Erro padrão	valor p
<i>Copaifera</i>	GEO	Na ₀₋₅	-0,05	0,03	0,04
		K ₀₋₅	-0,02	0,02	0,1
		K ₈₀₋₁₀₀	0,06	0,04	0,05
		Ca ₀₋₅	-0,8	0,34	0,4
		Ca ₀₋₅ ²	-0,22	0,13	0,05
		Ca ₀₋₅ ³	0,35	0,11	<0,1
		Ca ₅₋₂₅	-0,02	0,01	0,1
		Mg ₀₋₅	-0,02	0,01	0,01
		Al ₈₀₋₁₀₀	0,01	0,01	0,02
		CTC ₀₋₅	0,01	0,01	0,01
		CTC ₅₋₂₅	-0,01	0,01	<0,01
		V ₀₋₅	0,01	0,01	0,04
		m ₅₋₂₅	-0,01	0,01	0,01
		Declividade	-0,02	0,01	0,05
		Altitude	-0,36	0,23	0,05
		Altitude ²	-0,41	0,21	0,02

A covariável CTC foi significativa para a espécie *Copaifera langsdorffii*. Solos com maior CTC apresentam maior capacidade de manter os nutrientes na superfície do solo diminuindo a lixiviação (processo de extração de uma substância presente em componentes sólidos através da sua dissolução num líquido). Além disso, esta variável é calculada a partir de mais variáveis químicas e, portanto, é, de alguma forma, um resumo destas variáveis.

No modelo final da espécie *Copaifera langsdorffii* existem elementos químicos essenciais para a sobrevivência de todas as plantas, ditos macronutrientes, que são absorvidos em grande quantidade pela planta. Os macronutrientes que estão nesse modelo são o potássio, o cálcio e o magnésio, também conhecido como cátions básicos, importantes no processo de nutrição das plantas e a interação entre eles é fundamental, em termos de favorecer ou inibir a absorção pelas plantas. Saturação por bases é muito importante para conhecer o nível de fertilidade do solo.

A saturação por alumínio (m%) diminui a partir do momento em que se aumentam os níveis de Ca, Mg e K no solo, pois o aumento de teores de outros cátions faz com que diminua a % do cátion Al. Pode-se relatar também, que as plantas exigentes são menos tolerantes ao alumínio, que por sua vez é um cátion tóxico às células das plantas.

Outra covariável presente no modelo para a espécie *Copaifera langsdorffii* é

a declividade, que quanto maior, maior será o risco de erosão do solo, havendo com isso o empobrecimento do mesmo devido ao carreamento da camada superior e mais fértil. Esse solo com declividade tende a acumular água nos locais de menores altitudes. Pode-se inferir que talvez, nesses locais, ocorram maiores concentrações de nutrientes e maior densidade das árvores mais exigentes (BOARETTO, 2012).

Tabela 10 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis significativas do modelo geoestatístico para a espécie *Vochysia tucanorum*

Espécies	Modelo	Covariáveis	Estimativa	Erro padrão	valor p
<i>Vochysia</i>	GEO	$SAB_{Restante}$	-0,64	0,13	<0,01
		$SAB^2_{Restante}$	-0,04	0,12	0,36
		$SAB^3_{Restante}$	0,41	0,11	<0,01
		P_{0-5}	0,01	0,01	<0,01
		P_{5-25}	-0,04	0,01	<0,01
		V_{5-25}	-0,01	0,01	<0,01
		V_{80-100}	0,01	0,01	<0,01
		m_{0-5}	0,01	0,01	0,17
		m_{5-25}	-0,01	0,01	0,01
		Declividade	-0,02	0,01	0,01
		Altitude	0,01	0,01	0,04

No caso da espécie considerou também um nível de significância de 5% *Vochysia tucanorum* uma das covariáveis do modelo é a soma das áreas basais das espécies restantes e o único elemento macronutriente é o cálcio.

A covariável altitude pode influenciar nos fatores climáticos (temperatura, chuva, umidade relativa, luminosidade, entre outros) que consequentemente, afetarão o crescimento e a produção das espécie. Em altitudes maiores, há altas intensidades luminosas, provocando queimaduras nas folhas. No caso da declividade, que representa o grau de inclinação que o terreno exerce, influencia sobre a erosão. Os locais com maior declividade são mais suscetíveis à erosão. O comprimento do declive e o grau dele no terreno é influenciado pela velocidade com que a água escorre (BOARETTO, 2012).

Tabela 11 - Estimativa dos parâmetros, erro padrão e valor p das covariáveis do modelo CAR para a espécie *Xylopi aromatic*

Espécie	Modelo	Covariáveis	Estimativa	Erro padrão	valor p
<i>Xylopi</i>	CAR	pH H ₂ O ₅₋₂₅	0,01	0,01	0,84
		pH H ₂ O ₈₀₋₁₀₀	0,01	0,01	0,97
		MO ₀₋₅	-0,01	0,02	0,05
		MO ₅₋₂₅	-0,01	0,01	0,05
		MO ₈₀₋₁₀₀	0,01	0,01	0,99
		P ₀₋₅	0,01	0,01	0,54
		K ₀₋₅	0,01	0,01	0,05
		Ca ₀₋₅	-0,01	0,01	0,16
		Ca ₅₋₂₅	-0,01	0,01	0,04
		Al ₀₋₅	-0,02	0,01	0,25
		Al ₀₋₅ ²	-0,01	0,01	0,69
		CTC ₀₋₅	0,01	0,01	0,24
		CTC ₈₀₋₁₀₀	-0,01	0,01	0,86

No modelo da espécie *Xylopi aromatic* têm-se, como covariáveis significativas a 5%, o fósforo, potássio e cálcio, todos macronutrientes. O fósforo é um dos três macronutrientes exigidos em menores quantidade pelas plantas. Entretanto, é o nutriente mais empregado na adubação das culturas no Estado de São Paulo (BOARETTO, 2012). Isso ocorre porque ele interage fortemente com as partículas do solo, o que torna o fósforo indisponível ao vegetal.

A matéria orgânica (MO), refere-se a todo material resultante da decomposição de animais e vegetais, é de grande importância nos solos tropicais e subtropicais, pois influencia nas propriedades químicas, físicas e biológicas do solo, na capacidade de troca de cátions, por apresentar cargas (geralmente negativas). Quanto mais matéria orgânica maior a CTC do solo.

Tabela 12 - Valores dos parâmetros dos modelos para as espécies *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopi aromatic*

Espécies	Modelo	σ^2	ϕ	λ
<i>Copaifera langsdorffii</i>	Geoestatístico	0,01	20,66	
<i>Vochysia tucanorum</i>	Geoestatístico	0,01	61,15	
<i>Xylopi aromatic</i>	CAR			0,99

O valor de λ para o modelo CAR na Tabela 12 indica que existe autocorrelação, sendo que este valor varia de 0 à 1, em que para 0 não existe autocorrelação espacial, ou seja, o termo de erro não é espacialmente correlacionado.

Frade (2013) obteve no seu modelo final as covariáveis MO_{0-5} , MO_{80-100} , P_{0-5} , declividade e altitude, que foram selecionadas nas três espécies (*Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica*) e, neste trabalho, nenhuma covariável repetiu em todas as espécies, algumas covariáveis como m_{5-25} , declividade e altitude, foram significativas em duas espécies.

4.2 Análise Fatorial

A adequação da análise fatorial foi determinada pelo Kaiser-Meyer-Olkin (KMO). O teste de KMO teve um valor de 0,76, indicando que as variáveis estão correlacionadas e a análise fatorial apresentou um bom nível de adequação aos dados. 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.

Na Tabela 13 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 13, pode-se observar que o primeiro fator possui pesos mais altos nas covariáveis K_{0-5} , 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 S_{0-5} , S_{5-25} , S_{80-100} , $X2_{0-5}$, $X2_{5-25}$ e $X2_{80-100}$. O quarto fator extraído é composto pelas covariáveis MO_{0-5} , P_{0-5} , K_{0-5} e CTC_{0-5} . No quinto fator, as covariáveis foram $pH\ H_2O_{80-100}$ e V_{80-100} . As covariáveis $pH\ H_2O_{5-25}$ e Al_{5-25} compõem o sexto fator. O sétimo fator é composto apenas pela covariável Sódio na profundidade 0-5. Por fim, o oitavo fator é apenas pela covariável Altitude.

Os fatores extraídos pela análise fatorial foram utilizados como covariáveis nos modelos RM, GAM, TREE, CAR, SAR e geoestatístico. Assim, como na análise com as covariáveis originais, foram aplicadas as metodologias regressão linear múltipla com *stepwise*, modelo aditivo generalizado e árvore de regressão para selecionar os fatores mais significativos.

Na Tabela 15, os fatores representados por um "X" foram significativas a 20% de significância para os métodos *stepwise*, RM e GAM.

Tabela 13 - Resultados da Análise Fatorial Análise (AF)

nomes	Comunalidade	carga 1	carga 2	carga 3	carga 4	carga 5	carga 6	carga 7	carga 8	fator
pH H ₂ O ₅₋₂₅	0,77	0,00	0,05	0,14	0,15	-0,52	0,65	0,08	-0,15	6
pH H ₂ O ₈₀₋₁₀₀	0,76	-0,15	-0,01	0,11	0,01	-0,82	0,01	-0,22	0,03	5
MO ₀₋₅	0,76	0,01	-0,10	0,29	0,81	0,01	0,03	-0,08	-0,01	4
P ₀₋₅	0,75	0,02	-0,37	-0,05	0,73	0,15	-0,17	-0,12	-0,14	4
Na ₀₋₅	0,80	-0,11	-0,06	0,04	0,26	-0,13	0,01	-0,83	0,05	7
K ₀₋₅	0,68	-0,31	-0,46	-0,03	0,58	-0,03	0,08	0,05	-0,14	4
K ₅₋₂₅	0,73	-0,75	-0,19	0,06	0,24	0,00	-0,08	0,16	-0,20	1
K ₈₀₋₁₀₀	0,75	-0,80	0,02	0,14	0,14	-0,20	-0,13	0,08	-0,09	1
Ca ₀₋₅	0,84	-0,26	-0,73	0,22	0,32	-0,18	-0,18	-0,14	0,02	2
Ca ₅₋₂₅	0,84	-0,83	-0,05	0,23	0,05	-0,22	-0,10	-0,20	-0,01	1
Mg ₀₋₅	0,85	0,11	-0,84	0,06	0,34	0,10	-0,03	0,01	0,00	2
Al ₀₋₅	0,73	0,03	0,70	0,05	0,36	-0,10	-0,24	-0,19	0,06	2
Al ₅₋₂₅	0,81	-0,06	0,02	0,21	0,14	-0,13	-0,85	0,03	-0,04	6
CTC ₀₋₅	0,83	-0,28	-0,02	0,21	0,81	-0,09	-0,08	-0,20	0,02	4
CTC ₅₋₂₅	0,81	-0,51	0,03	0,36	0,39	-0,13	-0,44	0,14	-0,16	1
CTC ₈₀₋₁₀₀	0,67	-0,69	0,12	0,29	0,15	0,19	-0,13	-0,12	-0,09	1
V ₀₋₅	0,91	-0,04	-0,94	0,08	0,02	-0,07	-0,06	-0,10	0,00	2
V ₅₋₂₅	0,82	-0,79	-0,26	0,00	-0,12	0,01	0,20	-0,27	0,04	1
V ₈₀₋₁₀₀	0,71	-0,50	-0,09	-0,00	-0,08	-0,65	-0,09	0,09	0,07	5
m ₀₋₅	0,95	0,20	0,91	-0,12	-0,25	0,05	-0,06	0,04	0,03	2
m ₅₋₂₅	0,92	0,86	0,23	-0,12	-0,06	0,03	-0,30	0,09	0,06	1
m ₈₀₋₁₀₀	0,82	0,81	-0,05	-0,14	0,02	0,34	0,13	-0,07	0,04	1
S ₀₋₅	0,90	0,08	0,09	-0,80	-0,37	0,17	-0,09	-0,21	-0,19	3
S ₅₋₂₅	0,88	0,10	0,09	-0,78	-0,09	0,06	0,17	0,08	0,46	3
S ₈₀₋₁₀₀	0,81	0,24	0,05	-0,85	0,02	-0,03	0,06	0,11	0,06	3
Altitude	0,79	0,21	0,03	-0,16	-0,08	-0,05	-0,05	-0,07	0,84	8
X ₂₀₋₅	0,90	0,08	0,09	-0,80	-0,37	0,17	-0,09	-0,21	-0,19	3
X ₂₅₋₂₅	0,88	0,10	0,09	-0,78	-0,09	0,06	0,17	0,08	0,46	3
X ₈₀₋₁₀₀	0,81	0,24	0,05	-0,85	0,02	-0,03	0,06	0,11	0,06	3

Tabela 15 - Fatores significativos a 22% pelos métodos: STEP, RM, GAM e TREE

Variáveis	Método	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8
<i>Copaifera</i>	RM	X	X				X		X
	STEP	X	X				X		X
	GAM	X				X	X		X
	TREE	X	X	X	X		X	X	X
<i>Vochysia</i>	RM		X	X		X	X		
	STEP		X	X		X			
	GAM		X	X		X	X		
	TREE		X	X	X	X	X		X
<i>Xylopia</i>	RM	X	X	X			X	X	X
	STEP	X	X	X			X	X	X
	GAM	X	X	X			X	X	X
	TREE	X	X	X	X	X	X	X	X

Os fatores que se repetiram em pelo menos três métodos continuam no modelo. Antes de prosseguir, verifica-se se há necessidade de acrescentar um termo de segundo ou terceiro grau nos fatores. Pelo método GAM, é possível verificar, graficamente se as covariáveis tem algum termo (polinômio) de segundo ou terceiro grau em relação a variável resposta. Então, nos fatores que mostraram a existencia de efeitos não lineares, optou-se por acrescentar os polinômios de até terceiro grau.

Tabela 16 - Covariáveis pertencentes ao modelo de regressão linear múltipla para as espécies: *Copaifera langsdorffii*, *Vochysia tucanorum* e *Xylopia aromatica*

Espécies		Covariáveis
<i>Copaifera</i>	M1	Fator ₁ , Fator ₂ , Fator ₆ , Fator ₈
<i>Vochysia</i>	M2	Fator ₂ , Fator ₃ , Fator ₅ , Fator ₅ ² , Fator ₅ ³ , Fator ₆
<i>Xylopia</i>	M3	Fator ₁ , Fator ₂ , Fator ₃ , Fator ₃ ² , Fator ₆ , Fator ₇ , Fator ₈

Com os modelos M1, M2 e M3 verifica-se se existe a autocorrelação espacial pelo índice de Moran (Tabela 17).

Tabela 17 - O valor p do Índice de Moran para cada espécie

Espécie	Valor do Índice de Moran	valor p
<i>Copaifera langsdorffii</i>	0,31	< 0,01
<i>Vochysia tucanorum</i>	0,30	< 0,01
<i>Xylopia aromatica</i> ,	0,56	< 0,01

Observando a Tabela 17, conclui-se que, para todas as espécies, o índice de Moran foi significativo, ou seja, existe autocorrelação, justificando que os modelos podem ser estendidos para capturar tais padrões.

Figura 10 - Variograma empírico e envelope simulado para verificação de dependência espacial das espécies: *Copaifera langsdorffii* (a), *Vochysia tucanorum* (b), *Xylopia aromatica* (c)

A dependência espacial pode ser verificada por meio do gráfico, utilizando envelopes simulados e variogramas. Observando a Figura 10, alguns pontos estão fora do envelope simulado, indicando assim, uma dependência espacial para as três espécies, ou seja, as análises clássicas podem ser substituídas por análises espaciais.

Na Figura 11 tem-se, graficamente, os valores preditos dos modelos com parâmetro espacial para a espécie *Copaifera langsdorffii*.

Figura 11 - Comparação dos valores preditos dos modelos CAR, SAR e Geoestatístico com termo espacial para a espécie *Copaifera langsdorffii*

Para a espécie *Vochysia tucanorum*, os gráficos contendo os valores preditos dos modelos com parâmetro espacial, estão na Figura 12.

Figura 12 - Comparação dos valores preditos dos modelos CAR, SAR, e Geoestatístico com o termo espacial para a espécie *Vochysia tucanorum*

Os valores preditos dos modelos com parâmetro espacial da espécie *Xylopia aromatica* estão na Figura 13.

Figura 13 - Comparação dos valores preditos dos modelos CAR, SAR e Geoestatístico com termo espacial para a espécie *Xylopia aromatica*

Observando as Figuras 11, 12 e 13, conclui-se que os valores preditos estão correlacionados. É interessante comparar os modelos, pois, os métodos têm diferenças e assim, verifica-se qual ajusta-se melhor. Na Tabela 18, comparou-se os modelos com e sem parâmetro espacial pelo AIC.

Tabela 18 - Comparando os modelos RM, SAR, CAR e geoestatístico pelo AIC

Espécies	Método	Covariáveis incluídas	parâmetros	AIC
<i>Copaifera</i>	RM		5	-283,21
	SAR	Fator ₁ , Fator ₂ , Fator ₆ ,	7	-323,15
	CAR	Fator ₈	7	-331,12
	GEO		12	-337,04
<i>Vochysia</i>	RM		7	-267,66
	SAR	Fator ₂ , Fator ₃ , Fator ₅ ³ ,	9	-282,06
	CAR	Fator ₅ ² , Fator ₅ , Fator ₆	9	-295,78
	GEO		14	-312,31
<i>Xylopia</i>	RM	Fator ₁ , Fator ₂ , Fator ₃ ² ,	8	-536,26
	SAR	Fator ₃ , Fator ₆ , Fator ₇ ,	10	-741,61
	CAR	Fator ₈	10	-765,18
	GEO		15	-760,60

Analisando a Tabela 18, o modelo mais adequado para as espécies *Vochysia tucanorum* e *Xylopia aromatica* foi o modelo CAR e para a espécie *Copaifera langsdorffii*, o geoestatístico.

Tabela 19 - Estimativas dos parâmetros, erro padrão e valor p dos fatores significativas a 5% dos modelos geoestatístico e o CAR

Espécies	Modelo	Covariáveis	Estimativa	Erro padrão	valor p
<i>Copaifera</i>	GEO	Fator ₁	-0,02	0,01	0,06
		Fator ₂	-0,02	0,01	0,02
		Fator ₆	0,01	0,01	0,08
		Fator ₈	0,02	0,01	<0,01
<i>Vochysia</i>	CAR	Fator ₂	-0,01	0,01	0,67
		Fator ₃	-0,03	0,01	<0,01
		Fator ₅	-0,31	0,16	0,06
		Fator ₅ ²	-0,18	0,12	0,16
		Fator ₅ ³	-0,38	0,12	<0,01
		Fator ₆	0,01	0,01	0,32
<i>Xylopia</i>	CAR	Fator ₁	0,01	0,01	0,96
		Fator ₂	-0,01	0,01	0,97
		Fator ₃	-0,06	0,06	0,36
		Fator ₃ ²	-0,05	0,06	0,32
		Fator ₆	0,01	0,01	0,89
		Fator ₇	0,01	0,01	0,05
		Fator ₈	-0,01	0,01	0,02

As variáveis que apresentam maior peso nos fatores e que foram significativas no modelo geoestatístico a 5% para a espécie *Copaifera langsdorffii* são Ca₀₋₅, Mg₀₋₅, Al₀₋₅, V₀₋₅, m₀₋₅ e Altitude. No caso da *Vochysia tucanorum*, as variáveis significativas

no modelo CAR são S_{0-5} , S_{5-25} , S_{80-100} , X_{20-5} , X_{25-25} , X_{80-100} , pH H_2O_{80-100} e V_{80-100} . Para a *Xylopia aromatica* as com maior peso nos fatores são Altitude e Na_{0-5}

Tabela 20 - Valores dos parâmetros para os modelos geoestatístico e CAR

Espécies	Modelo	σ^2	ϕ	λ
<i>Copaifera langsdorffii</i>	GEO	0,01	48,84	
<i>Vochysia tucanorum</i>	CAR			0,92
<i>Xylopia aromatica</i>	CAR			0,99

Observando a Tabela 20, o valor de λ no modelo CAR indica que existe autocorrelação espacial, ou seja, o termo de erro não é espacialmente correlacionado.

Em síntese, é possível comparar os resultados obtidos utilizando tanto a análise fatorial com os resultados, como as covariáveis originais (nesta dissertação e na de Frade (2013)), em todos os casos, adicionando o efeito espacial aos modelos, verificou-se que existem padrões na variabilidade que não são capturados pelas covariáveis, e a que mais se repetiu, foi a Altitude.

5 CONSIDERAÇÕES FINAIS

Os valores preditos indicaram que os resultados obtidos com os diferentes modelos estão correlacionados mas não são equivalentes, diferenciando-se entre si. Por esta razão, é interessante comparar os modelos, verificar qual se ajustou melhor e se as diferenças alteram as conclusões práticas no contexto do estudo. Para as espécies *Copaifera langsdorffii* e *Vochysia tucanorum* o melhor modelo foi o geoestatístico, e para a espécie *Xylopia aromatica* foi o autoregressivo.

Com o modelo final para cada espécie, considerando as covariáveis originais, foram identificadas altitude e declividade como as que mais se repetiram neste trabalho e também no modelo final de Frade (2013). Portanto, por terem sido detectadas pela maioria dos métodos, estas covariáveis são consideradas para a formação das espécies de plantas nessa área.

No caso dos fatores, o modelo mais adequado para as espécies *Vochysia tucanorum* *Xylopia aromatica* foi o modelo condicional autoregressivo e para a espécie *Copaifera langsdorffii* o modelo geoestatístico. As covariáveis que tiveram maior peso no modelo, considerando os fatores, foram o cálcio, alumínio, pH H₂O, magnésio e saturação por alumínio.

Para continuidade dos estudos relacionados ao problema tratado nessa dissertação, deve-se aprofundar nas análises dos dados, na análise geoestatística e em dados de área e considerar outras variáveis ambientais para evidenciar os agrupamentos e a intensidade das relações ecológicas. Pode-se considerar a possibilidade de uma análise das quatro áreas abrangidas pelo projeto de parcelas permanentes.

Do ponto de vista metodológico, pode-se buscar investigar, para mais detalhes, as causas determinantes para as diferenças dos resultados entre os diversos métodos.

Neste estudo diferentes métodos não apresentaram uma mensagem consistente quais covariáveis seriam importantes, isto indica que essas covariáveis não foram capazes de explicar a área basal das espécies e a existência de padrão espacial indica que há outros fatores não capturados por essas covariáveis que podem contribuir para explicação da área basal das espécies consideradas.

O estudo ilustra que, para análises deste tipo, apenas um método não é recomendado. O uso de diferentes métodos, associados à identificação de padrões consistentes nos resultados, parece oferecer uma alternativa mais segura para se obter conclusões mais precisas.

REFERÊNCIAS

- ABBAD, G.; TORRES, C. V. Regressão múltipla stepwise e hierárquica em Psicologia Organizacional: aplicações, problemas e soluções. **Estudo de Psicologia**, Natal, v.7, n.spe, p.19-29, 2002.
- ASSUNÇÃO, R. M. **Estatística espacial com aplicações em epidemiologia economia e sociologia**. São Carlos: Associação Brasileira de Estatística, 2001. 131p.
- AZEVEDO, P. R. M. **Modelos de regressão linear**. Natal: EDUFRN, v.1, 202p. 1997.
- BAILEY, T. C.; GATTREL, T. C. **Interactive spatial data analysis**. London: Prentice Hall, 1995.
- BARROS, N. F.; SILVA, O. M.; PEREIRA, A.R.; BRAGA, J. M.; LUDWIG, A. Análise do crescimento de *Eucalyptus saligna* em solo cerrado sob diferentes níveis de N. P. e K. no Vale do Jequitinhonha, MG. **IPEF**, n.26, p.13-17, 1984.
- BELL, J. F. Application of classification trees to the habitat preference of upland birds. **Journal of Applied Statistics**. Abingdon, v.23, p.349-359, Aug. 1996.
- BERG, E. van den; SANTOS, F. A. M. dos. Aspectos da variação ambiental em uma floresta de galeria em Itutinga, MG, Brasil. **Ciência Floretal**, Santa Maria, v.13, n.2, p.83-98, 2003.
- BIVAND, R. **spdep**. Spatial dependence: weighting schemes, statistics and models, 2014. R package version 2014-01-31.
- BOARETTO, A. E. **Interpretação de resultados**. 2012. Disponível em: <<http://apostilas.cena.usp.br/Boaretto/CEN5715-2012/>>. Acesso em: 24 out. 2013.
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; Stone, C. J. **Classification and regression trees**. California: Wadsworth International Group, 1984. 358p.
- CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. **Análise espacial de áreas**. 2004. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf>>. Acesso em: 20 nov. 2013.
- CAMARERO, J. J.; GUTIÉRREZ, E.; FORTIN, M. Spatial pattern of subalpine Forest-alpine grassland ecotones in the Spanish Central Pyrenees. **Forest Ecology and Management**, Amsterdam, v.134, n.4, p.1-16, Sept. 2000.
- CAMARGO, E. C. G. **Desenvolvimento, Implementação e Teste de Procedimentos Geoestatísticos (Krigagem) no Sistema de Processamento de Informações Georeferenciadas (SPRING)**. 1997. 105p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.
- CAPRETZ, R. L. **Análise dos padrões espaciais de árvores em quatro formações florestais do Estado de São Paulo, através de análises de segunda ordem, como função K de Ripley**. 2004. 79p. Dissertação (Mestrado em Ecologia de Agroecossistemas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2004.

CHONG, G.; JUN, C. Performance of some variable selection methods when multicollinearity is present. **Chemometrics and intelligent laboratory systems**, Dordrecht, n.78, p.103-112, July, 2005.

CLARK, I. The semivariogram. Part I. **Eng. and Min. J**, v.1, n.1, p.90-94, 1979.

CLAUSET, L. R., SOARES, D. **Paisagem Paulista: áreas protegidas**. São Paulo: Empresa das Artes, 1999. 185p.

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. 98p. Dissertação (Mestrado em Botânica) - Universidade de São Paulo, 2010.

CONDIT, R.; ASHTON, P.; BAKER, P. Spatial patterns in the distribution of tropical tree species. **Science**, Washington, v.288, p.1414-1418, May. 2000.

CORREIA, J. C.; SHARMA, R. D. Produtividade do algodoeiro herbáceo em plantio direto no Cerrado com rotação de culturas. **Pesquisa Agropecuária Brasileira**, Brasília, Rio de Janeiro, v.39, n.1, p.41-46, jan. 2004.

COSTA, A. **Qualidade do solo, crescimento e desenvolvimento do milho sob diferentes sistemas de manejo na integração lavoura-pecuária**. 2005. 91p. Dissertação (Mestrado em Ciência do Solo) - Universidade do Estado de Santa Catarina, Lages - SC, 2005.

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

DIGGLE, P. J.; RIBEIRO JR., P. J. **Model-based geostatistics**. USA: Springer Series in Statistics, 2007. 241p.

EMBRAPA - **Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos**. Brasília: Embrapa Produção de Informação, 1999. 412p.

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

FARAWAY, J. J. Extending the Lienar Model with R. In: **Trees**. London: CRC, 2006, chap.13, p.278-306.

FRADE, D. D. R. **Relações entre variáveis ambientais e espécies florestais por metodologias de processos pontuais**. 2013, p.67. Dissertação (Mestrado Ciências) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2013.

FURTINI NETO, A. E.; SIQUEIRA, J. O.; CURTI, N.; MOREIRA, F. M. S. Fertilização em reflorestamento com espécies nativas. In: GONÇALVES, J. L. M.; BENEDETTI, V. **Nutrição e fertilização florestal**. Piracicaba: IPEF, 2000, p.351-383.

GAMA, F. G.; SANTOS, J. R.; FREITAS, C. C.; MURA, J. C. Simpósio Brasileiro de Sensoriamento Remoto , 12, 2005. **Aplicação de dados polarimétricos de radar aerotransportado (banda P) para estimativa de parâmetros estruturais de florestas tropicais**. Disponível em:<<http://marte.sid.inpe.br/col/ltid.inpe.br/sbsr/2004/11.21.10.06/doc/4413.pdf>>. Acesso

em: 24 de jan. 2014.

GOSLEE, S. C. Behavior of vegetation sampling methods in the presence of spatial autocorrelation. **Plant Ecology**, Dordrecht, v.187, n.187, p.203-212, Dec. 2006.

HASTIE, T.; TIBSHIRANI, R. Generalized additive models. **Statistical Science**, Hayward, v.1, n.3, p.297-318, 1986.

HASTIE, T.; TIBSHIRANI, R. Generalized Additive Models, **Chapman and Hall**, London, 1990, p.329.

ILLIAN, J.; PENTTINEN, A.; STOYAN, H.; STOYAN D. **Statistical Analysis and Modelling of Spatial Point Patterns**. Wiley, England. 534p. 2008.

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. 136p. Tese (Doutorado Ecologia Aplicada) - Escola Superior de Agricultura "Luiz de Queiroz". Universidade de São Paulo, Piracicaba, 2008.

KRAINSKI, E. T. **Estrutura de covariância de modelos espaciais para dados de áreas**. 2008. p.48. Dissertação (Mestrado em Estatística) - Universidade Federal, Minas Gerais, 2008.

KULBACK, S.; LEIBLER, R.A. On information and sufficiency. **The Annals of Mathematical Statistics**, Washington, v.22, n.1, p.79-86, Apr. 1951.

MACIEL, M. N. M.; BASTOS, P. C. O; WATRIN, O. S. 2007. Disponível em: <<http://www.alice.cnptia.embrapa.br/bitstream/doc/402021/1/17251732.pdf>>. Acesso em: 25 nov. 2013.

MATÉRN, B. **Spatial variation: stochastic models and their application to some problems in forest surveys and other sampling investigations**. Stockholm: **Skogsforsknings Institut**. (Skogsforsknings Institut. Meddelanden fran Statens Skogsforsknings Institut), v.49, n.5, p.144, 1960.

MATOS, F. D. A. **Estimativa de biomassa e carbono em floresta ombrófila densa de terra firme na Amazônia central Manaus - AM, Brasil, por meio de dados de satélites de média e alta resolução espacial**. 2006, p.93. Tese (Doutorado em Ciências Florestais - Área Manejo Florestal) - Universidade Federal do Paraná, Curitiba, 2006.

MELLO, R. L. **Geoestatística aplicada ao inventário florestal**. 2004. 110p. Tese (Doutorado em Recursos Florestais) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2004.

MINGOTI, S. A. Análise fatorial. In: **Análise de dados através de métodos de estatística multivariada**. Belo Horizonte: UFMG, 2007, cap.4, p.99-138.

MOEUR, M. Characterizing spatial patterns of trees using stem-mapped data. **Forest Science**, Bathesda, v.39, n4, p.756-775, 1993.

MOREIRA, S. G.; KIEHL, J. C.; PROCHNOW, L. L.; PAULETTI, V. Calagem em sistema de semeadura direta e efeitos sobre a acidez do solo, disponibilidade de nutrientes e produtividade

de milho e soja. **Revista Brasileira de Ciência do Solo**, Viçosa, v.25, p.71-81, 2001.

NEWBERY, D. McC; RENSHAW, E.; BRUNIG, E. F. Spatial pattern of trees in kerangas forest, sarawak. **Plant Ecology**, Dordrecht, v.25, n.2, p.77-89, 1986.

ODUM E. P. **Ecologia**. Rio de Janeiro. Editora Guanabara, p.434, 1988.

PIERI, F. A.; MUSSI, M. C. M.; MOREIRA, M. A. Óleo de copaifera (*Copaifera sp.*: histórico, extração, aplicações industriais e propriedade industriais e propriedade medicinais. **Revista Brasileira de Plantas Mediciniais**, Botucatu - SP, v.11, n.4, p.465-472, 2009.

REZENDE, L. L. **Reabilitação de campos ferruginosos degradados pela atividade minerária o quadrilátero Ferrífero**. 2010. 62p. Dissertação (Mestrado em Solos e Nutrição de Plantas) - Universidade Federal de Viçosa - MG, Viçosa, 2010.

RIBEIRO JR, P. J.; DIGGLE, P. J. **goeR**. Analysis of geostatistical data, 2013. R package version 1.7-4. Disponível em: <http://cran.r-project.org/web/packages/geoR/geoR.pdf>. Acesso em : 10 marc. 2013.

RODRIGUES, R.R. Relatório temático do projeto parcelas permanentes. Laboratório de Ecologia e Restauração Florestal, ESALQ/USP. Piracicaba, 2002. Disponível em: <http://www.lerf.esalq.usp.br/old/parcelas/relatorio1.pdf>. Acesso em: 4 fev. 2013.

SECCO, D.; REINERT, D. J.; REICHERT, J. M.; ROS, C. O. Produtividade de soja e propriedades físicas de um Latossolo submetido a sistemas de manejo e compactação. **Revista Brasileira de Ciência do Solo**, Viçosa, v.28, n.5, p.797-804, set/out. 2004.

SOUZA, O. M; BARBIN, D.; RIBEIRO JR, P. J.; STAPE, J. L. Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados de um ensaio de espaçamento florestal e delineamento sistemático tipo Leque. **Revista Árvore**, Viços-MG, v.32, n.3, p.499-509, 2008.

TEAM, R Development Core. **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: 20 ago. 2013. Acesso em: 18 ago. 2013.

THERNEAU, T.; ATKINSON, B; RIPLEY, B. **rpart**. Recursive Partitioning and Regression Trees, 2014. R package version 4.1-8. Disponível em: <http://cran.r-project.org/web/packages/rpart/rpart.pdf>. Acesso em : 05 abr. 2013.

VERONA, J. D. **Classificação e monitoramento fenológico foliar da cobertura vegetal na região da floresta nacional do Tapajós - Pará, utilizando dados multitemporais do sensor thematic mapper (TM) do Landsat**. 2003. p.139. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisa Espacial, São José dos Campos, 2003.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R.F.; ALVAREZ, V. H.; SCHAEFER, G. R. (Ed.). **Tópicos em ciência do solo**. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000. v.1, p.1-54.

WEISBERG, S. **Applied linear regression**. New York, 1985, p.227.

WALLER, L. A.; GOTWAY, C. A. **Applied Spatial Statistics for Public Health Data**.

Wiley: Wiley Series in Probability and Statistics, 2004, 494p.

WOOD, S. N. **Generalized Additive Models: An Introduction with R**. Chapman and Hall, 2006.

WOOD, S. N. **mgcv**. Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation, 2014, R package version 1.8-3. Disponível em: <http://cran.r-project.org/web/packages/mgcv/mgcv.pdf>. Acesso em: 07 jul. 2013.