WILIAN FRANÇA COSTA

Segmentação Multiresolução Variográfica Ótima

São Paulo
2016

WILIAN FRANÇA COSTA

Segmentação Multiresolução Variográfica Ótima

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do
Título de Doutor em Ciências

Área de Concentração:
Engenharia de Computação

Orientador: Prof. Dr. Antonio Mauro Saraiva

São Paulo
2016

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Costa, Wilian França

Segmentação multiresolução variográfica ótima / W. F. Costa -- versão corr.
- São Paulo, 2016.
111 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo.
Departamento de Engenharia de Computação e Sistemas Digitais.

1.Geoprocessamento 2.Geoestatística 3.Biodiversidade 4.Agricultura de
Precisão 5.Variografia I.Universidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Computação e Sistemas Digitais II.t.

*Este trabalho é dedicado à minha mãe, Marlene de França Santana
e a minha avó, Natália de França
mulheres de luta, que me ensinaram a enfrentar de frente todos os desafios desta vida,
com ética, respeito e sempre, sempre, com muito amor.*

Agradecimentos

Agradeço a todas as pessoas e instituições que contribuíram direta ou indiretamente para a realização deste trabalho, em especial:

A Escola Politécnica da Universidade de São Paulo e a Fundação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo auxílio financeiro.

Ao Laboratório de Automação Agrícola, principalmente ao prof. Dr. Antonio Mauro Saraiva pela oportunidade concedida e confiança depositada, pela participação em diversos projetos que me trouxeram muito conhecimento nestes últimos 6 anos.

Ao prof. Dr. Bruno Carvalho Albertini, por acreditar no meu trabalho, pela dedicação e coragem de enfrentar novos desafios. Que a vida nos traga um sopro de renovação, sempre!

À prof^a. Dr^a. Sybelle Belay por todo o auxílio e incentivo, sem o qual parte deste trabalho não seria possível.

Ao prof. Dr. Jose Paulo Molin pelo fornecimento dos dados agrícolas.

À todos os funcionários aqui representados por Edson de Souza, Suzano Bitencourt, Nilton Carmo, Lourdes Keiko. Obrigado pela força e amizade.

Aos colegas e amigos que adquiri do LAA, Leonardo Campos, Michel J. M. Bieleveld, Allan Koch Veiga, Ivairton Santos, João Ferreira Filho, Roberto Fray da Silva, Raphael Gasparian Chinchilla. Sem dúvida, sem vocês a jornada teria sido bem mais difícil. Obrigado pelos trabalhos que desenvolvemos juntos e pelos momentos compartilhados que, sem dúvida, marcaram um período muito feliz da minha jornada neste planeta.

À minha mãe, Marlene de França Santana, por sempre ter me incentivado a ir atrás dos meus sonhos. Te amo mãe.

À minha avó, exemplo pra toda vida. Um milagre a cada dia! Obrigado por tudo. Também te amo.

À minha querida irmã, Katia Regina França (*in memoriam*). Você sempre estará presente. Não disse que conseguiria? :)

“O impossível é sempre mais sedutor que o possível.”
(Jean Paul Sartre)

Resumo

O desenvolvimento de soluções que auxiliem na extração de informações de dados oriundos de sistemas de sensoriamento remoto e outras geotecnologias são essenciais em diversas atividades, por exemplo, a identificação de requisitos para o monitoramento ambiental; a definição de regiões de conservação; o planejamento e execução de atividades de verificação quanto ao cumprimento e uso do espaço; o gerenciamento de recursos naturais; a definição de áreas protegidas e ecossistemas; e o planejamento para aplicação e reposição de insumos agrícolas. Neste contexto, o presente trabalho apresenta um método para parametrizar um algoritmo segmentador Multiresolution, de forma que os segmentos obtidos sejam os maiores possíveis dentro de limites pré-estabelecidos de heterogeneidade para os dados avaliados. O método faz uso de variografia, uma ferramenta geoestatística que apresenta uma estimativa de quanto duas amostras variam em uma região espacial, de acordo com a distância relativa entre elas. Mostra-se também como a avaliação de múltiplos variogramas pode ser empregada na delimitação de regiões quando combinada a este algoritmo de segmentação, desde que os dados estejam dispostos em uma grade amostral regularmente espaçada. O método desenvolvido utiliza o efeito pepita estimado para os atributos dispostos em camadas sobrepostas e quantifica a segmentação em dois momentos (ou médias) para identificar o valor do parâmetro espacial ótimo a ser aplicado no segmentador. Apresenta-se, como exemplos de aplicabilidade do método, três casos típicos desta área: (i) definição de zonas de manejo para agricultura de precisão; (ii) seleção de regiões para estimativas de degradação ambiental na vizinhança de ponto de coleta/observação de espécies; e (iii) a identificação de regiões bioclimáticas que compõem uma Unidade de Conservação da biodiversidade.

Palavras-chave: algoritmos. segmentação. método. multiresolution. multidimensional. multivariado. gradientes espaciais. biogeografia. agricultura de precisão. geoestatística. bioclima. conservação da biodiversidade. mapas. nicho ecológico.

Abstract

Information extraction of data derived from remote sensing and other geotechnologies is important for many activities, e.g., the identification of environmental requirements, the definition of conservation areas, the planning and implementation of activities regarding compliance of correct land use; the management of natural resources, the definition of protected ecosystem areas, and the spatial planning of agricultural input reposition. This thesis presents a parameter optimisation method for the Multiresolution segmentation algorithm. The goal of the method is to obtain maximum sized segments within the established heterogeneity limits. The method makes use of variography, a geostatistical tool that gives a measure of how much two samples will vary in a region depending on the distance between each one of them. The variogram nugget effect is measured for each attribute layer and then averaged to obtain the optimal value for spatial segmentation with the Multiresolution algorithm. The segments thus obtained are superimposed on a regularly spaced sampled grid of georeferenced data to divide the region under study. To show the usefulness of this method, the following three case studies were performed: (i) the delineation of precision farming management zones; (ii) the selection of regions for environmental degradation estimates in the neighbourhood of species occurrence points; and (iii) the identification of bioclimatic regions that are present in biodiversity conservation units.

Keywords: algorithm. segmentation. method. multiresolution. multidimensional. multivariate. spatial gradient. biogeography. precision farming. geostatistics. bioclim. biodiversity conservation. maps. ecological niche.

Lista de ilustrações

Figura 1 – Seleção de algumas das 67 variáveis PCNM que possuem autovalores positivos	31
Figura 2 – Índice de Moran calculado para as 67 variáveis PCNM	32
Figura 3 – Mapas gerados para algumas das 279 variáveis PCNM	33
Figura 4 – Índice de Moran calculado para as 279 variáveis PCNM	34
Figura 5 – Elementos que compõem um variograma	35
Figura 6 – Modelos de funções comumente empregados em variogramas	36
Figura 7 – Exemplo de relação entre o efeito pepita e a dispersão espacial dos dados	50
Figura 8 – Exemplo de aplicação do primeiro momento para avaliação da segmentação	51
Figura 9 – Avaliação do Primeiro Momento para as variáveis PCNM 1, 2, 5, 10, 20, 50, 100, 150 e 1274	54
Figura 10 – Avaliação da curva PM para todas as variáveis PCNM geradas em uma grade de 20 x 20 pontos	55
Figura 11 – Fluxograma com o algoritmo utilizado para o calculo da curva PM . .	56
Figura 12 – Fluxograma com o algoritmo utilizado para o calculo da curva SM . . .	56
Figura 13 – Localização dos pontos amostrados no talhão analisado	59
Figura 14 – Mapas gerados para as camadas de dados interpolados por meio da Krigagem Universal	60
Figura 15 – Resultado do teste Broken Stick para seleção das componentes principais para os dados não-normalizados e normalizados	63
Figura 16 – Curvas para a medida de inércia do <i>K-means</i> para os dados não-normalizados e normalizados no subespaço PCA	67
Figura 17 – Grupos para o PCA <i>K-means</i>	68
Figura 18 – Curva para o Primeiro Momento obtida para os mapas interpolados . .	69
Figura 19 – Segmentos obtidos para o $sp_{opt} = 1364$	70
Figura 20 – Gráfico PCA para os dados do talhão analisado e atribuição dos pontos na grade aos segmentos FSMM obtidos	72
Figura 21 – Mapas com as regiões delimitadas pelos grupos PCA <i>K-means</i> e Segmentos FSMM	73
Figura 22 – Mapa com a localização dos pontos de observação das espécies	77
Figura 23 – Acidentes geográficos na região próxima ao ponto P27	81
Figura 24 – PCA executado nos pontos selecionados para cada <i>buffer</i>	82
Figura 25 – PCA executado nos pontos selecionados para cada segmento	83
Figura 26 – Carta Imagem do Parque da Serra da Canastra. Imagem CBERS de setembro de 2006	87

Figura 27 – As primeiras 16 das 19 variáveis avaliadas para a obtenção dos segmentos na Serra da Canastra	88
Figura 28 – Avaliação da Curva PM para a Serra da Canastra	88
Figura 29 – Avaliação da Curva SM para a Serra da Canastra	89
Figura 30 – Segmentos FSMM ($\text{segments.mask} \Rightarrow S = \{s_1, \dots, s_{10}\}$) e valores esperados ($E = \{e_1 = \text{bio1}, \dots, e_{19} = \text{bio19}\}$) para as primeiras 15 das 19 variáveis avaliadas	89
Figura 31 – Imagem de satélite da região da Serra da Canastra sobreposta com os segmentos obtidos e indicação de áreas de prioridade de ação pelo MMA	90
Figura 32 – Gráfico screeplot para análise de broken-stick para os dados dos segmentos	91
Figura 33 – Gráfico biplot PCA - segmentos para a Serra da Canastra	92
Figura 34 – Mapa com segmentos obtidos pela Segmentação Multiresolução Variográfica Ótima para a região da Serra da Canastra. Cada segmento identifica uma região com diferentes bioclimas, de acordo com as variáveis biocimáticas do Worldclim	93

Lista de tabelas

Tabela 1 – Importância dos componentes PCA: dados normalizados	63
Tabela 2 – Importância dos componentes PCA: dados normalizados	63
Tabela 3 – Carregamentos: dados não-normalizados	64
Tabela 4 – Carregamentos: dados normalizados	64
Tabela 5 – Importância dos componentes: dados normalizados, inseridos produtivi- dade e altitude para o talhão analisado	70
Tabela 6 – Carregamentos: dados normalizados, inseridos produtividade e altitude para o talhão	71
Tabela 7 – Pontos de amostragem e contagem de riqueza de espécies por revisão bibliográfica	76
Tabela 8 – Camadas Bioclim/Worldclim	78
Tabela 9 – Fator espacial ótimo resultante da execução do algoritmo proposto e respectivo raio da janela de avaliação para os ambientes	80

Lista de abreviaturas e siglas

FSMM	Função Segmentadora Multirresolução Multivariada
Geo-OBIA	<i>Geo Object Based Image Analysis</i>
GLS	<i>Generalized Least Squares</i>
MMA	Ministério do Meio Ambiente
ODS	Objetivos de Desenvolvimento Sustentável
PCA	<i>Principal Component Analysis</i>
PCNM	<i>Principal Coordinates of Neighbor Matrices</i>
PM	Primeiro Momento
PNUD	Programa das Nações Unidas para o Desenvolvimento
SIG	Sistema de Informações Geográficas
SM	Segundo Momento
SVM	<i>Support Vector Machine</i>
UC	Unidade de Conservação
UNECE	<i>United Nations Economic Commission for Europe</i>

Lista de símbolos

$\gamma(h)$	Função variograma para uma distância h escalar calculada para um determinado modelo para o caso isotrópico
$\gamma(\vec{h})$	Função variograma calculada para uma distância dada pelo vetor $\vec{h}$
$\hat{\gamma}(\vec{h})$	Função variograma experimental calculada pelos pontos incidentes na distância dada pelo vetor $\vec{h}$
$\mathbb{B}$	Função segmentadora multiresolução multivariada
μ	Média ou esperança
$\bar{x}$	Média para as amostras
s	Desvio padrão para as amostras
s^2	Variância para as amostras
Δ	Matriz de autovetores, na qual cada autovetor representa uma variável PCNM
λ_i	Período do autovetor unidimensional na posição i

Sumário

1	INTRODUÇÃO	15
1.1	Motivação	17
1.1.1	Delimitação de zonas de gerenciamento diferenciado em agricultura de precisão	18
1.1.2	Impactos causados pelas mudanças climáticas e degradação do meio ambiente	19
1.1.3	Delimitação de regiões para conservação ou recuperação	19
1.2	Objetivo	20
1.2.1	Objetivo específico	20
1.2.2	Passos necessários para a obtenção do objetivo	20
1.3	Contribuições da tese	21
1.4	Estrutura da tese	21
2	FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA	23
2.1	Variáveis regionalizadas e a geoestatística	23
2.1.1	Fenômeno espacial	24
2.1.2	Heterogeneidade, homogeneidade e escala espacial	25
2.1.3	Amostra e métodos de amostragem	25
2.2	Modelagem de estruturas espaciais	26
2.2.1	O índice de Moran para teste de autocorrelação espacial	27
2.3	Dados Artificiais - Variáveis Espaciais Baseadas em Autovetores	29
2.3.1	Imagens geradas por funções PCNM	29
2.3.2	Análise variográfica	32
2.3.3	Anisotropia	37
2.4	Análise de texturas	37
2.5	Segmentação de Imagens	38
2.6	Trabalhos correlatos	38
3	SEGMENTAÇÃO MULTIRESOLUÇÃO VARIOGRÁFICA ÓTIMA	41
3.1	Método proposto	41
3.1.1	Função Segmentadora Multiresolução Multivariada	42
3.1.2	Aspectos utilizados para caracterização da heterogeneidade espacial	44
3.1.3	Fluxo de processamento da função segmentadora multiresolução multivariada - FSMM ($\mathbb{B}$)	46
3.1.4	Avaliação variográfica de $\mathbb{B}$	47
3.1.5	Delimitação de regiões pelo valor máximo do Primeiro Momento	53
3.1.6	Caracterização de janela ótima para segmentação por valor máximo do Segundo Momento	54

4	APLICAÇÕES E RESULTADOS	57
4.1	Estudo de caso: identificação de zonas de manejo / gerenciamento diferenciado em Agricultura de Precisão	57
4.1.1	Dados e área de estudo	58
4.1.2	Análise de componentes principais (PCA)	59
4.1.3	Abordagens	65
4.1.4	Resultados	69
4.2	Estudo de caso: degradação ambiental para a comunidade de anuros na Mata Atlântica	74
4.2.1	Caracterização do problema	75
4.2.2	Dados empregados (anuros, áreas de estudo, bioclimáticos)	76
4.2.3	Abordagens	78
4.2.4	Estimativa de uso do solo	81
4.2.5	Resultados análise de degradação aplicando os segmentos	85
4.3	Estudo de caso: identificação de região bioclimática para a Unidade de Conservação da Serra da Canastra	85
4.3.1	Dados e área de estudo	86
4.3.2	Avaliações das curvas PM e SM	86
4.3.3	Obtenção dos segmentos	87
4.3.4	Análise dos dados via PCA	91
5	CONSIDERAÇÕES E CONCLUSÕES	94
5.1	Considerações	94
5.2	Conclusões	96
5.3	Avaliação dos resultados e desafios do estudo	97
5.4	Implicações e trabalhos futuros	99
	Referências	100

1 Introdução

A identificação de requisitos para o monitoramento ambiental, a definição de regiões de conservação, o planejamento e execução de atividades de verificação quanto ao cumprimento e uso do espaço, o gerenciamento de recursos naturais, a definição de áreas protegidas e ecossistemas e o planejamento para aplicação e reposição de insumos agrícolas são apenas algumas das atividades que necessitam do desenvolvimento de soluções que possam auxiliar na extração de informações de dados georreferenciados e oriundos de sistemas de sensoriamento remoto. Estes dados são importantes para o desenvolvimento de estudos globais e locais sobre biodiversidade, conservação da natureza, segurança alimentar, impactos causados por desflorestamento, monitoramento da desertificação, entre outras áreas de aplicação. Com a evolução das diversas geotecnologias - como o uso de imagens de satélite de alta resolução, tomadas aéreas feitas por veículos aéreos não tripulados (*drones*) e a combinação do uso de sensores cada vez mais baratos, miniaturizados e com maior poder de processamento - existe uma crescente demanda por técnicas que possam combinar as informações geradas por estas tecnologias, e que permitam a análise dos dados obtidos de modo a se atribuir um valor semântico ao que se está sendo observado (ex.: áreas em que determinada doença se manifestou em uma cultura ou região de um bioma) e que também facilitem a análise dos dados. Estas análises, algumas vezes, necessitam de critérios objetivos para divisão e identificação dos fenômenos espaciais envolvidos. Esta divisão pode ser feita pelo uso de algoritmos de segmentação ou agrupamento dos dados obtidos para as regiões em análise, por meio da combinação de imagens, dados de sensores, análises laboratoriais, entre outras fontes.

A evolução dos denominados Sistemas de Informações Geográficas (SIG), dos bancos de dados com suporte ao processamento de dados espaciais, dos softwares de processamento estatístico, aliada ao crescimento de comunidades de projetos de software livre e associada à adoção de padrões abertos para interoperabilidade de dados, levaram a um incremento significativo do uso da informação geoespacial nas mais diversas áreas do conhecimento nos últimos anos (2016). Como exemplos, o projeto R, o Sistema Gerenciador de Bancos de Dados Espaciais Postgres/PostGIS e o software para SIG QGIS, fornecem uma ampla gama de algoritmos e ferramentas para geoprocessamento, combinação e geração de dados georreferenciados, em seus mais diversos formatos. Esta evolução facilita o desenvolvimento de pesquisas para combinação dos dados proveniente de sensoriamento e facilita o desenvolvimento de novos métodos de síntese e geração de conhecimento. Neste contexto, esta pesquisa desenvolveu uma técnica para segmentação de regiões, que pode ser empregada em dados oriundos de interpolação de medições de sensores, de dados de amostra de solos obtidos por análises laboratoriais, ou de modo genérico, de dados provenientes

de modelos que representem variações de alguma grandeza numérica georreferenciada para as regiões em análise. Estes dados, quando combinados em camadas sobrepostas, fornecem as informações necessárias para a separação em sub-regiões, de acordo com as variações internas dos fenômenos mensurados, que são dispostos em gradientes espaciais. Estes gradientes representam a autocorrelação dos dados presentes em um fenômeno espacializado.

Os gradientes temporais e espaciais apresentam-se em diversos fenômenos na natureza. Das variações topográficas entre regiões até as variações de pH, temperatura e umidade do solo, os gradientes estão presentes na natureza em diferentes escalas. Tais fenômenos naturais são conceitos desenvolvidos pelo ser humano para melhorar sua percepção do mundo. Desde a antiguidade, filósofos como Heráclito já apresentavam a importância do *conceito* por meio da definição de que tudo é formado pela alternância entre seus opostos. Heráclito foi um dos primeiros a estabelecer o que chamava de lei universal da Natureza, definida como a luta permanente e construtiva entre conceitos interdependentes e opostos. Segundo ele, nada existe sem o seu oposto (KAHN, 1981, p. 152). O *conceito*, apesar de estranho em uma primeira leitura, pode auxiliar no entendimento da abordagem proposta neste trabalho. Como exemplo, pode-se imaginar a delimitação de regiões por meio de vales e elevados em uma certa zona de um mapa topográfico. Existe um limite a partir do qual o elevado deixa de existir, e a região passa a ser classificada como vale. Estes limites, que definem a extensão dos gradientes espaciais, são denominados como **segmentos** e a eles associamos um valor semântico, por exemplo, “Vale”, “Elevado”, “Área de Baixo pH” ou “Região com Baixa Sazonalidade da Precipitação”.

Por meio da delimitação dos gradientes que representam a autocorrelação espacial dos dados, busca-se indicar a extensão de regiões pelo limite de sua heterogeneidade interna. Deste modo, cada sub-região ou segmento apresenta o que pode ser considerado como uma área que possui características similares, considerada portanto, homogênea dentro de seus limites. No método proposto nesta tese, a região é delineada em relação aos dados processados pelo algoritmo segmentador *Multiresolution*, que separa as regiões avaliando sua forma e atributos (valores), de acordo com os parâmetros e limites pré-estabelecidos de heterogeneidade interna para os segmentos.

Muitos dos avanços no entendimento de problemas relacionados a análises de dados espacializados ou georreferenciados tem sido conseguidos ao longo dos anos por meio do uso de análise de gradientes. Problemas em mineração, ecológicos, urbanos e agrícolas usam técnicas baseadas em estimativas de autocorrelação por meio de correlogramas ou variogramas para identificar estruturas e padrões espaciais. Estes padrões podem evidenciar a ocorrência de fenômenos que, de outro modo, não poderiam ser detectados ou simulados.

O presente trabalho apresenta um método para parametrizar o algoritmo segmentador, de modo a se obter os maiores segmentos possíveis dentro dos limites de

heterogeneidade estabelecidos. Nele, é apresentado como o variograma - uma ferramenta geoestatística para quantificar a autocorrelação espacial - pode ser combinado ao algoritmo de segmentação multidimensional (*Multiresolution*). Esta combinação pode ser utilizada para delimitação de regiões, desde que os dados estejam dispostos em uma grade amostral regularmente espaçada, como em dados georreferenciados em formato *raster*, por exemplo.

O método apresentado utiliza o efeito pepita (*nugget*) estimado para as camadas/atributos dos dados empregados, quantificado em dois momentos ou médias, para atribuição do parâmetro espacial ótimo, que é então aplicado ao segmentador para a região em análise.

1.1 Motivação

Entre 25 e 27 de setembro de 2015, as Nações Unidas estabeleceram um conjunto de Objetivos de Desenvolvimento Sustentável (ODS) global. Estes objetivos e metas visam alcançar o desenvolvimento sustentável, considerando as dimensões econômica, social e ambiental de forma equilibrada e integrada, para erradicar a pobreza e a fome em todos os lugares até o ano de 2030. Para tal, foram definidos 17 objetivos com 169 metas ([PNUD - Programa das Nações Unidas para o Desenvolvimento, 2015](#)). Deste conjunto, este trabalho procura contribuir com os seguintes objetivos e metas:

Objetivo 2: Acabar com a fome, alcançar a segurança alimentar e melhoria da nutrição, e promover a agricultura sustentável

Meta 2.4: até 2030, garantir sistemas sustentáveis de produção de alimentos e implementar práticas agrícolas resilientes, que aumentem a produtividade e a produção e ajudem a manter os ecossistemas. Estas práticas devem fortalecer a capacidade de adaptação às mudanças do clima, às condições meteorológicas extremas, secas, inundações e outros desastres, e melhorar progressivamente a qualidade da terra e do solo.

Objetivo 15: proteger, recuperar e promover o uso sustentável dos ecossistemas terrestres, gerir de forma sustentável as florestas, combater a desertificação, deter e reverter a degradação da terra, e deter a perda de biodiversidade.

Meta 15.1: assegurar a conservação, recuperação e uso sustentável de ecossistemas terrestres e de água doce interiores e seus serviços, em especial, florestas, zonas úmidas, montanhas e terras áridas, em conformidade com as obrigações decorrentes dos acordos internacionais (até 2020).

Meta 15.2: promover a implementação da gestão sustentável de todos os tipos de florestas, deter o desmatamento, restaurar florestas degradadas e aumentar substancialmente o florestamento e o reflorestamento globalmente (até 2020).

Meta 15.3: combater a desertificação, e restaurar a terra e o solo degradado, incluindo terrenos afetados pela desertificação, secas e inundações, e lutar para alcançar um mundo neutro em termos de degradação do solo (até 2030).

Meta 15.5: tomar medidas urgentes e significativas para reduzir a degradação de *habitat* e estacar a perda de biodiversidade. Até 2020, proteger e evitar a extinção de espécies ameaçadas.

Todos estes ODSs e suas respectivas metas precisarão em algum momento serem mensurados para guiar as políticas públicas que visam atingí-los. Parte desta mensuração será realizada em áreas específicas, levando a necessidade de alguma métrica objetiva para delimitação das regiões, seja por meio da análise dos dados georreferenciados coletados em cada local, ou pelo processamento de imagens de satélites de sensoriamento remoto. Neste contexto, alguns dos problemas que este trabalho pode auxiliar são: (i) a delimitação de zonas de gerenciamento diferenciado em agricultura de precisão; (ii) a determinação de impactos causados pelas mudanças climáticas e degradação do meio ambiente; e (iii) a delimitação de regiões para conservação ou recuperação.

1.1.1 Delimitação de zonas de gerenciamento diferenciado em agricultura de precisão

Conhecer a área plantada, suas deficiências e vantagens, é uma forma de se diminuir o impacto ambiental causado pelo uso de fertilizantes agrícolas, ao mesmo tempo em que se maximiza a produtividade. As pesquisas em Agricultura de Precisão têm proposto formas diferentes de se delimitar sub-regiões em um talhão agrícola para que sejam tratadas de modo homogêneo em relação a reposição de nutrientes do solo, executada após cada ciclo de colheita (MOLIN, 2002; MACHADO; BERNARDI; SILVA, 2004; COELHO, 2005; ZHANG et al., 2009). O método desenvolvido neste trabalho pode ser empregado para encontrar a melhor segmentação de uma determinada região ou área em estudo, por meio do uso dos dados edáficos (estrutura física e química do solo) ou edafoclimáticos disponíveis, desde que distribuídos ao longo do terreno. Neste caso, cada segmento representa uma Zona de Gerenciamento Diferenciado, onde a reposição de insumos e demais tratamentos podem ser executados igualmente por toda a sua extensão. Este problema está relacionado com o ODS-2, Meta 2.4, pois auxilia o agricultor a diminuir a utilização de insumos agrícolas, o que pode minimizar os impactos ao meio ambiente e também melhorar a produtividade.

1.1.2 Impactos causados pelas mudanças climáticas e degradação do meio ambiente

Para se medir os impactos causados por atividades humanas, é preciso selecionar as regiões alvo de acordo com algum critério. Quanto melhores e mais representativas dos ambientes forem as amostras, melhores serão os resultados das análises e aplicações em métodos de classificação ([WULDER; NELSON; SEEMANN, 2007](#)). Em se tratando de ambientes em que uma ou mais espécies foram observadas, a identificação da extensão da região por fatores bióticos e abióticos pode fornecer uma melhora para seleção de dados georreferenciados empregados na geração de modelos de distribuição de espécie (ou nicho ecológico) ([SKOV, 2000](#)). A identificação da extensão de regiões por limite de heterogeneidade dos dados observados também pode ser usada como critério para seleção do entorno de observações ou amostragem baseada em gradientes espaciais. O objetivo neste caso é o estabelecimento de limites de similaridade, com o propósito de mitigar o efeito da autocorrelação espacial entre as amostras por meio da identificação de quais amostras pertencem ao mesmo grupo, pois a autocorrelação pode ocasionar incremento em erros de estimação de modelos subsequentes (erros Tipo I - falso positivo) ([RITCHIE, 2009](#); [BOYLE; BOYLE, 2013](#); [KATAYAMA et al., 2014](#); [MARROT; GARANT; CHARMANTIER, 2015](#)). A abordagem proposta neste trabalho avalia o alcance da autocorrelação espacial por meio das múltiplas camadas/atributos de informação disponíveis para consulta. Este problema está associado ao ODS-15 pois pode ser utilizado para delimitar, analisar e monitorar as regiões de interesse (ex.: para acompanhar mudanças antropomórficas).

1.1.3 Delimitação de regiões para conservação ou recuperação

A expansão agrícola urbana e extrativista apresenta-se como um dos principais fatores responsáveis pela degradação do solo ([FAO; ITPS, 2015](#)) e consequente alteração dos habitats de espécies nativas ([BARNOSKY et al., 2011](#); [CEBALLOS et al., 2015](#)). Estudos apontam que estes fatores atuam como estressores biológicos, levando a um declínio acelerado da vida selvagem no planeta ([DUNFORD; FREEMARK, 2005](#); [BARNOSKY et al., 2011](#)), além de influenciar para o agravamento dos efeitos decorrentes das mudanças climáticas globais ([EHRlich; EHRlich, 2013](#)). Estas mudanças estão alterando importantes Serviços Ecossistêmicos como, por exemplo, o abastecimento de água, impactado negativamente pela falta de chuvas ([VITULE et al., 2015](#)), ou a geração de energia, impactada pelo baixo nível de usinas hidrelétricas. Tais alterações são consideradas significativas e, segundo [Ceballos et al. \(2015\)](#), iniciaram o processo da sexta extinção em massa no planeta, levando a taxas de extinção sem precedentes nos últimos 65 milhões de anos.

Para a produção alimentícia, a definição de regiões protegidas pode ser um diferencial positivo para a produtividade, pois cerca de 75% das culturas necessitam parcialmente

ou completamente de polinização animal. Agricultores podem se beneficiar com um incremento nos níveis de produção advindo de medidas conservacionistas, que preservem os habitats, a fauna e flora nativas. Isso pode ser conseguido delimitando áreas para conservação e, por consequência, preservando a biodiversidade regional ([LOEHRLEIN, 2013](#), p. 22). Este problema está inserido no contexto do ODS-15.

Independentemente dos objetivos traçados pelas Nações Unidas, a crescente abordagem do tema na literatura aponta que a conservação dos habitats e a mensuração dos impactos causados pelas atividades humanas na natureza é importante. Estas iniciativas carecem de soluções para proposição de regiões de recuperação, identificação e preservação de áreas degradadas. Neste contexto, o trabalho procura contribuir com a definição de um método que possa indicar a extensão das regiões baseadas nos dados disponíveis, servindo como ferramenta para suporte a decisão na definição de políticas para delimitação de regiões de proteção ambiental e entornos, auxiliando assim nas atividades envolvendo o planejamento espacial ([STEAD; UNECE - United Nations Economic Commission for Europe, 2008](#)). Este problema também está associado ao ODS-15.

1.2 Objetivo

O objetivo principal desta tese é apresentar um método de segmentação ótimo baseado no processamento de múltiplas camadas/atributos em dados georreferenciados, interpolados ou espectrais, que possam ser aplicados a problemas de delimitação de regiões geográficas.

1.2.1 Objetivo específico

O método proposto deve identificar a extensão de regiões com base nos gradientes espaciais, presentes no conjunto de dados georreferenciados para a região em estudo.

1.2.2 Passos necessários para a obtenção do objetivo

Para a obtenção do objetivo apresentado, consideramos que os seguintes passos devem ser executados:

1. A definição de um conjunto artificial de dados que definem fenômenos espacializados.
2. Estabelecer o conceito de segmentação ótima em mapas com informações espectrais ou interpoladas.
3. Avaliar a aplicabilidade do método para a definição de zonas de manejo agrícola.

4. Avaliar a aplicabilidade do método em dados de ocorrência de espécies para seleção de regiões, que serão então empregadas como suporte ao pós-processamento dos dados e geração de modelos.
5. Avaliar a aplicabilidade em problemas de planejamento espacial, utilizados como suporte a tomada de decisões em áreas de proteção e conservação da biodiversidade.

1.3 Contribuições da tese

Algumas das contribuições apresentadas neste trabalho são:

- A tese apresenta como um conjunto de dados artificialmente gerado pode ser utilizado no teste de um método de segmentação multivariada, baseadas em estimativas de autocorrelação e variância.
- Apresenta como a segmentação utilizando múltiplas camadas/atributos pode ser avaliada e parametrizada pelo uso combinado do algoritmo segmentador *Multiresolution* e análise variográfica.
- Apresenta como a segmentação pode ser feita de forma hierárquica por abordagem *bottom-up*, para detecção e delimitação de fenômenos espaciais presentes na região em análise. Também define uma forma de avaliação de regiões para identificação de estruturas espaciais e a necessidade de expansão da janela de avaliação, caso os critérios estabelecidos não sejam alcançados.
- Apresenta como a estimativa de impacto de degradação ambiental pode ser baseada em segmentação e identificação de regiões, por meio da extensão de autocorrelação espacial. Exemplo apresentado na seção 4.2.
- Apresenta uma abordagem para identificação de bioclimas de Unidades de Conservação (UC) da biodiversidade. Exemplo apresentado na seção 4.2.

1.4 Estrutura da tese

A tese está dividida em 5 capítulos. No capítulo 2 são apresentados os conceitos necessários para a definição do método proposto.

No capítulo 3, é apresentado o método proposto nesta tese, as funções utilizadas para avaliação da segmentação e os conceitos empregados para caracterização do parâmetro ótimo a ser empregado no algoritmo segmentador.

No capítulo 4 são apresentados alguns exemplos da aplicabilidade do método por meio de casos em que o algoritmo pode ser empregado e os resultados em cada

problema proposto. Devido à possibilidade de aplicação do método em diferentes contextos, optou-se por discutir os resultados em cada domínio de aplicação na medida em que são apresentados.

Finalmente, no capítulo 5, são apresentadas as conclusões obtidas ao longo de toda a pesquisa, além das recomendações e os pontos fortes e fracos do método proposto. Também são discutidas algumas possíveis aplicações para os trabalhos futuros.

2 Fundamentação teórica e revisão bibliográfica

Este capítulo apresenta os conceitos que suportam o método de segmentação proposto nesta pesquisa. Inicialmente é fornecida uma breve explanação sobre os conceitos e métodos necessários ao entendimento, definição e identificação de variáveis regionalizadas. Na sequência, são apresentados alguns dos métodos normalmente utilizados para identificação de estruturas espaciais. Também é abordado o problema de segmentação de imagens e análises de textura, tal como sua relação com a análise espacial. Ao final do capítulo, são apresentados alguns dos trabalhos correlatos relevantes para esta proposição.

2.1 Variáveis regionalizadas e a geoestatística

A Teoria das Variáveis Regionalizadas (MATHERON, 1971) surgiu a partir de uma serie de publicações do professor Georges Matheron, inspiradas principalmente nos trabalho de H. J. Wijs, professor de Universidade Técnica de Delft, Holanda, e Daniel G. Krige, engenheiro de minas na África do Sul (YAMAMOTO; LANDIM, 2013).

Adaptando-se a notação de Matheron, uma variável regionalizada é definida como uma função aleatória Z que descreve um fenômeno distribuído no espaço e apresenta uma estrutura espacial qualquer. Assim, temos que a função $Z(\mathbf{x})$ caracteriza o fenômeno regionalizado para qualquer ponto em $\mathbf{x} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ pertencente ao espaço analisado. Segundo Matheron, esta função é composta por dois aspectos: um aspecto estruturado, aqui denotado por $\bar{x}$, que representa a extensão das características de um fenômeno regionalizado (média); e um aspecto aleatório, aqui representado por $e(\mathbf{x})$, que representa a irregularidade e a imprevisibilidade das variações de um ponto a outro (resíduo), representados na Eq. (2.1) (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2013).

$$Z(\mathbf{x}) = \bar{x} + e(\mathbf{x}) \quad (2.1)$$

Define-se variável regionalizada como uma variável aleatória em que a cada ponto no espaço é associada uma distribuição de probabilidades de ocorrência de valores. Nela, cada ponto em $\mathbf{x}$ é caracterizado pela variável aleatória $Z(\mathbf{x})$, sua média $\bar{x}$, variância s^2 e sua função de distribuição acumulada. Deste modo, define-se o conjunto $Z(\mathbf{x}) = \{z(\vec{x}_1), z(\vec{x}_2), \dots, z(\vec{x}_n)\}$ como a função aleatória, processo aleatório ou processo estocástico, no qual cada variável $z(\vec{x}_i)$ representa uma amostra da região em análise.

No estudo do comportamento das variáveis regionalizadas, entendidas como resultados de um processo estocástico, podem ser utilizados métodos de análise de continuidade

espacial. Estes métodos partem do pressuposto de que valores próximos aos locais de amostragem tendem a ser autocorrelacionados. Deste modo, para que este estudo seja feito, é necessário assumir alguns graus de estacionalidade para a função aleatória, definidos como estacionalidade de 1ª e 2ª ordem. Uma variável regionalizada possui estacionalidade de 1ª ordem se mantiver os mesmos parâmetros ($\bar{x}$ e $e(\mathbf{x})$) para qualquer que seja a sua posição. Deste modo, sua média ou esperança devem ser invariáveis, portanto não dependem da localização de $\vec{x}_i$ (Eq. (2.2)).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n Z(x_i) = E(Z(x_1)) = E(Z(x_2)) = \dots E(Z(x_n)) \quad (2.2)$$

Igualmente importante, a estacionalidade de 2ª ordem indica que, além de média ($\bar{x}$) constante, a variância (s^2) existe e depende apenas de uma distância h entre dois pontos distintos.

Assumindo estes dois tipos de estacionalidade, a Teoria das Variáveis Regionalizadas tem por objetivo o estudo e a representação estrutural deste tipo de variável. A técnica vem sendo empregada principalmente para resolução de problemas de estimativa de qualquer variável que dependa do espaço e apresente, além do caráter aleatório, um caráter estrutural (WEBSTER; OLIVER, 2007). Assume-se aqui que o caráter estrutural define a distribuição de uma variável no espaço analisado. Ao estudo das variáveis regionalizadas por meio de técnicas estocásticas dá-se o nome de Geoestatística, uma das subáreas da Estatística Moderna.

Os métodos denominados geoestatísticos em geral buscam identificar e descrever a estruturação espacial dos dados, estabelecendo uma função de correlação espacial (CHILES; DELFINER, 2012). Historicamente, a partir de 1980, a metodologia geoestatística passou a ser aplicada em diversos domínios além da mineração, área na qual se originou. Estudos em Agricultura de Precisão, Análise Espacial de Crimes, Cartografia, Climatologia, Ecologia de Paisagem, Engenharia Florestal, Epidemiologia, Geologia e Geotecnia, tiveram significativos avanços com a aplicação destas técnicas (YAMAMOTO; LANDIM, 2013).

Para qualquer estudo geoestatístico, é necessária a construção de variáveis regionalizadas pela amostragem ou coleta de um conjunto de informações observadas ao longo do espaço em que o fenômeno espacial ocorre (DRUCK, 2004, cap. 3). O objetivo é inferir as propriedades do fenômeno para identificar sua estrutura, limites ou organização no espaço. Neste contexto, são importantes os conceitos de fenômeno espacial, heterogeneidade e amostras, apresentados nas próximas subseções deste capítulo.

2.1.1 Fenômeno espacial

O fenômeno espacial pode ser caracterizado como o conjunto de valores para a variável de interesse, que define sua distribuição e a variabilidade espacial em um domínio

de duas (2D) ou três dimensões (3D). O fenômeno, portanto, representa estatisticamente todos os valores possíveis de uma população da qual uma amostra pode ser extraída (YAMAMOTO; LANDIM, 2013, p. 19).

2.1.2 Heterogeneidade, homogeneidade e escala espacial

Define-se a variação de um fenômeno espacial ao longo do espaço analisado como Heterogeneidade Espacial. Os métodos de análise espacial procuram mensurar a magnitude e a extensão desta heterogeneidade por meio da identificação de sua correlação espacial na busca pela identificação de estruturas espaciais de qualquer tipo (BORCARD; GILLET; LEGENDRE, 2011, p. 321). Como ferramentas para a identificação destas estruturas, normalmente são empregadas técnicas baseadas em correlogramas e variogramas (LEGENDRE, 1990). Todavia, a heterogeneidade de um fenômeno está intimamente ligada a sua escala. Assim, uma região aparentemente homogênea em uma escala, pode se mostrar significativamente heterogênea quando observada em uma resolução superior, devido à maior quantidade de pixels por unidade de área (FITCH et al., 2015).

A homogeneidade e a heterogeneidade podem ser consideradas como extremos de uma escala contínua positiva, definida em termos de heterogeneidade mínima e máxima, em que heterogeneidade zero equivale a homogeneidade. Assim, quanto maior ou mais detalhada for a região em estudo, maior será sua heterogeneidade. Entretanto, quanto maior for a heterogeneidade de um fenômeno, maior é a complexidade do modelo utilizado para sua representação, conseqüentemente, sua estimação será mais difícil. Neste sentido, é importante encontrar um equilíbrio entre a quantidade de amostras ou número de pixels por unidade de área e a necessidade de detalhamento do fenômeno em análise.

Fenômenos naturais espaciais dificilmente apresentam homogeneidade, assim buscamos neste trabalho encontrar limites para a heterogeneidade baseados em autocorrelação espacial e sobreposição de gradientes. O objetivo é que as regiões assim delimitadas possam ser tratadas como similares ou “homogêneas” em toda a sua extensão, respeitando os limites pré-estabelecidos.

2.1.3 Amostra e métodos de amostragem

Consideramos uma amostra como sendo um subconjunto do fenômeno espacial com maior densidade possível dentro da região em análise. O principal objetivo da amostragem é reproduzir a distribuição e variabilidade espaciais do fenômeno estudado. Existe uma incerteza associada em qualquer estimativa baseada em amostragem de pontos, e os métodos geoestatísticos caracterizam-se por oferecerem as incertezas associadas a estas estimativas.

Pode-se classificar a amostragem em três tipos: aleatória simples, aleatória estrati-

ficada, e sistemática (YAMAMOTO; LANDIM, 2013, p. 20). Para a amostragem aleatória simples, um conjunto de N pontos são posicionados aleatoriamente e sem sobreposição na região de interesse. Na amostragem aleatória estratificada, a região de interesse é subdividida em células de dimensões fixas. Para cada célula contendo um conjunto de pontos, um ponto é selecionado aleatoriamente. Ao término do processo, a quantidade de pontos selecionados corresponde ao número de células pelo qual a região foi subdividida. Similarmente, a amostragem sistemática é caracterizada quando os pontos são amostrados em uma malha regular definida sobre a região em estudo (YAMAMOTO; LANDIM, 2013, p. 20).

Neste trabalho, as camadas de informação representam regiões amostradas em grades regulares sobrepostas. Cada grade regular forma um mapa de bits compondo uma imagem *raster*, sendo que cada *pixel* representa uma amostra de área na região de interesse.

2.2 Modelagem de estruturas espaciais

No contexto de dados relacionados a estudos de conservação da biodiversidade, o ato de identificar estruturas espaciais em dados ecológicos está relacionado com a busca pela evidência da influência de algum processo, seja ele causado por força do ambiente ou por fatores bióticos. Estas estruturas espaciais podem se apresentar em diferentes escalas. A identificação destas escalas e a modelagem das estruturas espaciais correspondentes é um objetivo há muito buscado por ecologistas (BORCARD; GILLET; LEGENDRE, 2011). Esta busca se dá pois as escalas podem ser relacionadas a variáveis explanatórias representando hipóteses estabelecidas *a priori*, ou ajudar no estabelecimento de novas hipóteses que possam oferecer explicações para o processo observado (BORCARD; GILLET; LEGENDRE, 2011, p. 238).

As abordagens pioneiras para modelagem de estruturas espaciais utilizavam apenas as coordenadas como variáveis preditoras em métodos baseados em regressão linear Legendre (1990). Estas técnicas evoluíram para o uso de Análise Multivariada de Superfície de Tendência, técnica na qual as coordenadas são combinadas em polinômios formados pelos dados explicativos, modelando estruturas curvilíneas. Na prática, esta técnica é restrita a polinômios de segundo e terceiro graus, devido ao incremento substancial da complexidade computacional e tempo dispendido para processamento dos modelos (BORCARD; GILLET; LEGENDRE, 2011). Devido a esta limitação computacional, a utilização destas técnicas limita-se a captura ou modelagem de fenômenos espaciais simples, sendo consideradas pouco eficientes para lidar com dados combinados em diferentes escalas (BORCARD; GILLET; LEGENDRE, 2011, p. 238).

Para identificar as diferentes escalas que um fenômeno espacial pode ocorrer Borcard e Legendre (2002), Borcard et al. (2004) propuseram um método denominado Coordenadas

Principais de Matrizes Vizinhas (*Principal Coordinates of Neighbour Matrices* - PCNM). Este método permite a modelagem de qualquer estrutura espacial por meio do uso de variáveis espaciais ortogonais (linearmente independentes) combinadas por ordenação canônica, para amostragem das regiões analisadas (BORCARD; GILLET; LEGENDRE, 2011). Neste trabalho, estas variáveis foram utilizadas para avaliar os efeitos da segmentação em estruturas espaciais genéricas. Cada uma destas variáveis PCNM, gerada para uma grade pré-estabelecida, possui o seu respectivo índice de Moran previamente calculado. Estas variáveis formam mapas que são usados para avaliar a técnica proposta para escolha do valor ótimo para o parâmetro de segmentação espacial (sp). Este parâmetro é o responsável por definir o tamanho dos segmentos no algoritmo empregado. Assume-se, neste caso, a definição de que o melhor parâmetro de segmentação será aquele que apresentar os maiores segmentos para a região estudada, e que estes segmentos obtidos apresentam máxima heterogeneidade entre eles e considerável homogeneidade interna (baixa variabilidade intrasegmento), se consideradas todas as variáveis disponíveis para a área em estudo.

O Índice de Moran é aqui empregado como medida de autocorrelação, dispersão e estrutura dos gradientes presentes nos mapas artificiais gerados, e é apresentado na seção 2.2.1. O método utilizado para gerar estas variáveis artificiais, que simulam diferentes tipos de gradientes espaciais, em diferentes escalas, por meio do PCMN, é apresentado na seção 2.3.

2.2.1 O índice de Moran para teste de autocorrelação espacial

Ao estudar a distribuição de fertilizantes no solo como um sistema estocástico espacialmente distribuído, Patrick Alfred Pierce Moran desenvolveu um índice para caracterizar a autocorrelação espacial de um sinal por meio de sua relação de proximidade espacial e similaridade de valores (MORAN, 1950). Este índice, denominado I de Moran, avalia se os dados espacialmente distribuídos apresentam padrões tendendo a agrupamentos ou a dispersão. Adotando-se a notação utilizada por Borcard, Gillet e Legendre (2011) o índice é caracterizado pela equação Eq. (2.3) como:

$$I(d) = \frac{\frac{1}{W} \sum_{h=1}^n \sum_{i=1}^n w_{hi}(y_h - \bar{y})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \text{ para } h \neq i \quad (2.3)$$

onde

y_h e y_i representam os valores de uma variável y para os pares h e i de uma matriz das distâncias calculadas entre locais.

$\bar{y}$ representa a média de y .

d classe de distâncias os quais são calculados os pares.

w_{hi} peso para o qual é atribuído o valor 1 caso pertença a alguma classe d de distâncias e 0 caso contrário.

W é o número de pares pontos utilizados em cada classe.

Para a interpretação do índice, valores próximos a 1.0 para I indicam a presença de agrupamentos, ou seja, autocorrelação espacial positiva. Já valores tendendo a -1.0 (autocorrelação negativa), indicam dispersão estruturada ou organizada. Valores próximos a 0 indicam aleatoriedade, portanto ausência de correlação espacial (GOODCHILD, 1986). Entretanto, para fins de avaliação de significância estatística, o escore padronizado (z -score) e o valor p devem ser calculados. A hipótese nula para o teste indica que não existe padrão de aglomeração para a variável na região avaliada. A hipótese nula pode ser rejeitada se o valor p for significativamente pequeno (usualmente inferior a 0,05) e valor de z suficientemente grande (pelo menos 1,5 vezes o desvio padrão) (LEVINE, 2010).

O valor esperado para o índice de Moran (E_I) é dado pela Eq. (2.4),

$$E_I = -\frac{1}{(n-1)} \quad (2.4)$$

a partir da comparação entre os valores de I e E_I é possível identificar como os dados estão auto-correlacionados. A constatação de que o valor de I é superior ao esperado E_I é uma indicação de que o fenômeno apresenta autocorrelação positiva, provavelmente apresentando agrupamentos (LEVINE, 2010). E_I será próximo de zero quando o valor de n (número total de pares observados) for grande. Neste trabalho, esta medida é utilizada como parte do método empregado para geração de variáveis regionalizadas, apresentados na seção 2.3.1.

Existem outros testes de autocorrelação, como por exemplo, os índices C de Geary (GEARY, 1954) e G de Getis (GETIS; ORD, 2010). Estes métodos são amplamente utilizados em diversos estudos que buscam identificar relações espaciais em dados como, por exemplo, em análises biogeográficas (KNEGT et al., 2010) e em estudos criminalísticos (LEVINE, 2010). Estas técnicas geralmente são avaliadas com o uso de correlogramas. O uso e interpretação destes métodos não são abordados neste trabalho por não serem essenciais ao entendimento da abordagem proposta. Recomenda-se o estudo de Bivand, Pebesma e Gómez-Rubio (2013) para um maior aprofundamento.

Outro método, normalmente utilizado em Geoestatística, é a avaliação variográfica, apresentado na seção 2.3.2 e utilizado como técnica de avaliação estrutural espacial para a Segmentação Multiresolução Variográfica Ótima, apresentado na seção 3.1.

2.3 Dados Artificiais - Variáveis Espaciais Baseadas em Autovetores

A utilização de dados artificiais neste trabalho tem como objetivo simular fenômenos espaciais completamente representados na janela de avaliação. Para tal, são consideradas como sendo variáveis regionalizadas os gradientes gerados pela combinação de senoides que apresentam frequências incrementadas progressivamente, em combinações de diferentes fases relativas. Aplicadas aos eixos x e y de uma imagem 2D, estas senoides geram figuras complexas, formando gradientes que simulam diferentes tipos de autocorrelação espacial. Estas senoides são geradas pelo método denominado Coordenadas Principais das Matrizes Vizinhas (*Principal Coordinates of Neighbor Matrices* - PCNM), um caso especial de uma classe de métodos denominada Mapas de Autovetores de Moran (*Moran Eigenvector Maps* - MEM) (BORCARD; GILLET; LEGENDRE, 2011). O método PCNM utiliza a diagonalização da matriz de pesos para relações espaciais (neste caso distância relativa entre os pontos de amostra). Desta matriz são extraídos os autovetores que maximizam o índice de correlação de Moran (DRAY; LEGENDRE; PERES-NETO, 2006), apresentado na seção 2.2.1.

2.3.1 Imagens geradas por funções PCNM

As funções PCNM são obtidas por meio de análise de coordenadas principais (GOWER, 1966) de uma matriz contendo as distâncias pareadas entre os pontos de amostragem. Estes pontos, dispostos em uma grade regular, representam os *pixels* das imagens *raster* utilizadas. Cada imagem gerada representa uma camada a ser processada posteriormente. Os autovalores positivos associados aos autovetores correspondem a representação euclidiana da matriz de distâncias truncadas. O algoritmo para geração dos mapas pode ser resumido pelos seguintes passos (DRAY; LEGENDRE; PERES-NETO, 2006; ZUUR; IENO; SMITH, 2007):

1. Inicialmente deve-se computar a matriz representando as distâncias euclidianas pareadas entre os n pontos de amostragem ($D = [d_{ij}]$).
2. Na sequência, deve-se escolher um valor limite t para construir uma matriz de distâncias truncadas que será aplicado a matriz $D = [d_{ij}]$ usando a Eq. (2.5). Neste caso o valor escolhido representa a distância entre pontos da grade amostral, indicado como uma unidade do mapa ($t=1$).

$$\mathbf{D}^* = \begin{cases} d_{ij} & \text{se } d_{ij} \leq t \\ 4t & \text{se } d_{ij} > t \end{cases} \quad (2.5)$$

3. Em seguida, deve-se executar a análise de coordenadas principais para a matriz das distâncias truncadas $\mathbf{D}^*$. A análise consiste da diagonalização da matriz Δ como segue:

$$\Delta = \left[-\frac{1}{2} \left(d_{ij}^{*2} - d_{i.}^{*2} - d_{.j}^{*2} + d_{..}^{*2} \right) \right] \quad (2.6)$$

em que:

$$d_{i.}^{*2} = \frac{1}{n} \sum_{j=1}^n d_{ij}^{*2}, \quad (2.7)$$

$$d_{.j}^{*2} = \frac{1}{n} \sum_{i=1}^n d_{ij}^{*2}, \quad (2.8)$$

$$d_{..}^{*2} = \frac{1}{n} \sum_{i=1}^n d_{i.}^{*2}, \quad (2.9)$$

onde:

$d_{i.}^{*2}$ é a média calculada para a linha i da matriz $\mathbf{D}^*$,

$d_{.j}^{*2}$ é a média calculada para a coluna j da matriz $\mathbf{D}^*$,

$d_{..}^{*2}$ é a média calculada para a matriz $\mathbf{D}^*$ - soma da média de todas as linhas calculadas pela Eq. (2.7).

Os autovetores de Δ são escalonados pela raiz de seus respectivos autovalores ($\sqrt{\lambda_k}$) para se obter as coordenadas principais. Como a matriz de distância foi truncada, seus autovalores apresentam valores negativos, impossibilitando sua completa representação no espaço euclidiano. Neste caso a abordagem PCNM apenas utiliza os autovalores positivos para a obtenção das coordenadas, representando a matriz $\mathbf{D}^*$ no espaço euclidiano. Segundo [Dray, Legendre e Peres-Neto \(2006\)](#), nestes mapas os autovetores associados a valores altos para autovalores possuem significativa autocorrelação positiva. Esta autocorrelação pode ser utilizada para descrever estruturas espaciais em que cada autovetor representa um ou mais fenômenos espaciais periódicos. Quanto maior o autovalor, maior o fenômeno em relação ao mapa, ou seja, menor a sua frequência em relação a sua distribuição espacial na região observada.

No caso de uma grade com pontos igualmente espaçados, o resultado das variáveis PCNM representa uma série de senóides com períodos decrescentes. Considerando um conjunto contendo n pontos regularmente espaçados por um intervalo s , [Guénard et al. \(2010\)](#) mostra que, no caso unidimensional, o período λ_i da senoide gerada pelo autovetor correspondente ao autovalor ordenado decrescentemente na posição i é dado pela Eq. (2.10).

$$\lambda_i = \frac{2(n + s)}{i + 1} \quad (2.10)$$

Como exemplo, apresentamos os gráficos para o caso unidimensional na Fig. 1 e o respectivo cálculo para o Índice de Moran, Fig. 2.

Figura 1 – Seleção de algumas das 67 variáveis PCNM que possuem autovalores positivos construídas por meio o uso da i -ésima coluna da matriz Δ para os 100 pontos equidistantes. As primeiras 49 variáveis apresentam índice I de Moran maior que E_I , indicando que estas variáveis modelam correlações espaciais positivas

Fonte: Adaptado de [Borcard, Gillet e Legendre \(2011\)](#).

Para simulação de variáveis regionalizadas, foram utilizadas as variáveis PCNM calculadas para uma grade 20 x 20 que representam a aplicação bidimensional da técnica, denominados neste caso como Mapas de Autovetores de Moran. Alguns dos mapas gerados são apresentados na Fig. 3. Estes mapas foram utilizados como dados experimentais para avaliação do método de segmentação multiresolução variográfica ótima apresentado na seção 3.1. O incremento do índice da variável PCMN gera um mapa contendo fenômenos de menor escala espacial em relação ao tamanho da grade de avaliação. O cálculo do Índice de Moran para as 279 variáveis PCMN é apresentado na Fig. 4.

Figura 2 – Índice de Moran calculado para as 67 variáveis PCNM. Cada autovetor i representa uma coluna da matriz Δ , obtida para o vetor de 100 pontos equidistantes. As primeiras 49 variáveis apresentam índice I de Moran maior que E_I , indicando que estas variáveis modelam correlações espaciais positivas, conforme apresentado na seção 2.2.1)

Fonte: Adaptado da função PCNM de Borcard, Gillet e Legendre (2011).

2.3.2 Análise variográfica

Diversas aplicações em Ciências da Terra, Sensoriamento Remoto e Processamento de Imagens apresentam uma vasta quantidade de dados regularmente amostrados em grade como característica principal dos seus conjuntos de dados. Esta característica viabiliza a aplicação de diversas técnicas de análise de continuidade espacial e autocorrelação destes dados, permitindo a caracterização dos fenômenos espaciais observados (SRIVASTAVA; PARKER, 1989).

Uma destas técnicas, utilizada em Geoestatística, é a Análise Variográfica. Esta análise é feita por meio da construção de um gráfico denominado variograma em que a variância é representada no eixo da ordenada e as classes de distância na abscissa. Neste gráfico são avaliadas as variâncias dos dados, para uma certa separação espacial entre os pontos amostrados (KOURGLI; BELHADJ-AISSA, 2004; BACHMAIER; BACKES, 2011). O variograma caracteriza a variabilidade espacial de uma variável aleatória pela avaliação das diferenças entre as amostras na medida que a distância entre seus pares

Figura 3 – Mapas gerados para algumas das 279 variáveis PCNM a partir de uma grade regular de 20 x 20 pontos. Cada autovetor i indexa uma coluna da matriz Δ , obtida para uma matriz de 20 x 20 pontos que compõem uma grade contendo 400 pontos equidistantes. Na figura temos alguns dos mapas gerados para os autovetores na posição i , nomeados de L_i

Fonte: Convertido para *raster* a partir das funções apresentadas em [Borcard, Gillet e Legendre \(2011\)](#).

Figura 4 – Índice de Moran calculado para as 279 variáveis PCNM geradas em uma matriz 20×20 , representando uma grade de 400 pontos equidistantes. Cada autovetor i indexa uma coluna da matriz Δ , obtida para a grade. As primeiras 209 variáveis apresentam I de Moran maior que E_I . Quanto mais próximo de 1, mais positiva é a correlação espacial medida, portanto menor a dispersão espacial aleatória, conforme apresentado na seção 2.2.1)

Fonte: Adaptado da função PCNM de [Borcard, Gillet e Legendre \(2011\)](#).

aumentam.

Considerando a estacionalidade de segunda ordem e a hipótese intrínseca, que assumem que a função de autocorrelação pode ser aplicada em toda a superfície em estudo, temos que o processo gerador das amostras é uma função aleatória $Z(\vec{x})$ composta pela média e resíduo (Eq. (2.1)), e possui uma média constante (Eq. (2.2)). Assim o variograma pode ser definido pela Eq. (2.11) como:

$$\gamma(\vec{h}) = \frac{1}{2} E(Z(\vec{x}) - Z(\vec{x} + \vec{h}))^2 \quad (2.11)$$

em que $\vec{x}$ e $\vec{x} + \vec{h}$ representam posições no espaço separados pelo vetor $\vec{h}$.

Em termos numéricos, a função variograma é calculada em unidades de distância. Tipicamente, esta função é estimada por meio da aproximação a uma função previamente estabelecida a um variograma estimado pelas amostras, denominado variograma experimental. Sob estas condições, assume-se que a variância de Z é constante e que a correlação espacial de Z não dependa de $\vec{x}$, mas somente da distância definida por $\vec{h}$. Então pode-se estimar a correlação entre pontos amostrados $\{z(\vec{x}_i), z(\vec{x}_j)\}$ que possuam distâncias próximas a $\|\vec{h}\|$ (assumindo-se isotropia para o fenômeno). Definindo-se $N(\vec{h})$ como o número de pares de amostras $\{z(\vec{x}_i), z(\vec{x}_i + \vec{h})\}$ obtidas para uma distância $\|\vec{h}\|$, temos que o variograma experimental é dado por:

$$\hat{\gamma}(\vec{h}) = \frac{1}{2} \frac{1}{N(\vec{h})} \sum_{i=1}^{N(\vec{h})} [z(\vec{x}_i + \vec{h}) - z(\vec{x}_i)]^2 \quad (2.12)$$

Na prática, são utilizadas distâncias aproximadas, principalmente em problemas em que a grade amostral é irregular. Geralmente os variogramas tendem a um nível estacionário de variância C , identificado como **patamar**. A presença desta característica estrutural do variograma indica que os dados apresentam estacionalidade de segunda ordem, Fig. 5. A distância a em que este patamar é alcançado é denominada de **alcance**. Pares com distância de separação maiores que o alcance (a) são considerados não correlacionados espacialmente. A descontinuidade próxima à origem, indicada por C_0 (nunca interceptando

Figura 5 – Elementos que compõem um variograma. No exemplo: efeito pepita ($C_0 = 2$), patamar ($C = C_0 + C_1 = 6$) e alcance ($a = 8$)

Fonte: Adaptado de Legendre e Legendre (2012).

o zero no eixo das ordenadas) é denominada de **efeito pepita**, e representa o efeito de variações espaciais em pequena escala para a superfície analisada. Esta medida pode ser útil para identificar a necessidade de redefinição da grade de amostragem para intervalos menores, a presença de erros de amostragem ou erros nas medições, ou ainda a dispersão dos fenômenos espaciais no mapa.

Apesar do variograma experimental fornecer uma boa descrição sobre a continuidade espacial da variável, ele representa apenas valores discretos amostrados para as classes de distâncias específicas utilizadas. Para a o uso em técnicas de inferência espacial (interpolação) ou para simulação do processo observado, o variograma precisa fornecer uma estimativa da variância para todas as distâncias intermediárias. Esta aproximação pode ser feita por meio do uso de uma função contínua no variograma. Esta função deve apresentar um valor estimado para a variância mínima e máxima que se aproxime dos dados do variograma experimental (Eq. (2.12)). Um método amplamente utilizado é inferir, a partir da aproximação à um modelo paramétrico aos dados obtidos do variograma experimental, a função contínua que melhor representa o variograma (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2013, p. 224). Por motivo de simplificação de notação adotamos $\|\vec{h}\| = h$. Alguns

dos modelos paramétricos comumente utilizados podem ser vistos na Figura 6.

Figura 6 – Modelos de funções comumente empregados em variogramas

Fonte: Adaptado de Legendre e Legendre (2012).

As equações características destes modelos são dadas por:

- Modelo esférico: $\gamma(h) = C_0 + C_1 \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right]$ se $h \leq a$; $\gamma(h) = C$ se $h > a$.
- Modelo exponencial: $\gamma(h) = C_0 + C_1 \left[1 - \exp \left(-3 \frac{h}{a} \right) \right]$.
- Modelo gaussiano: $\gamma(h) = C_0 + C_1 \left[1 - \exp \left(-3 \frac{h^2}{a^2} \right) \right]$.
- Modelo efeito furo: $\gamma(h) = C_0 + C_1 \left[1 - \exp \left(1 - \frac{\sin(ah)}{ah} \right) \right]$.
- Modelo linear: $\gamma(h) = C_0 + bh$ sendo que o parâmetro b representa a inclinação da curva para o modelo. Um modelo linear com patamar C pode ser definido adicionando-se a seguinte condição ao modelo: $\gamma(h) = C$ se $h \geq a$.
- Efeito pepita puro: $\gamma(h) = C_0$ se $h > 0$; $\gamma(h) = 0$ se $h = 0$. A segunda parte é aplicada para a estimativa pontual. Na prática, as observações possuem o tamanho da granularidade da amostragem, ou seja, sempre haverá uma distância mínima entre os pontos amostrados. Assim, o erro para a escala observada será sempre superior a zero, portanto, o valor obtido para C_0 será sempre superior a 0 (LEGENDRE; LEGENDRE, 2012).

Devido ao escopo deste trabalho, não apresentamos outros modelos de uso menos frequentes. Maiores informações sobre o uso destes modelos podem ser encontradas em livros de geostatística, como [Bivand, Pebesma e Gómez-Rubio \(2013\)](#), [Druck \(2004\)](#) ou [Yamamoto e Landim \(2013\)](#).

2.3.3 Anisotropia

A anisotropia é o fenômeno pelo qual a função de autocorrelação espacial não é igual para todas as direções geográficas avaliadas ([MATHERON, 1971](#)). Assim, a variabilidade ou a distribuição espacial das variáveis ocorre mais intensamente em uma direção e menos em outra, uma característica relativamente comum em fenômenos da natureza ([CAMARGO; FELGUEIRAS; MONTEIRO, 2001](#)). Podemos classificar a anisotropia em dois tipos básicos: geométrica e zonal. Na anisotropia geométrica, o alcance do variograma muda de acordo com a direção, mas o patamar continua o mesmo. Para a anisotropia zonal o patamar muda com a direção, enquanto o alcance permanece inalterado. Na natureza, um fenômeno pode apresentar uma destas formas de anisotropia ou uma combinação destes tipos. A definição de heterogeneidade adotada pelo método proposto neste trabalho indiretamente considera a anisotropia, por delimitar o fenômeno pela sua similaridade de valores e forma. Assim, emprega-se indiretamente neste trabalho a aproximação pela média dos patamares a um modelo isotrópico, pois utiliza-se somente o efeito pepita medido pela aproximação ao modelo Gaussiano como medida de variância máxima e dispersão dos dados, como apresentado na seção [3.1.4](#).

2.4 Análise de texturas

A análise de textura é o processo pelo qual uma imagem é caracterizada pela quantificação de como as texturas que formam a imagem são compostas em função da variação espacial da intensidade dos *pixels* ([ABDULMUNIM; PETROLEUM, 2012](#)). Esta análise pode ser utilizada para detecção de bordas e segmentação de imagens, e possui aplicações em diversas áreas, como sensoriamento remoto, classificação de objetos e processamento de imagens médicas. As técnicas de para análise de texturas podem ser classificadas como [Kourgli e Belhadj-Aissa \(2004\)](#):

- **Estruturais:** utilizam métodos geométricos ou de morfologia matemática para obter os padrões que possam descrever as informações na imagem. Ex.: aplicação de filtros na imagem, detecção de borda, tesselação de Voronoi.
- **Estatísticas:** aplica métodos estocásticos para se obter as propriedades que possam descrever a distribuição espacial dos *pixels* na imagem. Ex.: técnicas para estimativa

de autocorrelação espacial, como correlogramas e variogramas; análise de frequências e histogramas.

- **Transformadas:** utiliza transformadas para análise no domínio da frequência, como a transformada de Fourier ou a transformada Wavelets para se obter as características descritoras da textura.

Kourgli e Belhadj-Aissa (2004) mostram como a análise variográfica pode ser utilizada na aplicação de segmentação em imagens monocromáticas. No trabalho, os autores descrevem como a extensão de influência, obtida pela medição da distância (em *pixels*) para o qual se obtém o primeiro valor máximo para a semi-variância, pode ser utilizada para se parametrizar as dimensões da matriz aplicada ao filtro Sobel para detecção de bordas.

Neste trabalho, aplicamos a análise variográfica para mensurar como a segmentação se comporta em cada nível, em um segmentador multivariado. Para isto, utilizamos o efeito pepita como elemento caracterizador das camadas/atributos e do efeito da segmentação.

2.5 Segmentação de Imagens

A segmentação de imagens é a técnica de processamento de imagens que tem como objetivo dividir a imagem em objetos “coerentes” (MUELLER; SEGL; KAUFMANN, 2004; KAGANAMI; BEIJI, 2009).

Pode-se dividir os algoritmos de segmentação de imagens em duas classes (MUELLER; SEGL; KAUFMANN, 2004; SABA; VALADANZOUJ; MOKHTARZADE, 2013):

- Segmentação baseada em detecção de borda (edge-based): neste tipo de segmentação são indicados os locais onde ocorrem mudanças abruptas na imagem.
- Segmentação baseada em divisão por regiões (region-based): as abordagens para este tipo de segmentação são baseadas em fatores limitantes, crescimento de regiões e divisão ou fusão de regiões. Esta é a abordagem adotada neste trabalho, pois o problema atacado é a segmentação para se obter regiões consideradas como “homogêneas”, contendo a maior extensão possível em cada segmento.

2.6 Trabalhos correlatos

O uso de medidas de variância em dados georreferenciados não é um conceito novo e vem sendo empregado em imagens de sensoriamento remoto e segmentação pelo menos desde a década de 1980 (WOODCOCK; STRAHLER, 1987). Por outro lado, a combinação de medidas de variância para a segmentação de imagens multiespectrais (imagem de

satélite composta de n-bandas espectrais, portanto um tipo de dado multivariado) é relativamente recente. Parte dos trabalhos recentes com este tipo de dado, têm como base o conceito da variância local (MARTHA et al., 2011; DRĂGUȚ; EISANK, 2012). A medida da variância local, proposta inicialmente por Woodcock e Strahler (1987), é efetuada calculando-se a média em uma máscara móvel de 3 x 3 *pixels* na imagem. A média do desvio padrão para todas as máscaras é utilizada como medida de variância local para a imagem. Recentemente, a técnica foi utilizada por Drăguț e Eisank (2012) para automatizar a escolha do parâmetro espacial (*sp*) empregado no algoritmo segmentador *Multiresolution* (BAATZ; SCHÄPE; STROBL, 2000). No entanto, pode-se destacar que o método por variância local não aplica nenhuma função que quantifique geoestatisticamente a estrutura espacial resultante do incremento do parâmetro *sp* empregado na segmentação. A técnica se limita ao uso, como critério de seleção deste parâmetro, a primeira ocorrência de queda no valor da variância estimada, quando incrementado o fator espacial. Esta heurística sujeita o método a mínimos locais, possivelmente não identificando o valor ótimo para a região em análise.

Outro método definido por Castilla (2003), empregado em imagens de satélite, utiliza medidas de gradientes de luminância para estimar a magnitude dos gradientes espaciais, calculado para os eixo vertical e horizontal da imagem. A técnica faz uso do algoritmo segmentador Watershed (VINCENT; SOILLE, 1991) para efetuar a fusão das regiões, de acordo com as magnitudes dos gradientes. Como desvantagem, o método apresenta a necessidade de aplicação de filtros de suavização na imagem, para minimizar os efeitos dos mínimos locais dos gradientes.

A principal diferença entre os métodos existentes e o proposto neste trabalho é a avaliação variográfica do resultado obtido para a segmentação. Considera-se o valor esperado em cada segmento, mensurado em cada camada/atributo e o respectivo efeito pepita - obtido pela aproximação ao modelo gaussiano - empregado como parâmetro descritor da estrutura espacial presente, resultante da segmentação. Esta medida avalia não somente a dispersão dos valores esperados nos segmentos, como também sofre o efeito da variância total dos dados. Devido à característica de continuidade deste modelo, pequenas variações nos agrupamentos ocasionam mudanças nos valores para o efeito pepita e patamar. Assim, esta avaliação permite não somente identificar quando a variação entre os segmentos alcançou sua máxima dispersão, como também que a maior variância entre eles foi alcançada, portanto, que o melhor agrupamento por segmento foi obtido, como apresentado na seção 3.1.4 deste trabalho.

Em Ecologia e Biogeografia, os cientistas têm analisando por décadas a influência de gradientes ambientais em processos bióticos (SANDERS; RAHBEK, 2012). Conceitos como o de Nicho Ecológico (GRINNELL, 1917), Coenoclinas (GAUCH; WHITTAKER, 1972), Hipótese do Distúrbio Intermediário (CONNELL, 1978) e Gradiente de Estresse

(BERTNESS; CALLAWAY, 1994), são alguns dos conceitos que procuram explicar as mudanças em fatores bióticos por meio de gradientes ambientais abióticos. Muitos destes testes são baseados em análises de vetores de distância, cujo objetivo é a identificação dos efeitos de tendência causados pela autocorrelação espacial, principalmente por meio do arranjo espacial dos dados, a mensuração e quantificação do fenômeno espacial observado, e sua possível influência nos dados coletados. Borcard et al. (2004) desenvolveu uma técnica que utiliza mapas simulados com diferentes formas de autocorrelação espacial, em diferentes escalas, para extrair os principais descritores de um fenômeno espacial, técnica denominada Mapas de Moran ou *Principal Coordinates of Neighbour Matrices* (Seção 2.3.1). A técnica é empregada em Ecologia e Biogeografia para estimar e considerar o efeito da autocorrelação nos dados e melhorar os modelos obtidos (MARROT; GARANT; CHARMANTIER, 2015).

O método proposto neste trabalho pode fornecer uma ferramenta para estimar a extensão da autocorrelação espacial, facilitando a separação das regiões analisadas e o planejamento de coletas para melhoramento de modelos de distribuição de espécies.

Em Agricultura de Precisão, Zane et al. (2013) efetua a delimitação de Zonas de Manejo agrícola aplicando um algoritmo segmentador por crescimento de região. Em seu método, a segmentação é feita para uma única variável, e o variograma é calculado dentro cada cada segmento, para se obter o alcance. Além disso, a segmentação proposta pelo autor necessita da escolha de pontos iniciais e uma indicação prévia do número desejado de zonas. **Por outro lado, o método proposto neste trabalho possui a vantagem de apresentar como resultado do processo a quantidade “natural” de regiões em que o talhão pode ser dividido, não necessitando de uma definição prévia para o número de zonas ou a indicação de pontos iniciais aleatórios.**

3 Segmentação Multiresolução Variográfica Ótima

Neste capítulo apresentamos o método proposto nesta tese para identificação de estruturas espaciais através da delimitação dos gradientes para as variáveis regionalizadas, ao qual denominamos Segmentação Multiresolução Variográfica Ótima. A abordagem busca identificar um valor ótimo para o parâmetro de segmentação espacial (sp) aplicado ao algoritmo segmentador *Multiresolution* (BAATZ; SCHÄPE; STROBL, 2000) modificado para apresentar como saída, além dos segmentos, os valores esperados, denominado Função Segmentadora Multirresolução Multivariada (FSMM). Esta função considera as múltiplas camadas de informação sobrepostas, onde cada uma representa diferentes fenômenos espaciais para uma mesma região. A função avalia fatores relativos a forma e valores nos pontos para a geração dos segmentos criados com base na grade amostral empregada. Apesar do algoritmo *Multiresolution* ter sido definido inicialmente para o uso em imagens de sensoriamento remoto, a generalização da técnica por meio da definição desta função permite o emprego em diversos tipos de dados, desde que seja utilizada uma grade regular como entrada do algoritmo. Esta restrição é necessária em decorrência da avaliação de heterogeneidade de forma, conforme apresentado na seção 3.1.2.

No capítulo propomos quais características uma segmentação ótima deve apresentar em aplicações com dados georreferenciados numéricos para a delimitação dos fenômenos espaciais. A técnica proposta permite a mensuração adequada dos efeitos do segmentador para os gradientes de autocorrelação e a subsequente avaliação dos limites destes gradientes por meio da variância e dispersão dos dados, caracterizado principalmente, pelo efeito pepita presente no variograma aproximado ao modelo gaussiano.

3.1 Método proposto

O método proposto para a identificação de estruturas espaciais neste trabalho modifica a técnica de segmentação *Multiresolution* (BAATZ; SCHÄPE; STROBL, 2000), adicionando a análise variográfica das camadas resultantes do processo de segmentação para a obtenção do parâmetro de segmentação espacial (sp) considerado ótimo. Entende-se como ótimo, neste caso, o menor valor de sp para o qual a média dos valores de C_0 , obtidos em todas as camadas resultantes do processo de segmentação, atinge o seu primeiro máximo, ou no caso de ser possível a expansão da região avaliada, um máximo global. Deste modo, um variograma é construído para cada camada resultante do processo de segmentação, em seguida é calculada a média entre os valores de C_0 obtidos para cada

camada processada. A técnica busca identificar os segmentos que melhor representam a estrutura da região em estudo, por meio de seus gradientes espaciais.

Nos casos em que a região investigada apresenta apenas parte da região disponível, a técnica pode indicar a necessidade de expansão da janela de avaliação (*bounding box*). O método foi desenvolvido para ser aplicado em dados amostrados, dispostos em grades regulares sobrepostas. Portanto, a grade deve ser padronizada e constante entre as várias camadas para que ofereça o suporte necessário ao processamento. Os dados de entrada podem ser resultantes de métodos de interpolação espacial ou imagens *raster* contendo múltiplas camadas/atributos/bandas espectrais de dados provenientes de imagens de satélites de sensoriamento remoto.

3.1.1 Função Segmentadora Multiresolução Multivariada

A busca pela identificação de regiões similares por proximidade para múltiplos fatores de escala pode trazer diversos avanços nas análises geoespaciais, principalmente se forem considerados métodos interativos e de processamento *on-line* para geração de modelos. Para a redução de dimensionalidade dos dados e a geração de novas relações, objetos e segmentos podem ser identificados por meio dos gradientes de valores próximos. Estes gradientes representam a autocorrelação espacial e podem ser utilizados para delimitar as regiões por meio da identificação de seu alcance espacial, não somente pelos valores, mas também pela forma como o fenômeno se apresenta no espaço.

Denominamos Função Segmentadora, a função que, a partir de um determinado fator espacial (sp), atue como fator limitante para a heterogeneidade espacial - mensurada em termos de formato no espaço e de limites de variação dos atributos para dados georreferenciados em cada segmento - que quando aplicada aos dados dispostos em múltiplas camadas espacialmente sobrepostas, retorne a segmentação destes dados, fornecendo os limites para cada alcance de autocorrelação.

O termo multiresolução (ou multirresolução) refere-se às diferentes escalas (ou resoluções espaciais) em que um determinado objeto ou fenômeno pode estar representado em uma imagem. Um método pode ser definido como multiresolução quando pode ser adaptado as diferentes escalas que um fenômeno espacial pode se apresentar nos dados analisados (BAATZ; SCHÄPE; STROBL, 2000). Adotamos o termo multivariada pois a função segmentadora pode utilizar N variáveis espacialmente sobrepostas.

Para o problema de delimitação de regiões, seja por meio da identificação de gradientes espaciais ou objetos em uma imagem obtida de tomadas aéreas feitas por *drones* ou imagens de satélite multi ou hiperespectrais, a segmentação da imagem é um dos primeiros passos para a realização de inferências e atribuição de semântica aos dados observados. Uma das abordagens utilizadas para análise de imagens de sensoriamento

remoto é a técnica denominada *Geo Object Based Image Analysis* (Geo-OBIA) (HAY; CASTILLA, 2008). Esta técnica utiliza como passo inicial a segmentação da imagem em análise, de modo que se obtenha um conjunto de segmentos não sobrepostos, compostos por conjuntos significativamente homogêneos de pontos ou *pixels* para que, posteriormente, possa ser atribuído algum valor semântico a cada segmento (BLASCHKE, 2010). A técnica é utilizada em problemas de sensoriamento remoto, em que são processadas imagens multiespectrais, obtidas de satélites de observação terrestre. Como resultado final, os objetos são identificados a partir geração de regiões homogêneas em relação a todas as bandas espectrais disponíveis. Para a obtenção destes segmentos um algoritmo popular é o *Multiresolution* (BAATZ; SCHÄPE; STROBL, 2000).

O processo para a segmentação das imagens efetuado pelo *Multiresolution* pode ser descrito como uma técnica de fusão de regiões na imagem. Inicialmente cada ponto/*pixel* representa um objeto ou região. Em cada iteração os objetos são fusionados aos pares considerando como critério de decisão para a fusão a homogeneidade local. Esta homogeneidade é descrita como a similaridade (espacial e espectral) em relação aos objetos adjacentes. Para cada nova fusão, uma função de custo para a operação é avaliada. Se o valor deste custo for menor do que um certo limiar previamente definido, a fusão ocorrerá. O processo termina quando não há mais objetos passíveis de fusão (BAATZ; SCHÄPE; STROBL, 2000).

A partir dos segmentos gerados, é possível efetuar uma análise exploratória sobre os dados, para a execução de inferências sobre as hipóteses previamente levantadas, e que expliquem e atribuam significado aos segmentos obtidos. Nesta fase, podem ser utilizados algoritmos de classificação supervisionada como o gerador de árvores de decisão C4.5, a análise de componentes principais (PCA, do inglês, *Principal Component Analysis*) ou máquina de vetor de suporte (SVM, do inglês, *Support Vector Machine*); ou não-supervisionada como o K-means ou Verossimilhança Máxima (NAVULUR, 2006).

Outros parâmetros, além do fator espacial (*sp*), podem ser aplicados em algoritmos de segmentação. Estes parâmetros atuam como fatores limitantes durante o processamento dos respectivos atributos das regiões em análise (HAY; CASTILLA, 2006). Algumas ferramentas exploram o ajuste automático do parâmetro *sp* e outras, no entanto, utilizam segmentos de referência delineados manualmente para ajuste da função objetivo, com por exemplo o trabalho de Achancray et al. (2014). Em nosso caso, ajustamos o parâmetro sem auxílio externo. Todavia, os aspectos referentes aos ajustes de outros parâmetros não serão tratados neste trabalho. Recomendamos a leitura de Espindola (2007) e Ferreira (2011) para o leitor interessado em detalhes de ajustes de outros parâmetros.

3.1.2 Aspectos utilizados para caracterização da heterogeneidade espacial

Definição 3.1.1 (Fator de Fusão de Segmentos). A definição adotada para heterogeneidade possui um componente de avaliação espacial e outro para avaliação de similaridade de valores para cada ponto amostral, posição ou área (representada pelo *pixel* em uma imagem bidimensional). A função, definida por Baatz, Schäpe e Strobl (2000), é representada pela Eq. (3.1), como:

$$f = w_{val}h_{val} + (1 - w_{val})h_{form} \quad (3.1)$$

em que

f representa o fator de fusão para os segmentos (objetos);

h_{val} representa o fator de heterogeneidade de valor. Em Sensoriamento Remoto utilizado como fator espectral, portanto relativo ao valor espectral para o *pixel* ou conjunto de pixels na imagem obtida. Em nosso método representa um valor numérico;

h_{form} representa o fator de heterogeneidade de forma. Considera-se forma como a configuração dos atributos em relação a sua distribuição espacial;

w_{val} representa o fator peso ou importância relativa entre estes fatores.

Definição 3.1.2 (Fator de Heterogeneidade de Valor). O fator heterogeneidade de valor (h_{val}) avalia a diferença entre os desvios-padrão para cada novo segmento em relação ao segmento vizinho. Este fator é definido na Eq. (3.2) onde (HAPP et al., 2009):

$$h_{val} = \sum_c w_c (n_{s_3} \sigma_c^{s_3} (n_{s_1} \sigma_c^{s_1} - n_{s_2} \sigma_c^{s_2})) \quad (3.2)$$

no qual

s_1 é o segmento selecionado,

s_2 é o segmento candidato a fusão,

s_3 é o segmento resultante da união entre o segmentos s_1 e s_2 ,

c representa a camada de informação (ou banda espectral em Sensoriamento Remoto),

w_c é o peso definido para cada camada (ou banda),

$\sigma_c^{s_i}$ é o desvio padrão, relativo ao segmento s_i em cada camada c ,

n_{s_i} é o número de *pixels* pertencentes a cada segmento s_i .

Definição 3.1.3 (Fator de Heterogeneidade de Forma). O fator de heterogeneidade de forma h_{form} (Eq. (3.3)) é composto por dois atributos: compacidade (h_{comp}) e suavidade de forma (h_{smo}). A importância relativa entre estes fatores é ajustada pela componente w_{comp} como segue (HAPP et al., 2009):

$$h_{form} = w_{comp}h_{comp} + (1 - w_{comp})h_{smo} \quad (3.3)$$

onde

h_{comp} é definido como o cálculo para cada segmento s_i da razão entre o perímetro do segmento (l_{s_i}) e a raiz quadrada no número de *pixels* ($\sqrt{n_{s_i}}$). O fator é calculado para os segmentos s_1 , s_2 e s_3 durante o processo de avaliação da fusão. A forma completa para o cálculo do fator é dado pela Eq. (3.4):

$$h_{comp} = n_{s_3} \frac{l_{s_3}}{\sqrt{n_{s_3}}} - \left(n_{s_1} \frac{l_{s_1}}{\sqrt{n_{s_1}}} + n_{s_2} \frac{l_{s_2}}{\sqrt{n_{s_2}}} \right) \quad (3.4)$$

h_{smo} é definido como o cálculo para cada segmento s_i da razão entre o perímetro do segmento (l_{s_i}) e o perímetro de seu retângulo envolvente mínimo (b_{s_i}). A forma completa para o cálculo do fator é dado pela Eq. (3.5):

$$h_{comp} = n_{s_3} \frac{l_{s_3}}{b_{s_3}} - \left(n_{s_1} \frac{l_{s_1}}{b_{s_1}} + n_{s_2} \frac{l_{s_2}}{b_{s_2}} \right) \quad (3.5)$$

Definição 3.1.4 (Dados de Entrada - Camadas). Define-se $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{nc}\}$ como sendo o conjunto formado pelas nc camadas ($\mathbf{c}_i$) de dados completamente sobrepostos espacialmente e dispostos em uma grade regular (por exemplo: no caso de uma imagem de satélite, este conjunto seria uma imagem multi ou hiperespectral contendo nc bandas espectrais).

Definição 3.1.5 (Pesos para as Camada de de Entrada). Define-se o $W = \{w_1, w_2, \dots, w_{nc}\}$ como o conjunto formado pelos pesos atribuídos a cada camada de entrada da Def. 3.1.4. O peso de cada camada é aplicado na Def. 3.1.2 (w_c).

Definição 3.1.6 (Máscara Identificadora de Segmentos). Define-se o conjunto $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{no}\}$ como sendo o resultado final da segmentação para o algoritmo *Multi-resolution* (Eq. (3.1)) dado os parâmetros iniciais. Cada elemento $\mathbf{s}_i$ representa um dos no segmentos obtidos ao término do processamento.

Definição 3.1.7 (Camadas de Valor Esperado (Média) para os Segmentos). Define-se $\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{nc}\}$ como o conjunto contendo as nc camadas ($\mathbf{e}_i$), cada uma possuindo os valores esperados (média) resultantes do processamento do algoritmo *Multiresolution* (Eq. (3.1)) para a obtenção os segmentos $\mathbf{S}_{(sp)}$.

Definição 3.1.8 (Valor Esperado para o Segmento ao Término da Segmentação). Cada camada $\mathbf{e}_i$ pode ser descrita como uma função em que, dado a camada $\mathbf{c}_i$ e o segmento $\mathbf{s}_j$, retorna média dos pontos em $\mathbf{c}_i$ que fazem intersecção com os segmentos obtidos $\mathbf{s}_j$. Deste modo, definimos a Eq. (3.6) como:

$$\mathbf{e}_i(\mathbf{c}_i, \mathbf{s}_j) = \mu(\mathbf{s}_j \cap \mathbf{c}_i), \quad \text{para } i = 1, \dots, nc \quad \text{e } j = 1, \dots, no. \quad (3.6)$$

Definição 3.1.9 (Função Segmentadora Multirresolução Multivariada (FSMM)). Por fim, define-se como Função Segmentadora Multirresolução Multivariada ($\mathbb{B}$) como a execução do algoritmo *Multiresolution* modificado para apresentar o resultado como apresentado na Eq. (3.7):

$$\mathbb{B}(\mathbf{C}, \mathbf{W}, w_{val}, w_{comp}, sp) = \{\mathbf{S}_{sp}, \mathbf{E}_{sp}\} \quad (3.7)$$

A fusão entre os segmentos ocorrerá sempre que $f < sp^2$.

Neste estudo estamos interessados em medir o efeito do parâmetro sp , pois a alteração dos pesos relativos ou das camadas pode modificar significativamente o resultado final. Assim, adotamos neste trabalho que todas as camadas possuem sempre os mesmos pesos, os fatores de heterogeneidade de forma e valor apresentam a mesma contribuição relativa, assim como os fatores de compacidade e suavidade de forma. Deste modo, assume-se para todo o restante do texto que:

- todos os pesos das camadas são iguais e unitários, ou seja, $\mathbf{W} = \{w_1 = 1, w_2 = 1, \dots, w_{nc} = 1\}$;
- o peso relativo entre o fator de valor e de forma apresenta importância igual para fatores, ou seja, $w_{val} = 0,5$;
- peso relativo entre compacidade e suavidade de forma indica importância igual entre os fatores, ou seja, $w_{comp} = 0,5$.

3.1.3 Fluxo de processamento da função segmentadora multirresolução multivariada - FSMM ($\mathbb{B}$)

Inicialmente cada segmento (objeto) é definido como um único *pixel*. O segmento cresce na medida que ele é fusionado ao seu vizinho sempre que a função que minimiza a heterogeneidade incorporada (f) resulta em um valor inferior ao quadrado do fator de escala (sp) informado no início da execução. O processo continua até que os segmentos não possam mais crescer (BAATZ; SCHÄPE; STROBL, 2000; HAPP et al., 2009).

Assim, o processo de segmentação procura maximizar a heterogeneidade entre os segmentos, minimizando a heterogeneidade intrasegmento, por meio do critério de heterogeneidade ajustável dado pela Eq. (3.1). O ajuste do critério é feito pela escolha

do parâmetro de escala (sp), dos pesos atribuídos a cada uma das camadas (w_c), do peso atribuído ao fator de avaliação de valor das amostras (w_v) e do peso para o fator de avaliação de compactidade da forma (w_{comp}). O parâmetro sp tem impacto direto no tamanho dos segmentos, deste modo, quanto maior for o seu valor, maiores serão os segmentos.

Ao final do processo cada segmento resultante é enumerado sequencialmente, sendo que, cada posição de *pixel* na imagem irá receber uma marcação indicando a qual segmento este pertence. Para todo segmento s_j pertencente a $\mathbf{S}$ será associado a média para o conjunto de pontos que fazem intersecção em cada uma das camadas c_i pertencentes a $\mathbf{C}$, resultando nos segmentos $e_i(c_i, s_j)$ pertencentes a $\mathbf{E}$, sendo que, este valor depende do resultado da execução do *Multiresolution* para os parâmetros iniciais do algoritmo (Def. 3.1.9).

Relação da segmentação com a *Geo-Object Based Image Analysis*

Na abordagem Geo-OBIA a segmentação é a etapa inicial da análise da imagem e que oferece suporte para as etapas subsequentes (HAY; CASTILLA, 2008). Nestas etapas, a cada segmento pode ser atribuído um significado ou selecionado de acordo com o aspecto em análise (exemplo: tipo de vegetação, classe de produção ou qualidade do fruto de uma cultura ou região apresentando alguma doença, tipos de região de acordo com o clima). Esta atribuição pode ser feita por meio do uso de técnicas de classificação supervisionadas ou não-supervisionadas. A classificação dos objetos selecionados é feita de acordo com as camadas de informação utilizadas para efetuar a segmentação.

Para a aplicação clássica do algoritmo *Multiresolution* (problemas de classificação de imagens de sensoriamento remoto) o resultado final do processamento é apresentado em formato vetorial (tipicamente em arquivos ESRI shapefile). Neste formato cada segmento pode também apresentar algumas medidas em relação as camadas utilizadas (ex.: média dos valores dos *pixels* para cada camada, entropia máxima, entre outras).

A versão mais conhecida do algoritmo *Multiresolution* é encontrada no software eCognition (NUSSBAUM; MENZ, 2008). Nesta pesquisa utilizamos como referência para nossa implementação em R, uma versão modificada da função Baatz disponível na biblioteca Terralib (CÂMARA et al., 2008).

3.1.4 Avaliação variográfica de $\mathbb{B}$

Para caracterizar a variabilidade e a distribuição dos segmentos em relação a janela de avaliação, este trabalho propõe a avaliação variográfica das camadas resultantes da FSMM ($\mathbf{E}_{sp}$).

Devido ao fato do efeito pepita representar a variação aleatória para os fenômenos

de menor escala espacial (BOHLING, 2007), por meio de sua caracterização é possível mensurar as alterações das variabilidades mínimas entre os segmentos gerados e também avaliar o efeito da segmentação em relação a cada camada de dados de entrada em $\mathbb{B}$, Eq. (3.7). Para tal, a adoção de um modelo para a aproximação do variograma experimental (ex.: modelo gaussiano, esférico,...) sensível a distribuição espacial dos dados é essencial.

Diferentemente da abordagem tradicional, em que a análise variográfica é executada para a identificação do alcance e patamar, para uso na estimativa dos pesos utilizados no processo de interpolação via Krigagem (YAMAMOTO; LANDIM, 2013), nesta abordagem necessitamos apenas medir o efeito do agrupamento dos pontos para quantificar as modificações geradas pelo segmentador na autocorrelação espacial das camadas nas diferentes delimitações de tamanho dos segmentos. Para que isto seja feito, o modelo para a aproximação do variograma experimental deve utilizar uma função que apresente aproximação assintótica na origem (variância mínima) e também para o patamar (variância máxima) de modo a garantir que a configuração seja influenciada não somente pela amplitude de variação do variograma (patamar), que indica a variância máxima dos dados de cada camada, mas também pelo alcance (distância em que a autocorrelação espacial é considerada mínima para os pares de pontos da grade). Consequentemente, a função é utilizada deve sempre apresentar um valor máximo e mínimo para o variograma, e ser sensível a homogeneidade a pequenas distâncias, para que o efeito do agrupamento dos segmentos possa ser devidamente mensurado. Modelos de variogramas com estas características são denominados modelos de transição (HAINING, 1993). Nesta abordagem adotamos como modelo de transição o modelo gaussiano, por apresentar estas características previamente referidas e, portanto, maior sensibilidade e estabilidade quando calculados diversos variogramas para mapas diferentes.

O método assume que sempre estará disponível a melhor grade amostral possível para os dados, ou seja, o maior número disponível de pontos equidistantes para a região analisada. A geração dos segmentos contendo as médias (E) para cada camada e_i é avaliada do seguinte modo:

Para cada camada e_i segmentada, a aproximação para o variograma experimental ao modelo gaussiano, por meio do Método dos Mínimos Quadrados Generalizados (GLS - *Generalized Least Squares*) (AITKEN, 1936) foi empregada, utilizando o software R (HIEMSTRA et al., 2009). Em mapas que apresentam anisotropia, esta função avalia as amostras como um único variograma contendo todas as amostras. Deste modo, a função aproxima o variograma experimental anisotrópico, a um modelo isotrópico.

Usando a abordagem de Hiemstra et al. (2009) para a aproximação da função Gaussiana ao variograma, são aplicados os parâmetros iniciais para o algoritmo de aproximação ao variograma experimental: para o patamar (*sill*) o valor médio entre a semi-variância máxima e sua mediana; para o alcance (*range*) 1/10 da diagonal para o retângulo envol-

vente dos dados; e o valor inicial do efeito pepita (*nugget*) é definido como o valor mínimo para a semi-variância experimental calculada.

Nestas condições, o efeito pepita (C_0) apresenta, além da variabilidade para a mínima distância entre amostras, também mensura a dispersão dos segmentos em relação a janela de avaliação e ao parâmetro espacial empregado (sp). Assim, quanto mais compactos espacialmente os segmentos gerados (ou menos dispersos), menor será o efeito pepita. Como exemplo a Fig. 7 apresenta este fenômeno. Ressalta-se que para a aproximação empregando outras curvas, o efeito não necessariamente se apresenta, podendo até mesmo apresentar efeito pepita zero.

Curvas caracterizadoras para o valor ótimo de sp

Tendo como ponto de partida a combinação da avaliação pelo variograma experimental aproximado a curva Gaussiana, juntamente com o processo de segmentação apresentado na seção 3.1.2, é possível medir o efeito global da segmentação em todas as camadas, além da dispersão dos segmentos. Para fins de análise do comportamento do segmentador pode-se caracterizar o efeito da segmentação em dois momentos avaliados progressivamente até o maior valor possível de sp para a janela de avaliação, obtido pelo resultado de Eq. (3.7) até que S_{sp} apresente um número de segmentos inferior ou igual a que três, sendo que este limite foi estabelecido para se evitar uma possível situação em que o resultado da segmentação seria um único segmento contendo toda a área avaliada, o que resultaria em uma grade homogênea contendo o mesmo valor em todos os pontos, fator que impossibilitaria o cálculo do variograma.

Definição 3.1.10 (Primeiro Momento (PM)). O primeiro momento refere-se ao valor médio do efeito pepita calculado para todas as camadas resultantes do processo de segmentação para o parâmetro sp , representado pela Eq. (3.8).

$$\bar{\gamma}_{\text{pepita}}(sp) = \frac{1}{nc} \sum_{i=1}^{nc} C_0(\mathbf{e}_i), \text{ com } \mathbf{e}_i \in \mathbf{E}_{sp} \quad (3.8)$$

em que:

$\bar{\gamma}_{\text{pepita}}(\mathbf{sp})$ é o valor médio do efeito pepita C_0 calculado para cada camada $\mathbf{e}_i$ (Def. 3.1.7), obtida do resultado da execução da função dada pela Eq. (3.7) para um determinado fator de segmentação espacial sp (Def. 3.1.9).

Para exemplificar a caracterização da segmentação por meio do primeiro momento, utilizaremos a camada nomeada **L10** na Fig. 3 gerada para uma matriz 50 x 50 para melhor visualização. Como resultado da execução foram identificados 5 segmentos, apresentados na Fig. 8.

Figura 7 – Exemplo de relação entre o efeito pepita e a dispersão espacial dos dados. Na figura são apresentados os mapas e seus respectivos variogramas experimentais e resultantes da aproximação para a curva Gaussiana pelo método GLS utilizando a função do pacote automap para o software R (HIEMSTRA et al., 2009)

Fonte: Exemplo adaptado de Deutsch e Journel (1998, p. 303). No exemplo original, o variograma para o mapa (e) foi aproximado ao modelo esférico. Neste o modelo Gaussiano empregado no GLS não apresenta convergência em decorrência da elevação exponencial dos valores experimentais.

Figura 8 – Exemplo de aplicação do primeiro momento para avaliação da segmentação. O gráfico apresentado em (b) indica o valor máximo para o parâmetro sp , identificado pelo primeiro máximo quando o valor apresenta um patamar

(a) Segmentos para o mapa L10 com $sp=96$

(b) Primeiro momento calculado para o mapa L10 da Fig. 3

Fonte: Autor.

Definição 3.1.11 (Segundo Momento (SM)). Para avaliar dados representando regiões reais não assumindo *a priori* nenhuma afirmação sobre a região amostrada, este trabalho propõe verificar a existência do fenômeno espacial na janela de avaliação por meio da análise do efeito acumulado das segmentações sucessivas e os valores ótimos encontrados para a curva PM. Assim, definimos o Segundo Momento (Eq. (3.9)) como:

$$\Sigma \bar{\gamma}_{pepita}(sp) = \frac{1}{sp} \sum_{i=1}^{sp} \bar{\gamma}_{pepita}(i) \quad (3.9)$$

Esta necessidade surge principalmente porque curva PM apresenta apenas o comportamento instantâneo para a média do efeito pepita, em outras palavras, avalia o resultado da execução da FSMM para o parâmetro sp empregado. Portanto, não considera o efeito acumulado da segmentação durante todo o processo. O efeito acumulado neste caso, funciona como um filtro passa-baixa, situação pelo qual é pouco sensível a

variações bruscas (alta frequência), apresentando apenas as maiores tendências para o sinal (baixa-frequência). Neste contexto, é necessário que se estabeleça uma métrica para avaliar o efeito acumulado da segmentação para o incremento do fator sp . Este fator pode evidenciar a tendência global de diminuição da variabilidade e a existência de um valor máximo para sp para a região em análise. Assim, esta curva auxilia na identificação de uma tendência global acumulada para os valores medidos. Deste modo pode-se detectar a necessidade de expansão da região avaliada, caso seja possível.

O SM tem como objetivo identificar a ocorrência de um máximo global de variância acumulada para o Primeiro Momento, ponto a partir do qual a tendência de diminuição dos valores de $\bar{\gamma}_{\text{pepita}}(i)$ pode ser evidenciado por meio da avaliação do comportamento da curva para todos os valores de sp $[1; sp_{\text{max}}]$. Esta tendência de diminuição na curva PM indica a existência de ao menos duas camadas contendo mais de um fenômeno parcialmente correlacionado espacialmente na janela de avaliação e que os segmentos gerados atingiram seu tamanho máximo. **Em outras palavras, dado que o resultado da segmentação delimita os segmentos que minimizam a variância interna (intra-segmento) e maximizam a variância externa (entre os segmentos), a variabilidade total atingiu seu valor máximo para todas as camadas no conjunto C processadas em $\mathbb{B}$.**

Definição 3.1.12 (Condição de Parada). A caracterização do ponto máximo global para $\Sigma\bar{\gamma}_{\text{pepita}}(sp)$ é dada pela detecção do ponto crítico positivo, definido pela primeira e segunda derivada discreta na curva SM como:

$$\frac{d\Sigma\bar{\gamma}_{\text{pepita}}}{dsp} = \frac{\Sigma\bar{\gamma}_{\text{pepita}_{sp+1}} - \Sigma\bar{\gamma}_{\text{pepita}_{sp-1}}}{2\Delta sp} = 0 \quad (3.10)$$

e

$$\frac{d^2\Sigma\bar{\gamma}_{\text{pepita}}}{dsp^2} = \frac{\Sigma\bar{\gamma}_{\text{pepita}_{sp+1}} - 2\Sigma\bar{\gamma}_{\text{pepita}_{sp}} + \Sigma\bar{\gamma}_{\text{pepita}_{sp-1}}}{(\Delta sp)^2} < 0 \quad (3.11)$$

A detecção deste ponto crítico pode ser empregada como fator auxiliar na detecção de condição de parada de processamento, pois indica que o maior valor para a curva PM foi detectado.

Como heurística adicional para a condição de parada de processamento, adotou-se, além do número mínimo de três segmentos, a ocorrência de valores para $\bar{\gamma}_{\text{pepita}}(i)$ constantes e inferiores a 0,01 por mais de 20 incrementos de sp .

Por meio da caracterização progressiva da segmentação, pode-se identificar a partir de qual valor para sp ocorre a variância máxima para os segmentos gerados, caracterizando assim, a extensão da dependência espacial para os segmentos na janela de avaliação.

Definição 3.1.13 (Parâmetro Espacial Ótimo sp_{opt}). Como o efeito pepita é sensível a dispersão espacial dos valores, o ponto máximo para a curva PM indica que existe um

valor de sp em que os segmentos obtidos apresentaram a maior dispersão possível para a máxima variância. Assim, detectadas as condições das Equações (3.10) e (3.11), temos que o parâmetro sp ótimo é dado pela Eq. (3.12) como:

$$sp_{\text{opt}} = \arg \max_{sp} (\bar{\gamma}_{\text{pepita}}(sp)) \quad (3.12)$$

em que sp_{opt} representa o parâmetro espacial ótimo obtido pela avaliação da curva PM por meio da detecção do ponto máximo global para o SM.

3.1.5 Delimitação de regiões pelo valor máximo do Primeiro Momento

A homogeneização de uma região por meio de um algoritmo segmentador baseado em inundação como o Multiresolution pode ser caracterizada pela queda do valor do Primeiro Momento para a camada avaliada (queda da variância mínima no mapa, ou máxima dispersão alcançada para os segmentos), na medida em que regiões com diferenças significativas para seus valores médios são fusionadas implicando, em última instância, na existência de um único valor médio para toda a área analisada. Em problemas onde é necessário avaliar múltiplas camadas/atributos/variáveis simultaneamente, espera-se que alguns dos fenômenos amostrados apresentem sobreposição espacial para os gradientes de correlação espacial, ainda que parcial. Deste modo, ao avaliar a FSM, o máximo para a curva PM indica que o máximo de autocorrelação foi obtido para os segmentos na janela de avaliação.

Para simular um conjunto de variáveis regionalizadas parcialmente correlacionadas, em outras palavras, que possuam sobreposição parcial de gradientes, utilizamos os dados apresentados na Fig. 3. Cada variável PCNM 2D empregada como entrada em $\mathbb{B}$ representa uma variável regionalizada com autocorrelação espacial conhecida. Quanto maior o índice da variável, maior a frequência espacial e menor a escala do fenômeno simulado. O que se observa a partir de testes executados neste dados é a existência de um valor máximo global para o Primeiro Momento ($\bar{\gamma}_{\text{pepita}}(sp)$), em variáveis que apresentem valores significativos para o Índice de Moran, tanto positivos quanto negativos, desde que as camadas apresentem fenômenos ao menos parcialmente distintos.

A Fig. 9 apresenta o gráfico com a avaliação de $\bar{\gamma}_{\text{pepita}}(sp)$ empregando as camadas geradas para as variáveis PCMN 1, 2, 5, 10, 20 e 1274 (última camada a apresentar Índice de Moran positivo) para uma grade de 50 x 50 pontos.

Para generalização do conceito proposto e visualização do comportamento da curva do PM para a detecção do valor ótimo de sp , executou-se a inserção gradual de todas as variáveis PCNM geradas para uma grade de 20 x 20 pontos (incremento unitário do índice i em $\Delta[i]$) no conjunto $\mathbf{C}$ processado em $\mathbb{B}$ seguida de avaliação do PM e SM. Conhecendo-se *a priori* o Índice de Moran para a variável regionalizada e sabendo-se que as variáveis com maior índice para o autovetor de Δ apresentam fenômenos com menor

Figura 9 – Avaliação do Primeiro Momento para as variáveis PCNM 1, 2, 5, 10, 20, 50, 100, 150 e 1274

Fonte: Autor.

escala espacial, portanto com maior frequência de ocorrência na janela, observou-se que o valor máximo global para o PM pode ser detectado com pelo menos três camadas. O resultado é apresentado na Fig. 10.

Uma vez estabelecidas todas as definições necessárias para a obtenção do valor ótimo para sp , o algoritmo utilizado para o cálculo da curva PM é apresentado no fluxograma da Fig. 11.

3.1.6 Caracterização de janela ótima para segmentação por valor máximo do Segundo Momento

Caso seja possível a expansão da região analisada, pode-se utilizar as condições de parada (Def. 3.1.12) para se avaliar qual foi a condição detectada. Caso o ponto máximo tenha sido detectado, ou seja as equações (3.10) e (3.11) forem verdadeiras, não é necessário a expansão da janela de avaliação, pois o ponto máximo para a curva PM foi detectado, portanto o valor de sp_{opt} pode ser obtido.

Caso o ponto máximo não tenha sido detectado, a janela de avaliação deve ser aumentada para que uma região maior seja avaliada. Quando pelo menos 3 ou mais variáveis apresentarem fenômenos espaciais que apresentam autocorrelação espacial parcial, a condição de parada poderá ser detectada, fornecendo o parâmetro sp_{opt} para a região selecionada. O algoritmo utilizado para o cálculo da curva SM é apresentado no fluxograma da Fig. 12.

Concluindo, a Segmentação Multiresolução Variográfica Ótima é definida para o valor de sp quando a curva PM apresenta valor máximo e não é possível a expansão da janela de avaliação. Quando a expansão é possível, é definida pelo máximo para a curva SP, quando a curva SM apresenta um máximo global.

Figura 10 – Avaliação da curva PM para todas as variáveis PCNM geradas em uma grade de 20×20 pontos. É possível de se detectar o máximo global para o PM a partir de 3 ou mais variáveis (Def. 3.1.12). Esta detecção é indicada pela mensagem “Stop Condition = TRUE” em cada figura

(a) PM avaliado para as 2 primeiras variáveis PCNM geradas

(b) PM avaliado para as 3 primeiras variáveis PCNM geradas

(c) PM avaliado para as 84 primeiras variáveis PCNM geradas

(d) PM avaliado para as 209 primeiras variáveis PCNM geradas

Fonte: Gerado pelo autor a partir da função `features` do pacote `features` para R (VARADHAN et al., 2011)

Figura 11 – Fluxograma com o algoritmo utilizado para o calculo da curva PM. Por meio desta curva o valor de sp_{opt} é obtido (diretamente, quando não é possível se expandir a janela de avaliação ou com auxílio da curva SM, quando a região avaliada puder ser expandida)

Fonte: Autor.

Figura 12 – Fluxograma com o algoritmo utilizado para o calculo da curva SM. Esta curva apresenta a inércia em relação a curva PM. Por meio da curva SM o valor de sp_{opt} é obtido pela detecção de um valor máximo global para a curva PM. Caso não seja detectado, a janela de avaliação deve ser expandida e o processo de cálculo das curvas SM e PM deve ser reiniciado

Fonte: Autor.

4 Aplicações e Resultados

Neste capítulo apresentamos alguns exemplos de aplicabilidade para a Segmentação Multiresolução Variográfica Ótima. Em cada um dos exemplos, a Análise de Componentes Principais (PCA) foi utilizada como ferramenta para redução de dimensionalidade para análise e geração de gráficos que permitissem visualizar como os dados estão organizados. O objetivo da redução de dimensões é facilitar a análise e interpretação dos dados, ao mesmo tempo em que se procura preservar a maior parte da informação (quantificada em termos de variação) contida nos mesmos. São apresentados três exemplos. O primeiro apresenta como a técnica pode auxiliar na definição de Zonas de Manejo agrícola. No segundo, apresenta-se uma aplicação no domínio da biogeografia em que o método pode ser empregado para selecionar regiões relacionadas ao nicho ecológico das espécies em análise. O último exemplo apresenta como a técnica pode auxiliar na tomada de decisão na definição de áreas prioritárias para a conservação da biodiversidade.

4.1 Estudo de caso: identificação de zonas de manejo / gerenciamento diferenciado em Agricultura de Precisão

Como exemplo para aplicação do algoritmo proposto, utilizamos o problema de identificação de Zonas de Gerenciamento Diferenciado ou Zonas de Manejo em Agricultura de Precisão (AP). Assume-se neste contexto a definição de AP como sendo um método de gestão do sistema de produção agrícola que se utiliza de tecnologia de informação para a identificação da variabilidade, buscando um tratamento adequado que maximize a sustentabilidade dos retornos (MACHADO; BERNARDI; SILVA, 2004, p. 199). A AP é formada por três componentes: obtenção dos dados em escala adequada, interpretação e análise dos dados, e implementação de gerenciamento em escala e tempo adequados (OLIVER, 2010).

A busca e identificação dos fatores que causam a variabilidade na produção e possibilitam o planejamento e a busca de soluções para o convívio e quantificação financeira destas variabilidades é um dos propósitos principais da AP. Normalmente a variabilidade é identificada pela presença de manchas nos mapas de produtividade agrícola (MACHADO; BERNARDI; SILVA, 2004, p. 200). Este objetivo norteia a idéia de Zonas de Manejo ou Zonas de Gerenciamento Diferenciado. O conceito foi definido por Doerge (1999) como a separação de um talhão em sub-regiões que representam uma combinação homogênea de fatores. Cada sub-região pode receber uma quantidade fixa dos insumos, visando maximizar o aproveitamento e diminuir os impactos ambientais. Assim, o objetivo das

Zonas de Manejo é caracterizar os limites de variabilidade, para então tratar cada região como uma área homogênea, que possa ser gerenciada separadamente, de modo a aumentar a produtividade, minimizar o consumo de insumos e maximizar os retornos financeiros (KHOSLA et al., 2010).

Neste exemplo, propomos a identificação das variações de nutrientes e textura do solo por meio da Segmentação Multiresolução Variográfica Ótima. A abordagem pode ser empregada como uma ferramenta de análise que facilite aos técnicos e engenheiros agrícolas a identificação das regiões presentes dentro de um talhão e os principais fatores responsáveis pelas diferenças entre as regiões. Considerando que a produtividade de uma Zona de Manejo seja fruto de um conjunto de condições que apresentam um considerável grau de homogeneidade, a identificação e delimitação destas regiões pode ser um fator importante para as subsequentes análises e levantamento dos fatores predominantes em cada zona delineada. Para um maior aprofundamento sobre as diferentes definições e métodos existentes para caracterização de Zonas de Manejo recomendamos a leitura de Oliver (2010) ou Santos (2014).

4.1.1 Dados e área de estudo

Os dados foram obtidos de uma fazenda particular no estado do Paraná, no município de Tibagi, em um talhão de trigo semeado em setembro de 2008. Um total de 67 amostras de solo foram coletadas em março de 2009, logo após a colheita, em uma área de 1,6 km x 1,2 km (133 ha aproximadamente). Nas amostras foram estimados por meio de análise laboratorial os níveis de: pH, Acidez potencial - H+AL (mmolc/dm³), Alumínio - Al (mmolc/dm³), Cálcio - Ca (mmolc/dm³), Magnésio - Mg (mmolc/dm³), Potássio - K (mmolc/dm³), Soma de bases trocáveis - SB (%), Capacidade de troca de cátions - CTC (mmolc/dm³), Fósforo - P (mg/dm³), Carbono - C (g/dm³), Matéria orgânica - MO (g/dm³), Porcentagem de saturação por bases - V (%), Silte (%), Areia (%) e Argila (%). A distribuição das amostras é apresentada na Fig. 13.

Para possibilitar a utilização pelos algoritmos, os dados de pH, H+Al, Al, Ca, Mg, K, MO, P, C, Silte, Areia e Argila foram interpolados. Foi utilizada uma grade com células de 30 m x 30 m. O algoritmo de interpolação empregado foi a Krigagem Universal (KU), pois algumas camadas apresentaram tendência linear, como recomendado por Diggle e Ribeiro Jr. (2010). A KU foi empregada por meio do uso da função `interpolateBlock` pertencente ao pacote `intamap` (HENGL et al., 2007) para o software R (R Core Team, 2015). A função foi parametrizada para utilizar o pacote `automap` (HIEMSTRA et al., 2009) para cálculo automático do variograma empregado. Ao utilizarmos a parametrização automática, assumimos que os variogramas gerados em mapas que apresentam anisotropia são aproximados a modelos isotrópicos segundo o método definido por Hiemstra et al. (2009). O resultado da interpolação (Fig 14) foi multiplicado por 1000 para reduzir erros

Figura 13 – Localização dos pontos amostrados no talhão analisado

Fonte: Costa et al. (2013).

de aproximação e diminuir o tempo de processamento.

Devido ao denso e elevado número de amostras coletadas (intervalo médio entre pontos inferior a 2 m, e total > 90.000 pontos) para os dados de produtividade de massa seca e altitude disponibilizadas para o talhão, estes foram interpolados utilizando o método Inverso da Distância ao Quadrado implementado pela função `idw` do pacote `gstat` para R (PEBESMA, 2004). Os dados de produtividade e altitude não são utilizados durante a segmentação por meio do FSMM.

4.1.2 Análise de componentes principais (PCA)

A análise de componentes principais é o mais antigo método de ordenação e também o mais conhecido e com diversas aplicações em Geociências (LANDIM, 2011). Por ser uma das técnicas mais comuns na literatura e amplamente empregada para análise multivariada será utilizada como heurística de redução de dimensionalidade e visualização das relações entre os dados. A técnica permite o estudo de múltiplas variáveis quantitativas intercorrelacionadas por meio da projeção dos dados aos eixos que maximizam a soma do quadrado das correlações (ABDI; WILLIAMS, 2010). Esta soma é o autovalor de um determinado eixo e representa o maior grau possível de correlação entre as variáveis e cada eixo, denominado componente.

Para cada componente, a porcentagem de variabilidade explicada por eixo será igual ao autovalor dividido pelo número de variáveis. O eixo com maior autovalor será o eixo principal, portanto, o que melhor representa as correlações entre as variáveis, o segundo eixo

Figura 14 – Mapas gerados para as camadas de dados interpolados por meio da Krigagem Universal. Utilizou-se os identificadores i_HAl para representar os dados interpolados a partir das amostras de H+Al, i_pH para as amostras de pH do solo, e assim por diante. Os valores estão multiplicados por 1000 e suas respectivas unidades estão descritas no início desta seção. O círculo em vermelho indica uma possível região de interesse, por apresentar um padrão similar em pelo menos 6 dos 12 mapas gerados

será o que possui o segundo maior valor e assim sucessivamente. A componente principal é o arranjo que melhor representa a distribuição dos dados e a componente secundária é a segunda com maior representação da variabilidade, sendo perpendicular a principal, e assim por diante. O número total de componentes principais é igual a quantidade de variáveis utilizadas, neste caso, a dimensionalidade do conjunto dos dados.

Cada ponto dos dados projetado nos eixos das componentes principais é denominado *escore*. O ângulo relativo entre cada variável (ex. *i_ph*) e um eixo de cada componente é denominado *carregamento* (*loading*). Por meio da projeção dos carregamentos multiplicado pelo respectivo desvio padrão da variável obtém-se a direção e o módulo do vetor que permite estimar a influência de cada variável nas amostras. Os autovetores correspondem as componentes principais e são formados pelo carregamento das variáveis originais. Os carregamentos podem ser considerados uma medida da importância relativa de cada variável em relação a cada componente principal, e seus respectivos sinais indicam relações diretas e inversamente proporcionais (LANDIM, 2011). Estatisticamente, a PCA representa a transformação de um conjunto de variáveis correlacionadas em um conjunto de combinações lineares que formam variáveis não correlacionadas - as componentes principais. Geometricamente, a PCA reposiciona o centroide dos pontos e rotaciona os eixos até apresentar uma configuração em que a variância máxima para os dados seja representada (CANGELOSI; GORIELY, 2007).

Pré-processamento: padronização ou normalização dos dados

Quando as variáveis apresentam diferentes escalas pode ser necessário uma normalização ou padronização preliminar, de modo que as variáveis apresentem média zero e variância unitária (transformação “z”). Deste modo, pode-se ressaltar as variâncias das variáveis que influenciam nas cargas fatoriais (*loadings*) para o subespaço PCA (LANDIM, 2011).

Por exemplo, algumas variáveis podem apresentar uma grande variância e outras pequenas. Neste caso, as variáveis com baixa variância apresentarão valores pequenos em suas cargas fatoriais para as componentes principais, enquanto que variáveis com significativa variância apresentarão valores expressivos em suas cargas fatoriais. A normalização faz com que todas as variáveis sejam representativas no subespaço PCA independente de sua escala. Por outro lado, se a escala relativa entre as variáveis for importante, a normalização pode não ser necessária.

Para comparação entre os efeitos da normalização nos diferentes padrões espaciais de delimitação das regiões obtidos, neste exemplo, as avaliações utilizam as variáveis normalizadas e não-normalizadas.

Outras transformações também podem ser aplicadas, como a transformada log ou o uso de pesos. A escolha do método depende do analista de dados e do conhecimento de

quais características cada transformação pode ajudar a ressaltar. Contudo, deve-se ter em conta de que toda a alteração nos dados modificará os autovalores e autovetores para a PCA.

Seleção das componentes principais

Existem diversos métodos, estatísticos ou heurísticos, para seleção das componentes principais mais relevantes. Alguns dos métodos são: o modelo de broken-stick, o teste de Kaiser-Gutman, o diagrama log-autovalor (LEV), o teste SCREE de Cartel, a validação cruzada, a técnica *bootstrapping* e a porcentagem acumulada da variância total. Uma descrição completa sobre estes métodos e outros pode ser encontrada em [Jackson \(2005\)](#) e também em [Cangelosi e Goriely \(2007\)](#).

Nesta análise utilizamos a primeira e a segunda componente principal (PC1 e PC2), caracterizadas como suficientes para representação da variabilidade dos dados pelo teste broken-stick ([JACKSON, 1993](#)), tanto para o caso em que os dados foram normalizados, quanto para a execução dos dados brutos. O método broken-stick apresenta como desvantagens não considerar o tamanho das amostras e subestimar as dimensões dos dados em alguns casos ([FRANKLIN et al., 1995](#)). Ainda assim, tem como vantagem ser um teste simples que, segundo [Jackson \(1993\)](#), apresenta bons resultados quando a amostragem possui regularidade, como é o caso dos dados dispostos em grade deste exemplo.

O teste de broken-stick é baseado nos autovalores dos dados aleatorizados. O método assume que se a variância total (soma dos autovalores) for dividida aleatoriamente entre os componentes, então a distribuição esperada seguirá o modelo broken-stick ([CANGELOSI; GORIELY, 2007](#)). Os autovalores considerados interpretáveis serão aqueles que excederem os valores gerados pela curva broken-stick ([CANGELOSI; GORIELY, 2007](#)). Assim, os componentes (eixos) selecionados são aqueles que apresentam variabilidade maior do que o esperado ao acaso. O resultado do teste executado para os dados do talhão, gerado pela execução da função `screeplot` - que implementa este teste no pacote `vegan` para o R - é apresentado na Fig. 15.

Neste caso, as componentes (PC1 e PC2) representam respectivamente 47,1 % e 39,6 %, totalizando 86,7 % da variabilidade para os dados não-normalizados (Tabela 1), e 44,7 % e 21,6 % respectivamente, totalizando 66,3 % da variância para o caso em que os dados foram normalizados (Tabela 2). Apesar de não explicar totalmente a variância, [Ferré \(1995\)](#) sugere que desde que a PCA seja utilizada para uso descritivo, como é o caso neste exemplo, uma solução que apresente menos dimensões do que o total é aceitável, pois não é possível de se afirmar qual é a verdadeira dimensionalidade dos dados.

Os carregamentos das componentes para os dados não-normalizados e normalizados são apresentados nas Tabelas 3 e 4 respectivamente. Os valores destacados representam as

Figura 15 – Resultado do teste Broken Stick para seleção das componentes principais para os dados não-normalizados e normalizados. Os gráficos demonstram que para ambos os casos somente as componentes PC1 e PC2 apresentam valores superiores a distribuição Broken-Sick

Fonte: Gráfico gerado pelo autor com auxílio da função `screeplot`, pacote `vegan` para R (OKSANEN et al., 2016).

Tabela 1 – Importância dos componentes PCA (até a 5 componente): dados não-normalizados

	PC1	PC2	PC3	PC4	PC5
Desvio padrão	13,80	12,65	6,27	3,30	1,37
Proporção da variância	0,47	0,39	0,09	0,02	0,00
Proporção acumulada	0,47	0,86	0,96	0,99	0,99

Fonte: Autor.

Tabela 2 – Importância dos componentes PCA (até a 5 componente): dados normalizados

	PC1	PC2	PC3	PC4	PC5
Desvio padrão	2,38	1,73	1,12	0,96	0,71
Proporção da variância	0,47	0,25	0,10	0,08	0,04
Proporção acumulada	0,47	0,72	0,83	0,90	0,94

Fonte: Autor.

relações de proporcionalidade (inversa ou direta) mais importantes. Para o caso dos dados não-normalizados, valores superiores a $|5, 0|$ foram considerados significativos e para os dados normalizados, $|0, 8|$.

A distinção entre os dois tipos é necessária pois as diferenças entre as variâncias altera o formato da transformação, consequentemente alterando o posicionamento das variáveis no subespaço. Se compararmos as variáveis significativas entre as duas tabelas, pode-se perceber que na Tabela 3 as variáveis de maior importância (maiores valores absolutos para as cargas) são: `i_hal`, `i_p`, `i_areia` e `i_argila`; e para a Tabela 4 as variáveis `i_ph`, `i_ca`, `i_mg`, `i_areia` e `i_argila` são as mais significativas além das diferenças consideráveis em amplitude dos valores das correlações. Esta alteração certamente modifica

Tabela 3 – Carregamentos: dados não-normalizados

	PC1	PC2	PC3	PC4	PC5
i_ph	0,19	0,15	0,02	0,12	0,02
i_hal	-4,02	-10,36	2,44	-0,43	-0,31
i_al	-0,53	-0,80	0,08	-0,04	-0,02
i_ca	3,07	1,77	0,44	1,09	-0,91
i_mg	1,74	1,01	0,24	0,63	-0,54
i_k	-0,01	0,16	0,03	0,27	-0,21
i_mo	1,66	-2,42	1,13	2,40	0,40
i_p	9,71	-6,06	-3,20	-0,41	0,00
i_c	0,96	-1,41	0,66	1,39	0,23
i_silte	-0,32	0,04	0,17	-0,18	-0,62
i_areia	-5,72	-1,41	-3,44	1,04	-0,09
i_argila	5,55	1,23	3,07	-0,30	0,16

Fonte: Autor.

Tabela 4 – Carregamentos: dados normalizados

	PC1	PC2	PC3	PC4	PC5
i_ph	0,90	0,10	0,01	0,25	-0,20
i_hal	-0,64	-0,71	0,10	-0,15	0,00
i_al	-0,77	-0,55	0,07	-0,16	-0,08
i_ca	0,96	0,08	0,10	-0,05	-0,05
i_mg	0,96	0,08	0,10	-0,05	-0,05
i_k	0,14	0,15	0,83	0,38	0,28
i_mo	0,29	-0,91	0,18	0,16	-0,17
i_p	0,45	-0,62	-0,18	-0,11	0,56
i_c	0,29	-0,91	0,18	0,16	-0,17
i_silte	-0,33	0,27	0,65	-0,57	-0,08
i_areia	-0,84	0,13	0,00	0,42	-0,01
i_argila	0,89	-0,17	-0,02	-0,33	-0,02

Fonte: Autor.

o resultado de algoritmos de classificação como o *K-means*, pois reposiciona os escores para cada ponto no subespaço, como apresentado na seção 4.1.3.

Visualização dos agrupamentos PCA

Por meio da proximidade relativa entre os centroides dos agrupamentos formados pelas projeções aos eixos das componentes principais selecionadas, pode-se visualizar quais conjuntos (segmentos ou grupos) apresentam maiores semelhanças. Deste modo, o gráfico biplot é utilizado aqui como ferramenta auxiliar na interpretação dos padrões espaciais obtidos.

4.1.3 Abordagens

Para verificar as diferenças entre o uso da FSMM em relação a uma das abordagens comuns para identificação de Zonas de Manejo, apresentamos neste seção uma comparação utilizando o algoritmo *K-means* (HARTIGAN; WONG, 1979), por ser uma das soluções amplamente utilizadas neste tipo de problema (SANTOS, 2014).

K-means

O algoritmo *K-means* é uma das muitas abordagens possíveis para a definição de Zonas de Manejo (KHOSLA et al., 2010; SANTOS, 2014). O algoritmo faz parte da classe de classificadores não supervisionados. Seu objetivo é dividir um conjunto de M amostras contendo N dimensões em K grupos disjuntos, de modo que a soma do quadrado das distâncias seja mínima em cada grupo (HARTIGAN; WONG, 1979). Estas somas representam as médias μ_j das amostras x_i para cada grupo G e seus respectivos centroides. O algoritmo então tende a encontrar os centroides que minimizam a inércia, dada pela Eq. (4.1):

$$\sum_{i=0}^n \min_{\mu_j \in G} (||x_i - \mu_j||^2) \quad (4.1)$$

Esta inércia é utilizada como uma medida de coerência interna entre os grupos. Segundo Arthur e Vassilvitskii (2007), esta medida possui alguns problemas elencados a seguir:

- A inércia assume que os grupos são convexos e isotrópicos, o que nem sempre é verdade. A resposta do algoritmo a grupos alongados ou distribuídos de forma irregular é pobre, portanto, não separa devidamente estes tipos de grupos. Este certamente é o caso deste exemplo, pois os dados coletados sobre os nutrientes ou textura do solo estão espalhados ao longo do terreno analisado.
- A inércia não é uma métrica normalizada, portanto, somente se pode afirmar que quanto menor o valor obtido, melhor é a estimativa. Em espaços de alta-dimensionalidade, a distância euclidiana representa uma hipersfera. Se considerarmos que o espaço d -dimensional forma um hipercubo, o volume do hipercubo se mantém constante, enquanto o volume de uma hipersfera torna-se cada vez menor ao adicionarmos novas dimensões. Em espaços com elevado número de dimensões, uma grande proporção dos dados utilizados para encontrar os grupos reside nos extremos do hipercubo que representa o espaço amostral da característica analisada (BEYER et al., 1999). Assim, as instâncias dos dados nos extremos do espaço amostral dificilmente são classificadas corretamente por classificadores que utilizam este tipo de medida de distância (Euclideana, Manhattam, Mahalanobis). Neles, a medida de distância tende a perder sua efetividade com o aumento da quantidade de

dimensões empregada na classificação dos grupos. Este fenômeno é conhecido como “Maldição da dimensionalidade (*Curse of dimensionality*)” (BELLMAN; DREYFUS, 2010).

Uma saída para minimizar estes efeitos é executar um pré-processamento dos dados, aplicando técnicas de redução de dimensionalidade, antes da execução do classificador. Uma destas técnicas pode ser a PCA, por fornecer os dados representados em um subespaço com número reduzido de dimensões. Estas dimensões são não correlacionadas e representam uma combinação linear dos dados originais em que maximiza-se a variabilidade (DING; HE, 2004).

Consideradas estas limitações do *K-means*, observa-se que:

1. Se o *K-means* for utilizado para delineamento das Zona de Manejo, não é recomendado obter os grupos diretamente dos dados sem um pré-processamento para redução da dimensionalidade, devido aos problemas previamente expostos.
2. Se os dados não forem normalizados, a medida de inércia do *K-means* apresentará resultados diferentes em relação aos normalizados, portanto, alterando os grupos.

Segundo Ding e He (2004), a redução de dimensões via PCA executa o agrupamento dos dados de acordo com a definição da função objetivo do *K-means*. Este fato melhora o desempenho do classificador.

Assim, os dados foram inicialmente processados via PCA, para somente então serem agrupados via *K-means* de 2 modos diferentes: 1) a PCA é executada nos dados brutos, sem normalização. 2) a PCA é executada nos dados normalizados para média 0 e variância 1. A geração das Zonas de Manejo com o *K-means* apresenta como vantagem a obtenção da segmentação já classificada em termos de similaridades dos grupos em relação a sua proximidade no subespaço transformado para as componentes principais empregadas.

Para identificar a quantidade ideal de agrupamentos, uma solução comum é comparar as inércias (somas dos quadrados das distâncias entre cada membro do grupo e o seu respectivo centroide) para um número crescente de grupos obtidos. A medida que o número de k de grupos é incrementado, o valor das inércias obtidas para cada grupo tende a diminuir, e esta soma tende diminuir na medida em que os grupos tendem a ser cada vez menores. O ponto a partir do qual a curva passa a ter um comportamento aproximadamente linear pode ser adotado como o número ótimo de grupos para o *K-means*. Neste ponto do gráfico, a curva apresenta uma inflexão (“cotovelo”) a partir da qual as diferenças entre os valores são relativamente pequenas (THORNDIKE, 1953), tendendo a linearidade. A curva para os dados normalizados (seção 4.1.2) é apresentada na Fig. 16 em que esta heurística aponta que os dados podem ser divididos em 5 agrupamentos. O

teste, neste caso, foi executado com 25 repetições para cada valor k de grupos empregado. Em cada repetição, os centroides dos grupos foram distribuídos aleatoriamente no início de cada estimativa.

Figura 16 – Curvas pra a medida de inércia do *K-means* para os dados não-normalizados e normalizados no subespaço PCA. O “cotovelo” da curva indica que a partir de 5 grupos o somatório das inércias passa a apresentar pequenas variações. Neste caso, $k = 5$ que pode ser definido como o número ótimo de grupos

Fonte: Autor.

Definidos o total de 5 agrupamentos, a classificação dos grupos para os dados não-normalizados e normalizados é apresentada na Fig. 17(a) e Fig. 17(b), respectivamente.

Segmentação Multiresolução Variográfica Ótima

Ressalta-se que, ao contrário do *K-means*, a FSMM não é um algoritmo classificador e apenas segmenta os dados de acordo com a definição de heterogeneidade apresentada na seção 3.1.2. A classificação dos segmentos pode ser feita posteriormente, seguindo os mesmos princípios aplicados em Geo-OBIA (BLASCHKE, 2010), possivelmente associando um valor semântico aos segmentos obtidos. No que tange a identificação de Zonas de Manejo, esta abordagem visa auxiliar na identificação das diferenças, de modo a oferecer ao agrônomo suporte para uma análise mais aprofundada sobre os motivos e processos que as ocasionam, e suas implicações na qualidade e produtividade da cultura em análise. Como exemplo, alguns dos valores semânticos aplicados aos talhões podem ser: locais em que certa variedade da cultura apresentou as características desejadas; ou regiões de maior ou menor produtividade; ou ainda, regiões onde a cultura foi mais vulnerável a uma certa praga ou doença.

Para encontrar os segmentos aplicando a técnica da Segmentação Multiresolução Variográfica Ótima, o primeiro passo é a identificação do parâmetro sp que deve ser

Figura 17 – Grupos para o PCA K-means.

Grupos K-means para os dados não-normalizados

(a) Dados não-normalizados

Grupos K-means para os dados normalizados

(b) Dados normalizados

Fonte: Autor.

aplicado ao FSMM. Para tal, a curva para o Primeiro Momento (seção 3.1.5) foi obtida para as camadas interpoladas das amostras de nutrientes e textura do solo (seção 4.1.1). Devido ao fato de se estar avaliando uma região completamente representada pelos dados obtidos, fato que impossibilita a expansão da janela de avaliação para o local, não foi necessário o uso da curva para o Segundo Momento. Neste caso, o valor ótimo é dado pelo ponto em que sp apresenta valor máximo para a curva PM. A curva obtida e o valor ótimo encontrado ($sp_{opt} = 1364$) são apresentados na Fig. 18.

Figura 18 – Curva para o Primeiro Momento obtida para os mapas interpolados. Os dados em verde representam os valores reais obtidos e os dados em azul representam os valores obtidos por meio da aproximação a uma curva spline. O ponto máximo ($sp_{max} = sp_{opt}$) para o PM indica qual o valor de sp deve ser empregado na FSMM para obter os segmentos mais representativos das variações no talhão

Fonte: Autor.

O próximo passo do método é a geração dos segmentos que servirão de ponto inicial para as análises subsequentes, empregando o sp_{opt} na FSMM, Eq. (3.1.9). O resultado é apresentado na Fig. 19.

4.1.4 Resultados

Para a análise dos dados, os segmentos FSMM obtidos foram indicados no gráfico biplot gerado para as duas componentes principais mais significativas obtidas dos dados normalizados, apresentado na seção 4.1.2. Ressaltamos que diferentemente da obtenção dos grupos via *K-means*, em que o subespaço PCA é utilizado pelo classificador, a segmentação FSMM ocorre com os dados brutos pois, seu objetivo é encontrar os segmentos em função da avaliação da heterogeneidade (de valores e forma), portanto, não é um algoritmo classificador de grupos. Neste caso, o gráfico biplot para o PCA permite visualizar como os segmentos se apresentam no subespaço transformado. Além disso, o uso das variáveis

Figura 19 – Segmentos obtidos para o $sp_{opt} = 1364$

Fonte: Autor.

normalizadas antes da aplicação da PCA, permite a comparação e a visualização dos vetores de influência (cargas), ainda que apresentem pequenas variações.

Como as Zonas de Manejo normalmente são utilizadas como ferramentas na busca da identificação dos fatores que influenciam a produtividade, as variáveis i_prod (produtividade de massa seca, em t/ha) e i_alt (altitude, em m) foram inseridas no PCA para fins de observação quanto ao comportamento destes fatores em relação ao conjunto de dados, principalmente as direções de correlação e amplitudes das variabilidades. A inserção destas variáveis não alterou significativamente as componentes principais (Tabela 5) e seus carregamentos (Tabela 6), se comparados aos dados das tabelas 2 e 4.

Tabela 5 – Importância dos componentes (até a 5 componente): dados normalizados, inseridos produtividade e altitude para o talhão analisado

	PC1	PC2	PC3	PC4	PC5
Desvio padrão	2,50	1,74	1,14	1,04	0,96
Proporção da variância	0,45	0,22	0,09	0,08	0,07
Proporção acumulada	0,45	0,66	0,76	0,83	0,90

Fonte: Autor.

Por exemplo, o carregamento da variável i_ph para a PC1 passou de 0,90 para 0,88 (Tabela 6). Nota-se que, apesar do valor para o carregamento da componente PC1 de i_alt ter sido próximo de 0,80, o que indica uma considerável correlação positiva com os dados, o fato de não ter havido grande mudança sugere que a altitude não explicaria sozinha o

Tabela 6 – Carregamentos para os dados normalizados, inseridos produtividade e altitude para o talhão analisado

	PC1	PC2	PC3	PC4	PC5
i_ph	0,88	0,04	-0,14	0,24	0,17
i_hal	-0,67	-0,67	0,01	-0,20	-0,11
i_al	-0,79	-0,50	0,03	-0,18	-0,12
i_ca	0,95	0,03	-0,15	0,01	-0,03
i_mg	0,95	0,03	-0,15	0,01	-0,03
i_k	0,12	0,12	-0,86	-0,05	0,23
i_mo	0,23	-0,93	-0,21	0,06	0,08
i_p	0,44	-0,63	0,22	-0,06	-0,04
i_c	0,23	-0,93	-0,21	0,06	0,08
i_silte	-0,31	0,28	-0,47	-0,61	-0,37
i_areia	-0,86	0,15	-0,09	0,27	0,27
i_argila	0,89	-0,20	0,07	-0,19	-0,21
i_prod	-0,12	-0,02	-0,22	0,63	-0,73
i_alt	-0,79	-0,18	-0,29	0,26	0,14

Fonte: Autor.

arranjo das variáveis, mas possui considerável correlação com os dados do conjunto. O resultado da análise é apresentado no gráfico biplot na Fig. 20.

Analisando os gráfico biplot da Fig. 20, pode-se perceber que:

- os segmentos que possuem características semelhantes, são representados próximos uns dos outros no subspaço PCA;
- a produtividade (i_prod) apresenta um pequeno aumento na direção do segmento 2, que por sua vez apresenta altos valores para a altitude (i_alt) e baixos valores para pH (i_ph), Ca (i_ca) e Mg (i_mg);
- o segmento 3 apresenta valores altos para a altitude (i_alt), Al (i_al) e H+Al (i_hal) e apresenta baixos valores de K (i_k);
- os segmentos 13, 10 e 11 apresentam altos valores para argila (i_argila) e baixos valores de areia (i_areia), sendo que o segmento 11 possui uma concentração maior de P (i_p);
- os segmentos 4 e 7 apresentam baixas concentrações de matéria orgânica (i_mo) e C (i_c). Por outro lado, o segmento 4 apresenta uma maior concentração de silte (i_silte) e o segmento 7 uma maior concentração de K (i_k).

Percebe-se também na Fig. 20, que os segmentos conseguiram separar grupos que de outro modo seriam difíceis de serem diferenciados, como no caso dos segmentos 4 e 7, por exemplo. A comparação final entre os mapas obtidos para o K-means e por meio da FSM é apresentada na Fig. 21.

Figura 20 – Gráfico PCA para os dados do talhão analisado e atribuição dos pontos na grade aos segmentos FSMM obtidos. Os dados de produtividade (*i_prod*) e altura (*i_alt*) foram adicionados a PCA, mas não foram utilizados na etapa de segmentação. As setas em preto representam as “cargas” (*loadings*) das variáveis para cada eixo. No centroide de cada grupo está o identificador do segmento o qual os pontos pertencem. A função *ordispider* do pacote *vegan* para R foi utilizada para obtenção gráfico

Fonte: Autor.

A principal vantagem da Segmentação Multiresolução Variográfica Ótima está no fato de que as regiões segmentadas apresentam o maior tamanho possível para os limites de mínimos de homogeneidade (ou máxima heterogeneidade). A área de cada segmento, gera o que tomamos a liberdade de definir como “segmentos naturais” em que a dispersão máxima para os gradientes espaciais foi alcançada, fator assegurado pelo máximo obtido para a curva PM.

A sobreposição dos segmentos FSMM no subespaço PCA ocorre não somente em decorrência da dimensionalidade dos dados, mas também do fato da segmentação respeitar as restrições de heterogeneidade espacial, além de representar a similaridade entre seus atributos. Aplicando-se o *K-means* nestes dados, as sobreposições parciais ocorreriam se

Figura 21 – Mapas com as regiões delimitadas pelos grupos PCA *K-means* e Segmentos FSMM. Os mapas apresentados nas figuras (a) e (b) apresentam desenhos contendo formas geométricas com aparências não-naturais. Por exemplo, as regiões 1 e 5 em (a) estão separadas quase que retilmente, além do problema de surgimento de pequenas “ilhas”, devido ao resultado do classificador

(a) *K-means*: dados não-normalizados (b) *K-means*: dados normalizados

(c) Segmentos FSMM

Fonte: Autor.

fossem adicionadas outras dimensões para processamento no algoritmo (outras componentes principais), mas sem considerar as restrições espaciais. Apesar do incremento de outras componentes também incrementar a variabilidade explicada, sua inserção intensificaria o efeito causado pela “maldição da dimensionalidade”, isto é, resultaria em um aumento na formação de “pequenas ilhas” no mapa gerado. Estas ilhas dificultam a utilização do mapa resultante e, por consequência, requer o emprego de filtros espaciais para reduzir a fragmentação das classes, para que sejam úteis ao uso agrícola (KHOSLA et al., 2010, p. 212). O fenômeno de formação de pequenas regiões pode ser observado nos mapas gerados para o K-means, apresentados na Fig. 21(a) e Fig. 21(b). Por sua vez, os segmentos FSMM dispensam a aplicação de filtros espaciais para remoção de pequenas regiões, pois as maiores áreas possíveis já foram obtidas.

Comparando-se os mapas gerados, pode-se visualmente verificar que o mapa apresentado na Fig. 21(c) apresenta padrões de segmentação aproximados aos grupos delineados nas Figs. 21(a) e 21(b). Observa-se também que o segmento de número 7 aproxima-se visualmente ao fenômeno destacado na Fig. 14 e que esta mesma região não é evidente para o K-means com os dados normalizados, Fig. 21(b), e manifesta-se parcialmente quando os dados não foram normalizados, Fig. 21(a), ao menos para as dimensões escolhidas. Podemos concluir que a segmentação FSMM apresenta uma aproximação dos padrões espaciais encontrados para os dois casos do K-means, e que também elimina a necessidade de aplicação de filtros espaciais para remover pequenas ilhas no mapa, além de apresentar o número ideal de segmentos para a região analisada, considerando este exemplo.

Acreditamos que o exemplo poderá ser generalizado para outras aplicações, principalmente pelo fato do método basear-se no efeito obtido do uso de combinações genéricas de imagens 2D formadas por Mapas de Moran (PCNM) (seção 3.1.5).

Este exemplo procurou evidenciar que a utilização da Segmentação Multiresolução Variográfica Ótima apresenta uma melhor divisão para a região analisada, além de não necessitar da normalização dos dados, por considerar as quebras de heterogeneidade, fator que não depende da escala.

4.2 Estudo de caso: degradação ambiental para a comunidade de anuros na Mata Atlântica

Esta seção apresenta as principais diferenças entre a amostragem por definição de um *buffer* de tamanho prefixado para avaliação das regiões próximas aos pontos de interesse e o uso de segmentos definidos para representarem as micro regiões bioclimáticas em que cada ponto de coleta está inserido. O termo bioclimático se justifica aqui pelo uso de dados climáticos associados as observações das espécies em análise. O termo *buffer* é empregado em SIG para denominar a região delimitada por uma distância fixa a alguma

geometria. Um *buffer* definido ao redor de um ponto forma um círculo de raio R , onde R é a distância informada. Neste caso, este tipo de *buffer* é utilizado para se obter estimativas sobre a região estudada. Neste exemplo, apresentamos uma comparação entre esta técnica e o uso de segmentos obtidos por meio da Segmentação Multirresolução Variográfica Ótima, como uma alternativa para obter estimativas que sejam representativas do ponto amostrado em termos dos dados disponíveis.

4.2.1 Caracterização do problema

A destruição e a fragmentação de habitats como consequência da expansão urbana e agrícola, são apontados como os principais fatores responsáveis pela perda global de biodiversidade (TILMAN, 2001; RYBICKI; HANSKI, 2013; RAMIADANTSOA et al., 2015). Com a crescente demanda agropecuária muitas áreas próximas a regiões protegidas tem se tornado um mosaico de plantações e pastagens, promovendo uma inevitável perda e fragmentação das regiões de floresta (LIRA et al., 2012). Regiões fragmentadas tendem a não possuírem as mesmas heterogeneidades ambientais se comparadas a cobertura vegetal original. Consequentemente, alguns dos fragmentos ainda que possuam os recursos e as condições físicas necessárias para a colonização e sobrevivência de uma espécie, não provêm os locais apropriados necessários para reprodução das espécies (LIRA et al., 2012).

Vasconcelos et al. (2014) aponta que, em virtude da diversidade de microhabitats presentes na Mata Atlântica, diversas espécies de anuros concentram-se em locais muito específicos, com algumas espécies observadas em apenas uma localidade, portanto, microendêmicas. Além disso, ele também aponta que a topografia pode ser um dos fatores determinantes para a definição de regiões biogeográficas.

Para poder se estimar o impacto antrópico causado nas regiões e espécies, um fator importante é a forma como será feita a amostragem da região ou seleção de áreas de interesse (LIRA et al., 2012). Este fator pode impactar diretamente nos resultados, possivelmente levando a diferentes conclusões sobre o mesmo problema, devido principalmente as diferentes formas de estruturação espacial da amostragem realizada (DÁTTILO; AES; IZZO, 2013).

O objetivo deste exemplo é apresentar como a Segmentação Multiresolução Variográfica Ótima pode ser empregada para avaliar as características próximas das regiões vizinhas ao ponto de coleta ou ocorrência das espécies. Neste caso, especificamente, apresenta como a estimativa de uso do solo na região próxima aos pontos de observação pode ser executada empregando uma métrica biologicamente mais significativa de proximidade, que permita comparar regiões por seus atributos bioclimáticos. Neste caso, utiliza-se os dados ambientais para se obter este significado biologicamente mais favorável a uma observação de espécies que coexistem na mesma região. Esta estimativa pode ser útil na busca de possíveis evidências de impactos nas comunidades das espécies investigadas,

principalmente ao se avaliar os fatores relacionados com as atividades humanas nas regiões onde foram identificadas comunidades de anuros (sapos), previamente selecionadas por um critério de semelhança ambiental.

Neste estudo não utilizamos variáveis que apresentem dados de altitude pois acreditamos que estas informações enviesariam os resultados, principalmente por existir correlação entre temperatura e altitude (HALBRITTER et al., 2013; LEHMAN, 2014).

Não pretende-se aqui apresentar uma análise completa dos fatores e impactos causados, mas apenas verificar a existência de possíveis correlações que ofereçam pistas para um maior aprofundamento futuro.

Para selecionar as regiões próximas aos locais, são apresentadas duas abordagens. Na primeira, a estimativa é feita por meio de um *buffer* de 10 km de raio, calculado em torno do ponto de ocorrência das espécies. Na segunda abordagem, ao invés de uma geometria de tamanho fixo, é gerado um segmento FSMM em que o ponto coletado está inserido. A heurística empregada neste caso foi a adoção do parâmetro espacial ótimo (sp_{opt}) mínimo comum a todos os pontos de coleta. Este fator foi obtido empregando a Segmentação Multiresolução Variográfica Ótima avaliada para cada ponto.

4.2.2 Dados empregados (anuros, áreas de estudo, bioclimáticos)

Dez ambientes pertencentes ao bioma Mata Atlântica tiveram a fauna de anuros compilada da literatura com auxílio de Fabrício Hiroiuki Oda, pesquisador no Laboratório de Ictioparasitologia da Universidade Estadual de Maringá e Sybelle Bellay, pesquisadora do Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura, Universidade Estadual de Maringá. Estes dez ambientes fazem parte de um total de 73 pontos analisados, do qual selecionamos apenas aqueles que estão no bioma Mata Atlântica pela classificação da região pelo Ministério do Meio Ambiente (MMA) (BRASIL. Ministério do Meio Ambiente, 2007). A descrição com as respectivas coordenadas geográficas dos pontos são apresentadas na Tabela 7.

Tabela 7 – Pontos de amostragem e contagem de riqueza de espécies por revisão bibliográfica

Ponto	Local	Riqueza (espécies)	Latitude	Longitude	Fonte
P19	Parque Estadual do Desengano	13	21° 52'S	41° 54'W	Siqueira et al. (2011)
P20	Parque Estadual Carlos Botelho	19	24° 01'S	47° 47'W	Moraes, Sawaya e Barrella (2007)
P21	Reserva Biológica de Duas Bocas	34	20° 16'51'S	40° 28'37'W	Prado e Pombal Jr. (2005)
P27	Estação Ecológica Juréia-Itatins	20	24° 31'S	47° 16'W	Narvaes, Bertoluci e Rodrigues (2009)
P33	RPPN Campo Escoteiro Geraldo Hugo Nunes	40	22° 34'33'S	43° 01'44'W	Silva-Soares et al. (2010)
P42	Condomínio Rio Sagrado	32	25° 29'55'S	48° 50'29'W	Armstrong e Conte (2010)
P43	Morro São João	16	22° 33'S	42° 01'W	Almeida-Gomes et al. (2008)
P57	Reserva Natural Salto Morato	42	25° 09'S	48° 16'W	Garey e Hartmann (2012)
P65	Área de proteção Ambiental de Jundiá	31	23° 11'	46° 56'W	Ribeiro, Egito e Haddad (2005)
P73	Universidade Estadual do Sudoeste da Bahia	15	14° 53'34'S	40° 48'9.2'W	Nery (2014)

Fonte: Autor.

A região, em toda a sua extensão delimitada pelas coordenadas (Long, Lat)

48°56'30.120"O, 25°36'0.000"S e 40°22'30.000"O, 14°47'30.120"S - WGS84, foi selecionada a partir das coordenadas dos pontos de coleta. Esta região apresenta como característica principal ser coberta por uma matriz urbano-agrícola formada por áreas públicas e privadas. Os pontos selecionados estão localizados em áreas remanescentes de Mata Atlântica e apresentam os seguintes climas segundo a classificação de Köppen: *Cfa* - clima subtropical úmido, *Am* - clima tropical de monção, *Aw* - clima tropical de estações úmida e seca (estação seca no inverno) (PEEL; FINLAYSON; MCMAHON, 2007; ALVARES et al., 2013), com níveis pluviométricos variando de 800 mm a 2100 mm (CONEGO, 2005). Estes pontos, identificados na Fig. 22, foram utilizados para verificar as diferenças entre os ambientes e as técnicas de seleção de áreas. Também buscou-se verificar a existência de possíveis relações entre a ocupação do solo as espécies de anuros nas áreas adjacentes a estes ambientes.

Figura 22 – Mapa com a localização dos pontos de observação da espécie. Ressalta-se a localização dos pontos dentro das regiões pertencentes aos bioma Mata Atlântica

Fonte: Composição entre o Mapa de Biomas do Brasil (BRASIL. Ministério do Meio Ambiente, 2007) e os pontos de coleta dos dados produzido pelo próprio autor.

Como variáveis ambientais, os dados do projeto Worldclim/Bioclim (HIJMANS et al., 2005) foram utilizados para se obter os segmentos FSM. Também foram utilizadas na comparação entre os gradientes, sua heterogeneidade e separabilidade dos dados em

relação aos ambientes, comparadas para as abordagens (segmentos FSMM e uso de *buffer* de tamanho fixo). Estes dados são derivados de medições de temperatura e pluviosidade no intuito de gerar variáveis biologicamente mais significativas para as espécies em questão (denominadas variáveis bioclimáticas). Estas variáveis representam tendências anuais (e.g.: temperatura anual média, precipitação anual), sazonalidade (e.g.: variação anual de temperatura e precipitação), fatores ambientais extremos ou limitantes (e.g.: temperatura mínima no mês mais frio, máxima no mês mais quente) (HIJMANS et al., 2005). As 19 variáveis (camadas) bioclimáticas disponibilizadas pelo projeto Worldclim e suas respectivas descrições são apresentadas na Tabela 8.

Tabela 8 – Camadas Bioclim/Worldclim e suas respectivas descrições

Variável	Descrição)
BIO1	Temperatura Média Anual
BIO2	Variação de Temperatura Média Diurna (Média mensal (max. temp. - min. temp.))
BIO3	Isotermalidade (BIO2/BIO7) (* 100)
BIO4	Sazonalidade de Temperatura (desvio padrão * 100)
BIO5	Temperatura Máxima no Mês Mais Quente
BIO6	Temperatura Mínima no Mês Mais Frio
BIO7	Variação Anual de Temperatura (BIO5-BIO6)
BIO8	Temperatura Média no Trimestre Mais Chuvoso
BIO9	Temperatura Média no Trimestre Mais Seco
BIO10	Temperatura Média no Trimestre mais Quente
BIO11	Temperatura Média no Trimestre Mais Frio
BIO12	Precipitação Anual
BIO13	Precipitação no Mês Mais Chuvoso
BIO14	Precipitação no Mês Mais Seco
BIO15	Sazonalidade da Precipitação (Coeficiente de Variação)
BIO16	Precipitação no Trimestre Mais Chuvoso
BIO17	Precipitação no Trimestre Mais Seco
BIO18	Precipitação no Trimestre Mais Quente
BIO19	Precipitação no Trimestre Mais Frio

Fonte: Autor.

4.2.3 Abordagens

Para verificar as diferenças entre a seleção por *buffer* e segmentos FSMM, executamos uma Análise de Componentes Principais (PCA) nos dados das variáveis, Tabela 8, selecionadas por meio da sobreposição espacial entre as camadas (BIO1 à BIO19) e as geometrias em análise.

Buffer

Para a estimativa de uso do solo no entorno de uma região de interesse foi definido um *buffer* de 10 km de raio ao redor de cada ponto de observação. Nesta abordagem assume-se que as regiões no entorno são relativamente similares ao local do ponto de coleta/observação. Por outro lado, se o *buffer* estiver próximo de regiões montanhosas, escarpadas ou costeiras, toda a região de seu polígono permanece sendo contabilizada, ignorando qualquer diferença topográfica, fator que pode possivelmente prejudicar as estimativas obtidas por este método.

Aplicando a Segmentação Multiresolução Variográfica Ótima

Como alternativa para a primeira abordagem propomos o uso de segmentos FSMM que correspondem a regiões bioclimáticas com maior similaridade (maior homogeneidade) em relação as áreas de registro de ocorrência das espécies. No que se refere ao uso de dados de ocorrência e observações de espécies, este limite define a extensão dos segmentos obtidos. Seu propósito é obter uma amostra de parte do nicho fundamental sobreposto das espécies em análise na vizinhança do ponto de coleta. O nicho fundamental pode ser caracterizado por como sendo o nicho definido pelo intervalo de condições e recursos no qual os indivíduos de uma espécie podem persistir (HUTCHINSON, 1957). Formalmente, Hutchinson (1957) define como um espaço abstrato n-dimensional, formado por variáveis físicas e biológicas, correspondendo a um estado no ambiente que permite a uma espécie existir por tempo indeterminado.

Adotamos aqui como uma aproximação das condições do nicho fundamental o uso de variáveis ambientais abióticas. Estas variáveis são empregadas para se obter informações que representem as condições climáticas pelas quais os indivíduos de uma espécie possam viver e se reproduzir, excluído-se as limitações resultantes de interações bióticas.

Os segmentos que representam a vizinhança dos ambientes foram obtidos por meio da avaliação das 19 camadas contendo as variáveis bioclimáticas, disponibilizadas pelo projeto Worldclim (HIJMANS et al., 2005). Estas camadas são normalmente utilizadas para modelagem de nicho ecológico fundamental (HIJMANS et al., 2005).

Para os 10 ambientes avaliados foram obtidos os valores ótimos para o fator espacial (sp_{opt}) e o resultado é apresentado na Tabela 9. A tabela também apresenta o raio a partir do ponto de coleta que define o tamanho da janela de avaliação, dimensionada pelo algoritmo apresentado na seção 3.1.6.

Após a execução do algoritmo, o menor valor obtido para sp_{opt} foi utilizado para segmentação da região calculada para o retângulo envolvente mínimo entre todos os pontos. Segmentos gerados com valores menores para o parâmetro sp estarão contidos nos segmentos gerados por valores maiores. Assim a segmentação pelo menor sp comum a

Tabela 9 – Fator espacial ótimo (sp_{opt}) e respectivo raio do buffer no qual é gerada a janela de avaliação para cada região

Ambiente	Raio (m)	sp_{opt}
P19	12000	40
P20	138000	1200
P21	74000	600
P27	14000	50
P33	74000	150
P42	138000	500
P43	11000	40
P57	42000	920
P65	74000	850
P73	18000	120

Fonte: Autor.

todos os pontos terá a menor área com as mesmas delimitações de homogeneidade para os segmentos gerados. Deste modo homogeneizamos a amostragem não mais pela definição de um raio mínimo para um *buffer* específico, mas ao invés disso, pelo menor parâmetro de segmentação espacial, neste caso, utilizado como menor fator comum de homogeneização espacial.

A Fig. 23 apresenta a diferença entre o uso do *buffer* e de um segmento delimitado pela heterogeneidade entre as camadas do projeto Bioclim. Nela estão representados os principais acidentes geográficos na região do ponto P27. Na imagem pode-se observar que o segmento FSMM seleciona uma região mais homogênea, se comparado com a área do *buffer*. Esta característica do segmento possibilita a identificação de regiões que apresentem espécies micro-endêmicas, uma das características das espécies presentes no bioma (VASCONCELOS et al., 2014).

A descrição completa do método utilizado para a obtenção dos segmentos pode ser verificada na sessão 3.1.

Comparação da heterogeneidade dos ambientes via PCA

Uma Análise de Componentes Principais (Principal Component Analysis - PCA) (ABDI; WILLIAMS, 2010) foi realizada usando os dados obtidos nos pontos pertencentes a cada *buffer* e a cada segmento. Com base nos dois eixos da ordenação (PCA1 e PCA2) foi verificado a separação entre os valores para as camadas pertencentes a cada geometria associada a um dos pontos de observação de ocorrência das espécies. O gráfico biplot para os dados do PCA dos *buffers* são apresentados na Fig. 24 e para os segmentos são

Figura 23 – Acidentes geográficos na região próxima ao ponto P27 segundo Sayre et al. (2014). O mapa apresenta o polígono gerado para o *buffer* calculado tendo como centro o ponto de coleta. Também apresenta o segmento obtido pela segmentação utilizando o menor fator espacial comum a todos os 10 pontos analisados ($sp = 40$). Foram mantidas as marcações originais da classificação dos dados no *dataset* com os termos em inglês, com a respectiva tradução entre parênteses

apresentados na Fig. 25.

Pode-se observar que no gráfico apresentado na Fig. 24, os ambientes apresentam como maior fator de separabilidade a latitude. Encontram-se aninhados basicamente em 4 grupos: 1) formado pelos pontos P27, 42 e 57; 2) formado pelos pontos 20, 33; 3) 19, 21, 43, 65; 4) 73. Por outro lado, para o gráfico da Fig. 25 é possível claramente se distinguir univocamente cada ambiente representado pelos segmentos e apontar quais as variáveis apresentam as maiores contribuições para o segmento. Ex.: P33 apresenta valores moderados para BIO6 (Temperatura Mínima no Mês Mais Frio); P73 apresenta valores significativos para a variável BIO3 (Isotermalidade).

4.2.4 Estimativa de uso do solo

Para estimar o uso do solo, empregamos o *dataset* publicado por Sayre et al. (2014). Este *dataset* possui dados sobre bioclimas, acidentes geográficos, litologia e cobertura vegetal de todos os continentes, organizados e inseridos por diversas instituições e pesqui-

Figura 24 – PCA executado nos dados selecionados para cada buffer obtido, representando a vizinhança de cada um dos 10 pontos de coleta no bioma Mata Atlântica

Fonte: Autor.

sadores no mundo. Deste *dataset* utilizamos os dados sobre cobertura vegetal, compostos por 23 classes que representam o uso ou ocupação do solo, com resolução espacial de 300 m para cada *pixel* da imagem (*raster*).

Índice de estimativa de uso do solo

Propomos a identificação e quantificação, em termos de ocupação e uso do solo, para regiões que deveriam estar preservadas, por meio de análise da sobreposição espacial entre as regiões identificadas no *dataset* como pasto, plantação e área urbana, e as regiões dos *buffers* ou segmentos FSMM gerados pela Segmentação Multiresolução Variográfica Ótima. Este exemplo busca evidenciar as diferenças entre a seleção por um critério simples e o método proposto, principalmente se forem consideradas as características do ambiente necessário para a sobrevivência das espécies analisadas.

Figura 25 – PCA executado nos dados selecionados para cada segmento obtido representando a vizinhança bioclimática de cada um dos 10 pontos de coleta no bioma Mata Atlântica

Fonte: Autor.

Considerando o bioma Mata Atlântica como uma região que idealmente formada apenas por florestas, a estimativa de uso do solo foi realizada por meio de consulta espacial que retornou a área total de sobreposição entre cada *buffer* ou segmento FSMM com as regiões identificadas no *dataset* como:

- “*Mostly Cropland* (Majoritariamente talhões agrícolas)” - conjunto representado pela letra *C*,
- “*Artificial or Urban Area* (Área Urbana ou Artificial)” - conjunto representado por *U*,
- “*Bare Area* (descampado)” - conjunto representado por *B*,

- “*Grassland, Shrub, or Scrub* (pastagem, arbustos e vegetação rasteira)” - conjunto representado por G .

Assim, a área interna de cada *buffer* ou segmento FSMM, que representa a vizinhança do ambiente, foi utilizada para estimar a quantidade de área (em km²) que não contém mais a flora nativa, identificada como locais de plantação, pastagem, descampado e área urbana.

Tomando-se P como o polígono que representa a região vizinha a cada ambiente i de onde foram obtidas as observações dos anuros, definimos o Índice de Estimativa de Uso do Solo I_{P_i} (Eq. (4.2)) como:

$$I_{P_i} = \frac{A_{B_i} + A_{C_i} + A_{G_i} + A_{U_i}}{A_{P_i}}, \quad i = 1, 2, \dots, n \quad (4.2)$$

em que:

P_i é o polígono representando o *buffer* ou segmento,

A_{P_i} é a área do polígono representando o ambiente i ,

A_{B_i} é a área resultante da intersecção entre P_i e B :

$$A_{B_i} = \text{Area}(P_i \cap B), \quad (4.3)$$

A_{C_i} é a área resultante da intersecção entre P_i e C :

$$A_{C_i} = \text{Area}(P_i \cap C), \quad (4.4)$$

A_{U_i} é a área resultante da intersecção entre P_i e U :

$$A_{U_i} = \text{Area}(P_i \cap U), \quad (4.5)$$

A_{G_i} é a área resultante da intersecção entre P_i e G :

$$A_{G_i} = \text{Area}(P_i \cap G), \quad (4.6)$$

Finalmente, a área total de ocupação em cada *buffer*/segmento i é dada por:

$$A_{P_i}^{\text{oc}} = A_{B_i} + A_{C_i} + A_{G_i} + A_{U_i} \quad (4.7)$$

4.2.5 Resultados análise de degradação aplicando os segmentos

Para avaliar um cenário de possível degradação da região e a existência ou não de correlação entre a área total dos segmentos e os índices de ocupação para os 10 ambientes na Mata Atlântica, executou-se uma regressão linear simples entre o índice de ocupação em cada segmento (I_{P_i}) e sua respectiva área total (A_{P_i}) e também entre $A_{P_i}^{oc}$ e A_{P_i} .

A relação observada para os dados avaliados indicou a existência de correlação entre a área do segmento A_{P_i} e a área ocupada $A_{P_i}^{oc}$ (ambas em km^2). Esta apresentou para o coeficiente de correlação de Spearman o valor $rs = 0.8545$, o que indica que as variáveis tendem a aumentar ou diminuir juntas com $P = 0.001$ (a correlação é válida e não é fruto de uma possível aleatoriedade dos dados).

Ao comparar a área total do polígono (km^2) e porcentagem da área de ocupação do segmento, a correlação manteve-se válida, apresentando o valor para o coeficiente de Spearman $rs = 0.6364$ (valores aumentando ou diminuindo juntos) com $P = 0.0478$ o que permite rejeitar a hipótese de que a correlação é fruto da aleatoriedade dos dados. Os valores obtidos corroboram a hipótese de que, quanto maior o segmento, maior será a ocupação medida, evidenciando um cenário de possível degradação em todas as regiões analisadas, pois, quanto maior o segmento, maior a quantidade de área ocupada por pasto, plantação, descampado ou área urbana. No caso do uso dos *buffers*, como a área é constante, esta correlação é inexistente.

Pode-se alegar que quanto maior a área do *segmento*, naturalmente maior será a ocupação. Contudo, estas as regiões deveriam estar preservadas. Além disso, os segmentos indicam a extensão das áreas por sua semelhança, o que pode ser importante, principalmente para as espécies endêmicas.

No exemplo, estamos assumindo que a classificação de uso do solo não apresenta erros do Tipo I e II para o *dataset* empregado.

4.3 Estudo de caso: identificação de região bioclimática para a Unidade de Conservação da Serra da Canastra

Estudos apontam que, tanto para a conservação, quanto para a necessidade de deslocamento ou reintrodução de espécies ameaçadas, um dos fatores primordiais é a similaridade entre os habitats originais em termos de recursos disponíveis e condições ambientais (ROE et al., 2010). Neste contexto, apresentamos neste exemplo um caso em que a aplicação da Segmentação Multiresolução Variográfica Ótima pode ser aplicada para definição da extensão de uma região protegida por meio de dados ambientais.

4.3.1 Dados e área de estudo

Segundo [Maia Neto et al. \(2007\)](#), o Parque Nacional da Serra da Canastra, localizado na região sudoeste do Estado de Minas Gerais, é uma das Unidades de Conservação (UC) que fazem parte do Sistema Nacional de Conservação da Natureza ([JEFFERY; FIRESTONE; BUBNA-LITIC, 2008](#)). A UC, de aproximadamente 200.000 ha, foi criada em 1972 e abrange os municípios de São Roque de Minas, Sacramento, Delfinópolis, São João Baptista do Glória, Capitólio e Vargem Bonita. A região é caracterizada por um clima Tropical, com duas estações bem definidas. O inverno (de abril a outubro) é seco, e o verão (de novembro a março) é chuvoso, com curtos períodos de seca. Classificação climática de Köppen Cwb ([PEEL; FINLAYSON; MCMAHON, 2007](#); [ALVARES et al., 2013](#)). Os totais pluviométricos anuais ficam entre 1.200 mm e 1.800 mm. A temperatura média anual fica em torno de 22° C e 23° C, com mínimas próximas de zero e máximas de 40° C. No que se refere a hidrologia, o parque insere-se entre a Bacia do Rio São Francisco e a Bacia do Rio Paraná (Rio Grande e Rio Paranaíba), sendo que a nascente do Rio São Francisco é uma das muitas nascentes e cursos d'água presentes na UC ([MAIA NETO et al., 2007](#)). As altitudes variam de 700 m a 800 m nos vales e acima de 1.000 m no topo dos chapadões (Chapadão da Canastra e Chapadão da Babilônia). Apenas 1/3 dos 200.000 ha (cerca de 71.525 ha) foram regularizados na área do Chapadão da Canastra, e nela situa-se a área oficial do parque. As outras áreas (130.000 ha) apesar de reconhecidas pelo decreto de criação (decreto 70.355 de 03 de abril de 1972) permanecem em situação pendente de regularização fundiária, gerando até hoje conflitos com a unidade ([MAIA NETO et al., 2007](#)). No plano operativo para prevenções a incêndios no parque, [Maia Neto et al. \(2007\)](#) relata que estes conflitos geralmente são manifestados por incêndios criminosos na área regularizada, o uso irregular da região protegida como pastagem para o gado e também a existência de atividades clandestinas de mineração. Atualmente, região do parque está classificada em três grandes regiões: zona de amortecimento, a área regularizada (no Chapadão da Canastra) e área não regularizada (Chapadão da Babilônia) como pode ser observado na Fig. 26.

Variáveis ambientais

Assim como no exemplo anterior, foram utilizados os dados do projeto Worldclim/Bioclim ([HIJMANS et al., 2005](#)) para se obter os segmentos FSMM. No processo para a avaliação do tamanho ideal da janela foram empregadas as 19 variáveis bioclimáticas, apresentadas na Tabela 8.

4.3.2 Avaliações das curvas PM e SM

Para encontrar o parâmetro espacial ótimo (sp_{opt}) empregado em $\mathbb{B}$, utilizou-se como janela inicial de avaliação o retângulo envolvente mínimo calculado para a região

Figura 26 – Carta Imagem do Parque da Serra da Canastra. Imagem CBERS de setembro de 2006

Fonte: Plano operativo de combate a incêndios florestais do Parque Nacional da Serra da Canastra (MAIA NETO et al., 2007).

legal da UC. Em cada avaliação, as curvas PM e SM são calculadas, buscando-se verificar a ocorrência das condições apresentadas na Def. 3.1.12. Enquanto esta condição de parada não ocorrer, a janela é expandida por um fator $d * 2$, em que o fator d é representado em unidades métricas para o mapa. A cada nova expansão d será atualizado para o valor expandido do *buffer*. O valor inicial empregado para a expansão do *buffer* foi 1000 m ($d = 1000$). Ao final da execução, o *sp* encontrado foi de 220, para uma janela expandida em 1000 m. A Fig. 27 apresenta 16 das 19 as variáveis avaliadas para a obtenção dos segmentos.

Os resultados das avaliações para as curvas PM e SM são apresentados nas figuras Fig. 28 e Fig. 29, respectivamente.

4.3.3 Obtenção dos segmentos

Os segmentos gerados para a janela de avaliação das curvas PM e SM são apresentados na Fig. 30.

Estes segmentos são usados apenas como referência na obtenção do valor de sp_{opt} . Entretanto, para uma visualização completa dos segmentos para a região, é necessário utilizar uma área significativamente maior em relação a utilizada na avaliação, pois deseja-

se obter a extensão completa dos segmentos, que de outro modo seriam limitados pelo espaço avaliado.

Figura 27 – As primeiras 16 das 19 variáveis avaliadas para a obtenção dos segmentos na Serra da Canastra

Fonte: Recorte de [Hijmans et al. \(2005\)](#).

Figura 28 – Avaliação da Curva PM para a Serra da Canastra

Fonte: Autor.

Figura 29 – Avaliação da Curva SM para a Serra da Canastra

Figura 30 – Segmentos FSMM (segments.mask $\Rightarrow S = \{s_1, \dots, s_{10}\}$) e valores esperados ($E = \{e_1 = \text{blo1}, \dots, e_{19} = \text{blo19}\}$) para as primeiras 15 das 19 variáveis avaliadas

Assim, para obter os segmentos finais para a UC, foi utilizado como referência o retângulo envolvente mínimo da junção dos limites do estado de São Paulo e Minas Gerais, delimitado pelas coordenadas (Long, Lat) 53°6'29.520"O, 25°18'39.240"S e 39°51'23.040"O, 14°13'58.440"S - WGS84. O resultado final é apresentado na Fig. 31. Nele estão indicados os segmentos que interceptam a região da UC. Também estão destacadas as regiões definidas pelo MMA como áreas prioritárias para a conservação, utilização sustentável e repartição dos benefícios da biodiversidade (hachuras brancas) (BETTIOL, 2007).

Figura 31 – Imagem de satélite da região da Serra da Canastra sobreposta com os segmentos obtidos e indicação de áreas de prioridade de ação pelo MMA

O objetivo das áreas demarcadas pelo MMA é a disponibilização de maiores informações sobre a biodiversidade dos biomas que as compõem, dentre elas o entorno do Parque Nacional da Serra da Canastra. Esta informação foi inserida no mapa com o objetivo visualizar o que o MMA considera que deve receber atenção e relacionar estas regiões com os resultados obtidos para os segmentos e a sua respectiva extensão de influencia bioclimática. Pode-se perceber, por exemplo, a importância dos rios para o clima, evidenciado pela inclusão de boa parte dos rios no entorno da UC nos segmentos selecionados, Fig. 31. Também é possível de se observar que, apesar de não terem sido utilizados dados de altitude, indiretamente a influência do relevo causa pequenas diferenças

nos micro-climas da região, impactando nos gradientes espaciais que são delineados pelos segmentos.

4.3.4 Análise dos dados via PCA

Assim como no exemplo anterior, os dados foram normalizados e a primeira e segunda componente principal foram selecionadas por serem apontadas como suficientes para se avaliar os dados pelo teste de broken-stick, cujo resultado é apresentado na Fig. 32. As componentes selecionadas representam 62,54 % e 20.47 % da variância total para os dados, respectivamente.

Figura 32 – Gráfico screeplot para análise de broken-stick para os dados dos segmentos

Fonte: Autor.

O gráfico biplot para os dados do PCA é apresentado na Fig. 33 e os respectivos segmento são identificados no mapa apresentado na Fig. 34.

Analisando os gradientes no gráfico biplot, Fig. 33. Pode-se observar, algumas relações entre as variáveis e segmentos, por exemplo:

- O segmento 967 (majoritariamente Chapadão da Canastra) apresenta valores altos para Precipitação Anual (bio12) e Precipitação no Trimestre Mais Chuvoso (bio16) e apresenta valores baixos para Temperatura Média Anual (bio1), Temperatura Média no Trimestre Mais Chuvoso (bio8), Temperatura Média no Trimestre mais Quente (bio10), Sazonalidade da Precipitação (bio15);
- O segmento 1010 (majoritariamente Chapadão da Babilônia) apresenta maiores valores para Precipitação no Mês Mais Chuvoso (bio13) e Precipitação no Trimestre

Mais Frio (bio19) e baixos valores para a Temperatura Média no Trimestre Mais Seco (bio9) e para Temperatura Média no Trimestre Mais Frio (bio11);

- O segmento 1018 apresenta valores altos para Precipitação no Mês Mais Seco (bio14) e Precipitação no Trimestre Mais Seco (bio17) e valores baixos para Temperatura Mínima no Mês Mais Frio (bio6).

Figura 33 – Gráfico biplot PCA para os dados dos segmentos FSMM para a região da Serra da Canastra

Fonte: Autor.

Assim, observando os resultados obtidos, é possível afirmar que no Chapadão da Canastra as chuvas são intensas e com baixa sazonalidade, além disso, a região apresenta temperaturas mais amenas, em relação as regiões no entorno. O Chapadão da Babilônia, por sua vez, apresenta uma maior precipitação nos meses mais frios, com baixas temperaturas nestes meses. Outras análises e técnicas podem ser aplicadas aos dados dos segmentos para se obter novas relações.

Figura 34 – Mapa com segmentos obtidos pela Segmentação Multiresolução Variográfica Ótima para a região da Serra da Canastra. Cada segmento identifica uma região com diferentes bioclimas, de acordo com as variáveis biocimáticas do Worldclim

Fonte: Autor.

5 Considerações e Conclusões

Neste capítulo apresentamos inicialmente algumas considerações sobre a pesquisa realizada (seção 5.1), seguido pelas conclusões (seção 5.2), avaliação dos resultados (seção 5.3) e finalmente as sugestões de trabalhos futuros (seção 5.4).

5.1 Considerações

Neste trabalho foi apresentado um método de segmentação ótimo baseado no processamento de múltiplas camadas/atributos em dados georreferenciados, interpolados ou espectrais, que pode ser aplicado em problemas de delimitação de regiões geográficas. Acreditamos que nossas hipóteses foram satisfeitas quando procuramos responder: qual algoritmo segmentador poderia ser utilizado como base para a pesquisa? Como quantificar o que é ótimo e quais aspectos são relevantes?

Respondidas estas questões, o trabalho procurou apresentar alguns exemplos de aplicabilidade do método proposto e, quando possível, uma comparação com outra técnica. Entretanto, o objetivo não foi efetivamente a resolução dos estudos de caso, mas exemplificar que nestes problemas o valor do parâmetro ótimo do segmentador pode ser obtido pelo valor máximo da curva PM. Assim, os segmentos obtidos podem ser utilizados para responder as questões relevantes de cada problema, motivação para que a hipótese fosse levantada.

Para definir os aspectos construtivos e operacionais para o método proposto - apresentados no capítulo 3 - esta pesquisa procurou identificar e selecionar qual algoritmo de segmentação existente seria o mais adequado ao uso. A técnica deveria oferecer suporte ao desenvolvimento da pesquisa e permitir o processamento de múltiplas camadas de dados sobrepostos sobre uma mesma área (multidimensional).

Pesquisamos alguns algoritmos de segmentação e identificamos uma abordagem que poderia auxiliar no processamento de dados com múltiplas camadas de informação: a técnica denominada *Geo Object Based Image Analysis* (Geo-OBIA), apresentada na seção 3.1. Esta técnica, empregada em imagens de Sensoriamento Remoto, é utilizada para processamento de imagens multi e hiperespectrais e faz uso de um passo inicial de pré-processamento dos dados, por meio da aplicação de um algoritmo segmentador para múltiplas bandas espectrais da imagem, com posterior processamento e atribuição de valores semânticos. Um dos algoritmos amplamente empregados na área é o *Multiresolution* e, devido à vasta gama de opções de ferramentas que implementavam este algoritmo, contando inclusive com uma implementação com código aberto disponibilizada pelo projeto TerraLib, ele foi o escolhido. Implementamos uma versão com pequenas alterações como um pacote para o

software R com o objetivo oferecer suporte aos testes. Nesta versão do algoritmo, além da atribuição das células ao seu respectivo segmento, também é retornada a média dos valores de cada segmento em sua respectiva camada. Deste modo, pode-se avaliar o impacto da segmentação no mapa para cada modificação dos parâmetros inseridos no segmentador. A versão em R do algoritmo facilitou o processamento de diferentes tipos de dados, desde que disposto em uma tabela, podendo inclusive, possuir valores nulos (comuns em dados *raster* de mapas reais). Nesta etapa da pesquisa, realizamos os primeiros testes com dados agrícolas, e o resultados foram publicados em [Costa et al. \(2013\)](#).

Selecionado o algoritmo segmentador e efetuada as devidas adaptações, buscou-se responder como definir o conceito de segmentação ótima dentro do escopo do trabalho para os dados disponíveis. Neste caso, foi necessário quantificar os fenômenos espaciais por meio da geoestatística e seus métodos de estimativa da autocorrelação espacial. Também recorremos aos métodos aplicados em estudos de biogeografia.

Em biogeografia, uma das técnicas amplamente empregadas no estudo de autocorrelação espacial é a utilização do Índice de Moran, que permite avaliar a distribuição dos valores em um mapa. Também vimos que [Borcard et al. \(2004\)](#) desenvolveu uma técnica baseada neste índice, denominada Mapas de Moran ou PCNM. Esta técnica utiliza mapas simulados com diferentes formas de autocorrelação espacial em diferentes escalas para extrair os principais descritores de um fenômeno espacial. Empregamos estes mapas como simulações de fenômenos espaciais 2D genéricos para avaliar o resultado do segmentador quando o parâmetro espacial é incrementado. Este parâmetro determina o tamanho dos segmentos para a avaliação da função de heterogeneidade adotada (de valor e espacial). Uma vez definidos os dados genéricos para representar fenômenos espaciais genéricos, ainda era necessária a definição de um método para mensurar os impactos da segmentação em cada camada/atributo, em relação a variabilidade e a estrutura espacial. Nesta etapa da pesquisa, buscamos identificar qual função poderia ser maximizada ou minimizada para caracterizar o valor ótimo para o parâmetro espacial (*sp*) a ser aplicado no segmentador.

Na etapa seguinte, aplicamos a análise variográfica para o teste de impacto do segmentador e respectivas modificações na distribuição espacial e avaliação da variabilidade em cada camada. Pode-se então medir a extensão da autocorrelação espacial (alcance), a variância mínima e a dispersão dos valores (efeito pepita), e a variância máxima (patamar). Entretanto, o modelo empregado para a aproximação do variograma deveria ser o mais estável possível (modelo de transição). Levando em consideração estes fatores, adotamos o modelo Gaussiano como função a ser aproximada pelo método dos mínimos quadrados generalizados (GLS), como descrito na seção 3.1.4. Então estabelecemos os critérios necessários a definição da segmentação ótima, por meio do teste com os dados artificiais gerados pelas funções PCNM - Mapas de Moran.

5.2 Conclusões

Por meio do uso dos Mapas de Moran sobrepostos como camadas/bandas de uma mesma imagem, verificou-se que o efeito pepita aplicado ao modelo Gaussiano, apresentava um valor máximo quando o limite de autocorrelação para os fenômenos sobrepostos era atingido nos mapas simulados, corroborando a hipótese de que é possível a definição de uma função indicadora para a segmentação ótima. Assim, o resultado final da segmentação, quando utilizado o valor de sp para este valor máximo, apresentou o melhor delineamento visual para as camadas avaliadas, considerando o fenômeno/mapa observado. O passo natural foi testar a aplicação para todas as variáveis geradas em uma grade de 20 x 20 células, inseridas sequencialmente, no intuito de observar a ocorrência do fenômeno, que se confirmou. Observamos que, havendo correlação espacial, haveria um valor máximo para a média do efeito pepita em todas as camadas processadas, sendo este efeito denominado Primeiro Momento (PM) - $\bar{\gamma}_{\text{pepita}}(sp)$, obtido para a curva Spline calculada para os dados brutos.

Quando utilizados dados reais, percebeu-se que nem sempre na janela inicial de avaliação havia um fenômeno completamente caracterizável. Em algumas situações a região selecionada somente apresentava parte da autocorrelação ou gradiente. Desta constatação, verificou-se a necessidade de expansão da janela de teste no mapa, caso fosse possível. Nos testes subsequentes, verificou-se que a correlação parcial no mapa poderia ser detectada pelo fato que a curva PM apresentaria um máximo, mas sem a presença de uma queda significativa nos valores obtidos que pudesse caracterizar como máximo global para a região avaliada. Assim, foi necessário a definição de um Segundo Momento (SM) - $\Sigma \bar{\gamma}_{\text{pepita}}(sp)$, que representasse a inércia para a curva PM. Deste modo, quando o valor de SM apresenta um máximo global, garante-se que na curva PM existe um máximo global, válido para a janela de avaliação. Utilizamos este fenômeno para expandir a região avaliada até que este máximo para a curva SM seja detectado. Assim, foi definido o processo para obtenção do parâmetro espacial ótimo (sp_{opt}) aplicado a FSMM ($\mathbb{B}$). Com a definição das curvas PM, SM e as condições para a sua obtenção, chegamos a definição do parâmetro ótimo aplicado ao segmentador, que nos permitiu denominar a técnica de **Segmentação Multirresolução Variográfica Ótima**.

No Capítulo 3, apresentamos como dados simulados de fenômenos espaciais podem ser aplicados na avaliação de uma técnica de segmentação multivariada (múltiplas camadas, bandas espectrais, atributos ou propriedades numéricas georreferenciadas, desde que dispostas em grade ou mapa de bits) por uso de avaliação variográfica. Por meio do comportamento mensurado e quantificado do segmentador aos efeitos nos variogramas em cada camada, definimos o conceito Segmentação Multirresolução Variográfica Ótima como a identificação do parâmetro espacial ótimo sp_{opt} aplicado ao segmentador $\mathbb{B}$ na região em análise. Este parâmetro, determina o tamanho ideal dos segmentos. O resultado é a divisão

da região em segmentos que apresentam um equilíbrio entre a maior dispersão (pois o efeito pepita é maior quando a dispersão também é maior) pela maior diferença entre as variâncias para as médias nos segmentos (objetivo do segmentador). Assim, $\mathbb{B}$ retorna os segmentos que apresentam máxima variabilidade entre eles para a região analisada, devido ao somatório do efeito pepita medido para a aproximação à curva Gaussiana ao resultado do segmentador em cada camada.

5.3 Avaliação dos resultados e desafios do estudo

O método proposto é uma heurística para identificar o melhor fator de segmentação para as regiões de interesse quando aplicado o algoritmo *Multiresolution*, portanto, não apresentamos uma prova formal de seu funcionamento em absolutamente todas as situações. Neste sentido, apresentamos alguns exemplos nos quais o método se mostrou válido. Intuitivamente, quando o segmento passa a agregar regiões muito diferentes, a variabilidade diminui e a dispersão das médias dos segmentos também. Este fato faz com que a curva PM passe a apresentar valores menores, na medida que o valor de sp é incrementado e aplicado em $\mathbb{B}$.

Caso o valor inicial para PM seja superior ao valor máximo da janela de avaliação, o ponto para o qual o gráfico apresenta o ponto no eixo sp do primeiro mínimo para a curva PM pode ser utilizado como fator espacial inicial para a área atual.

O cálculo do variograma para cada camada é computacionalmente custoso ($\mathcal{O}(n^2)$). Dependendo do tamanho da área avaliada, a identificação do parâmetro sp pode demorar para convergir. O emprego de técnicas de processamento paralelo pode ser de grande ajuda neste sentido, tanto para o segmentador, quanto para o cálculo do variograma.

Para exemplificar a aplicabilidade do método proposto, no Capítulo 4 foram apresentados 3 casos: a definição de Zonas de Manejo Agrícola, a seleção de regiões para amostragem em estudos de biodiversidade aplicados ao problema de avaliação da degradação ambiental de anuros na Mata Atlântica, e a identificação das regiões bioclimáticas que compõem uma Unidade de Conservação da biodiversidade aplicada a UC da Serra da Canastra, em Minas Gerais.

Para o problema de definição de Zonas de Manejo (seção 4.1), o método apresentou o comportamento previsto, ou seja, um valor máximo para a avaliação da curva PM que identifica o valor do parâmetro espacial ótimo (sp_{opt}) que deve ser utilizado no segmentador, como pode ser verificado na seção 4.1.3. Nesta seção também foram comparados os resultados dos mapas obtidos com a aplicação do algoritmo *k-means*. O mapa resultante da aplicação do método proposto apresenta como vantagem não apresentar pequenas “ilhas” na definição dos segmentos, além de delinear fenômenos espaciais que, de outro modo, poderiam ser mascarados pelo critério de escolha do número de agrupamentos para o

k-means.

No problema de avaliação de degradação ambiental para anuros na Mata Atlântica (seção 4.2), mostramos que o método permite a seleção de regiões que representam uma parte do Nicho Fundamental sobreposto para as espécies analisadas. Apresentamos, por meio dos gráficos para os dados analisados via PCA, que os fatores predominantes em cada região permitem separar os ambientes que, apesar de pertencerem ao mesmo bioma, apresentam características distintas. A separabilidade dos ambientes são fatores que podem auxiliar na identificação das características principais para cada região, o que pode ser relevante principalmente se consideradas as espécies endêmicas que vivem na Mata Atlântica. Assim, as regiões selecionadas podem ser consideradas como bioclimaticamente importantes para as espécies, de modo que as estimativas obtidas para os segmentos selecionados podem ajudar a evidenciar os fatores de maior impacto para as espécies em questão. Ressaltamos que, neste exemplo, os valores ótimos para cada ponto foram obtidos por motivos de verificação da validade do método. Como no exemplo buscava-se o valor mínimo comum para *sp*, qualquer ponto que apresentasse valor superior ao mínimo previamente registrado poderia ter sua avaliação interrompida. Neste caso o valor para todos seria calculado enquanto não ultrapassasse o mínimo registrado, valendo para o conjunto de pontos o menor valor de *sp* obtido. Esta pode ser uma das estratégias adotadas para se diminuir o tempo de processamento.

Para o último estudo de caso, mostramos como identificar a extensão das regiões bioclimáticas para a UC do Parque Nacional da Serra da Canastra (seção 4.3). O exemplo mostrou como os segmentos podem auxiliar na tomada de decisão na seleção de regiões dos quais a preservação ou degradação podem impactar diretamente na biodiversidade local. Procuramos mostrar que nem sempre os critérios de divisão política estabelecidos, a seleção de regiões para proteção ou de prioridade de ação, consideram o bioclima, que pode ser um aspecto importante para as espécies que vivem nestas regiões. Pode-se perceber, sem nenhuma análise por especialista, que os rios exercem uma grande influência no clima local, pois diversos segmentos os englobam. Também é possível notar que as regiões com perímetro urbano são englobadas em segmentos, o que evidencia o impacto das cidades no clima local.

Ainda sobre o teste na UC, é importante salientar que **não** foram utilizados dados de altitude, o que poderia favorecer o contorno de rios e outros acidentes geográficos naturais. No exemplo, foram selecionados apenas os segmentos que apresentavam alguma sobreposição com a área legal da UC. Caso tivéssemos considerado a região de amortecimento, mais segmentos teriam sido selecionados. Também é importante destacar que o método não procura substituir os critérios utilizados para delimitar uma UC, mas servir de ferramenta que possa complementar a análise ou auxiliar no estabelecimento de políticas de ações afirmativas, fornecendo a indicação de áreas prioritárias para o estabelecimento

de programas de valorização dos serviços ambientais (HUSSAIN; MILLER, 2014) para a preservação da biodiversidade e recursos naturais locais integrados à comunidade local.

5.4 Implicações e trabalhos futuros

O método proposto apresenta uma abordagem alternativa para a segmentação de regiões geográficas com dados multidimensionais numéricos georreferenciados. A combinação da avaliação variográfica e o uso dos valores esperados para cada segmento em cada camada, permitem a identificação da extensão da autocorrelação espacial e sua delimitação, gerando como produto final os segmentos que melhor representam a região, dada sua distribuição e maximizando a heterogeneidade entre os segmentos por meio da avaliação da curva PM. A expansão da janela de avaliação, quando possível, também permite encontrar a melhor escala de representação do fenômeno espacial predominante na região. Acreditamos que estas sejam as maiores contribuições deste trabalho.

O uso destas definições pode permitir a aplicação de novos modos de se criar modelos de distribuição espaciais de dados, uma vez que a avaliação passa a não ser mais pixel a pixel, mas segmento por segmento. Esta abordagem pode permitir o surgimento de novas técnicas de modelagem, principalmente se aplicada a dados de espécies. Ao mesmo tempo, carece de investigação sobre a geração de métricas para avaliação da qualidade dos modelos que utilizarão a técnica de segmentação ou objetos e não mais por valores de *pixels*. Além disso, a generalização do método necessita de uma prova formal.

Referências

ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433–459, jul. 2010. ISSN 1939-0068. DOI: 10.1002/wics.101.

ABDULMUNIM, M. E.; PETROLEUM, S. I. M. Developing the range operator in the segmentation methods depending on dermatology infection. **IJCIT**, v. 1, n. 1, p. 15–19, sep. 2012. ISSN 2277–0764. Disponível em: <<http://ijcit.com/archives/volume1/issue1/Paper010103.pdf>>. Acesso em: 18 jan. 2016.

ACHANCCARAY, P. et al. A free software tool for automatic tuning of segmentation parameters. **South-Eastern European Journal of Earth Observation and Geomatics**, v. 3, p. 707–712, 2014. ISSN 2241-1224.

AITKEN, A. C. On Least Squares and Linear Combination of Observations. In: ROYAL SOCIETY OF EDINBURGH, 4. **Proceedings of...** 1936. v. 55, p. 42–48. ISSN 0370-1646. DOI: 10.1017/S0370164600014346. Disponível em: <http://www.journals.cambridge.org/abstract_S0370164600014346>. Acesso em: 18 jan. 2016.

ALMEIDA-GOMES, M. et al. Herpetofauna of an Atlantic rainforest area (Morro São João) in Rio de Janeiro State, Brazil. In: ACADEMIA BRASILEIRA DE CIÊNCIAS. **Anais...** 2008. v. 80, n. 2, p. 291–300. ISSN 0001-3765. DOI: 10.1590/S0001-37652008000200007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0001-37652008000200007&lng=en&nrm=iso&tlng=en>. Acesso em: 18 jan. 2016.

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, dec. 2013. DOI: 10.1127/0941-2948/2013/0507.

ARMSTRONG, C. G.; CONTE, C. E. Assemblage of anurans (Amphibia, Anura) of an area of Atlantic Forest, South of Brazil. **Biota Neotropica**, v. 10, n. 1, p. 39–46, mar. 2010. ISSN 1676-0603. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1676-06032010000100003&lng=en&nrm=iso&tlng=pt>. Acesso em: 18 jan. 2016.

ARTHUR, D.; VASSILVITSKII, S. k-means++: the advantages of careful seeding. In: ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS, 18. **Proceedings of...** Society for Industrial and Applied Mathematics, 2007. p. 1027–1035. Disponível em: <<http://dl.acm.org/citation.cfm?id=1283494>>. Acesso em: 08 fev. 2016.

BAATZ, M.; SCHÄPE, A.; STROBL, J. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: BEITRÄGE ZUM AGIT-SYMPOSIUM SALZBURG, HERBERT WICHMANN VERLAG. **12. Conference...** Karlsruhe, Germany, 2000. (Angewandte Geographische Informationsverarbeitung), p. 12–23.

- BACHMAIER, M.; BACKES, M. Variogram or Semivariogram? Variance or Semivariance? Allan Variance or Introducing a New Term? **Mathematical Geosciences**, v. 43, n. 6, p. 735–740, ago. 2011. ISSN 1874-8961, 1874-8953. Disponível em: <<http://link.springer.com/10.1007/s11004-011-9348-3>>. Acesso em: 18 jan. 2016.
- BARNOSKY, A. D. et al. Has the Earth's sixth mass extinction already arrived? **Nature**, v. 471, n. 7336, p. 51–57, mar. 2011. ISSN 0028-0836. Disponível em: <<http://www.nature.com/nature/journal/v471/n7336/full/nature09678.html>>. Acesso em: 18 jan. 2016.
- BELLMAN, R. E.; DREYFUS, S. **Dynamic programming**. 1st ed. Princeton, New Jersey: Princeton University, 2010. (Princeton Landmarks in mathematics). ISBN 978-0-691-14668-3.
- BERTNESS, M. D.; CALLAWAY, R. Positive interactions in communities. **Trends in Ecology & Evolution**, v. 9, n. 5, p. 191–193, maio 1994. ISSN 0169-5347. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0169534794900884>>. Acesso em: 18 jan. 2016.
- BETTIOL, G. M. **Revisão Áreas Prioritárias para Conservação da Biodiversidade (importância biológica) - 2007**. Brasília: Projeto de Conservação e Utilização Sustentável da Diversidade Biológica Brasileira - PROBIO/MMA, 2007. Mapa Digital. Disponível em: <<http://mapas.mma.gov.br/geonetwork/srv/br/metadata.show?id=408>>. Acesso em: 2016-03-22.
- BEYER, K. et al. When Is “Nearest Neighbor” Meaningful? In: BEERI, C.; BUNEMAN, P. (Ed.). **Database Theory — ICDT’99**. Berlin: Springer Heidelberg, 1999, (Lecture Notes in Computer Science, 1540). p. 217–235. ISBN 978-3-540-65452-0 978-3-540-49257-3. DOI: 10.1007/3-540-49257-7_15.
- BIVAND, R. S.; PEBESMA, E.; GÓMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**. New York: Springer, 2013. ISBN 978-1-4614-7617-7 978-1-4614-7618-4. Disponível em: <<http://link.springer.com/10.1007/978-1-4614-7618-4>>. Acesso em: 18 jan. 2016.
- BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, n. 1, p. 2–16, 2010. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0924271609000884>>. Acesso em: 18 jan. 2016.
- BOHLING, G. C. Introduction to Geostatistics. **Kansas Geological Survey Open File Report no**, v. 26, n. 2007, p. 50, 2007. Disponível em: <<http://sh.st/st/787f28ed3e745c14417e4aec27303038/http://people.ku.edu/~gbohling/BoiseGeostat/IntroToGeostatistics.pdf>>. Acesso em: 18 jan. 2016.
- BORCARD, D.; GILLET, F.; LEGENDRE, P. **Numerical Ecology with R**. New York: Springer, 2011. ISBN 978-1-4419-7975-9, 978-1-4419-7976-6. Disponível em: <<http://link.springer.com/10.1007/978-1-4419-7976-6>>. Acesso em: 18 jan. 2016.
- BORCARD, D.; LEGENDRE, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. **Ecological Modelling**, v. 153, n. 1–2, p. 51–68, jul. 2002. ISSN 0304-3800. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0304380001005014>>. Acesso em: 18 jan. 2016.

- BORCARD, D. et al. Dissecting the spatial structure of ecological data at multiple scales. **Ecology**, v. 85, n. 7, p. 1826–1832, jul. 2004. ISSN 0012-9658. Disponível em: <<http://www.esajournals.org/doi/abs/10.1890/03-3111>>. Acesso em: 18 jan. 2016.
- BOYLE, T. J. B.; BOYLE, C. E. B. **Biodiversity, Temperate Ecosystems, and Global Change**. New York: Springer Science & Business Media, 2013. ISBN 978-3-642-78972-4.
- BRASIL. Ministério do Meio Ambiente. **Mapas de cobertura vegetal dos biomas brasileiros**. Brasília: Ministério do Meio Ambiente, 2007. Disponível em: <<http://mapas.mma.gov.br/mapas/aplic/probio/datadownload.htm>>. Acesso em: 15 dez. 2015.
- CÂMARA, G. et al. TerraLib: an open source GIS library for large-scale environmental and socio-economic applications. In: HALL, G. B.; LEAHY, M. G. (Ed.). **Open Source Approaches in Spatial Data Handling**. Berlin: Springer Heidelberg, 2008, (Advances in Geographic Information Science, 2). p. 247–270. ISBN 978-3-540-74830-4 978-3-540-74831-1. DOI: 10.1007/978-3-540-74831-1_12.
- CAMARGO, E. C. G.; FELGUEIRAS, C. A.; MONTEIRO, A. M. V. A importância da modelagem da anisotropia na distribuição espacial de variáveis ambientais utilizando procedimentos geoestatísticos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 10. **Anais...** Foz do Iguaçu, PR, 2001. v. 21, p. 395–402. Disponível em: <<http://leg.ufpr.br/lib/exe/fetch.php/pessoais:abtmartins:0395.402.280.pdf>>. Acesso em: 18 jan. 2016.
- CANGELOSI, R.; GORIELY, A. Component retention in principal component analysis with application to cDNA microarray data. **Biology Direct**, v. 2, p. 2, jan. 2007. ISSN 1745-6150. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1797006/>>.
- CASTILLA, G. **Object-oriented analysis of remote sensing images for land cover mapping**: conceptual foundations and a segmentation method to derive a baseline partition for classification. Tese (Ph. D. Thesis) — Polytechnic University of Madrid, Madrid, Spain, 2003. Disponível em: <http://oa.upm.es/133/01/07200302_castilla_castellano.pdf>. Acesso em: 27 mar. 2016.
- CEBALLOS, G. et al. Accelerated modern human-induced species losses: Entering the sixth mass extinction. **Science Advances**, v. 1, n. 5, p. e1400253, jun. 2015. ISSN 2375-2548. Disponível em: <<http://advances.sciencemag.org/content/1/5/e1400253>>. Acesso em: 18 jan. 2016.
- CHILES, J.-P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2nd ed. Hoboken, New Jersey: Wiley, 2012. (Wiley series in probability and statistics). ISBN 9780470183151.
- COELHO, A. M. **Agricultura de precisão: manejo da variabilidade espacial e temporal dos solos e culturas**. Sete Lagoas, MG: Embrapa Milho e Sorgo, 2005.
- CONEGO, J. G. L. **Precipitação anual média do Brasil (1960-1990)**. Brasília - Brazil: Agência Nacional de Águas, 2005. Disponível em: <<http://mapas.mma.gov.br/mostratema.php?temas=isoietas>>. Acesso em: 18 jan. 2016.

- CONNELL, J. H. Diversity in tropical rain forests and coral reefs. **Science**, v. 199, n. 4335, p. 1302–1310, mar. 1978. ISSN 0036-8075, 1095-9203. Disponível em: <<http://www.sciencemag.org/content/199/4335/1302>>. Acesso em: 18 jan. 2016.
- COSTA, W. F. et al. Applying multiresolution segmentation algorithm to generate crop management zones based on interpolated layers. In: EFITA-WCCA-CIGR CONFERENCE SUSTAINABLE AGRICULTURE THROUGH ICT INNOVATION, 2013, Turim, Itália. **Proceedings of...** Turim, Itália, 2013. p. 7. Disponível em: <<http://www.cigr.org/Proceedings/uploads/2013/0214.pdf>>. Acesso em: 5 mar. 2016.
- DÁTTILO, W.; AES, P. R. G.; IZZO, T. J. Spatial structure of ant–plant mutualistic networks. **Oikos**, v. 122, n. 11, p. 1643–1648, nov. 2013. ISSN 1600-0706. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0706.2013.00562.x/abstract>>. Acesso em: 18 jan. 2016.
- DEUTSCH, C. V.; JOURNEL, A. G. **GSLIB**: geostatistical software library and user's guide. New York: Oxford University Press, 1998. ISBN 0-19-510015-8.
- DIGGLE, P.; RIBEIRO JR., P. J. **Model-based Geostatistics**. New York: Springer New York, 2010. (Springer Series in Statistics). OCLC: 802786793. ISBN 978-0-387-48536-2 978-1-4419-2193-2.
- DING, C.; HE, X. K-means clustering via principal component analysis. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 21. **Proceedings of...** Madison, Wisconsin, USA: ACM, 2004. p. 29. DOI: 10.1145/1015330.1015408.
- DOERGE, T. Defining management zones for precision farming. **Crop Insight**, v. 8, n. 21, 1999.
- DRĂGUȚ, L.; EISANK, C. Automated object-based classification of topography from SRTM data. **Geomorphology**, v. 141–142, p. 21–33, mar. 2012. ISSN 0169-555X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169555X11006143>>. Acesso em: 18 jan. 2016.
- DRAY, S.; LEGENDRE, P.; PERES-NETO, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). **Ecological Modelling**, v. 196, n. 3-4, p. 483–493, jul. 2006. ISSN 03043800. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0304380006000925>>. Acesso em: 18 jan. 2016.
- DRUCK, S. **Análise espacial de dados geográficos**. Planaltina: Embrapa Cerrados, 2004. ISBN 8573832606 9788573832600.
- DUNFORD, W.; FREEMARK, K. Matrix Matters: Effects of Surrounding Land Uses on Forest Birds Near Ottawa, Canada. **Landscape Ecology**, v. 20, n. 5, p. 497–511, jul. 2005. ISSN 0921-2973, 1572-9761. Disponível em: <<http://link.springer.com/10.1007/s10980-004-5650-5>>. Acesso em: 18 jan. 2016.
- EHRlich, P. R.; EHRlich, A. H. Can a collapse of global civilization be avoided? In: ROYAL SOCIETY OF LONDON B: BIOLOGICAL SCIENCES. **Proceedings of...** 2013. v. 280, n. 1754, p. 20122845. ISSN 0962-8452, 1471-2954. Disponível em: <<http://rspb.royalsocietypublishing.org/content/280/1754/20122845>>. Acesso em: 18 jan. 2016.

ESPINDOLA, G. M. **Ajuste de parâmetros em algoritmos de segmentação de imagens por crescimento de regiões**. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, 2007.

FAO; ITPS. **Status of the World's Soil Resources (SWSR) - technical summary**. Rome: Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, 2015. ISBN 978-92-5-108960-6.

FERRÉ, L. Selection of components in principal component analysis: A comparison of methods. **Computational Statistics & Data Analysis**, v. 19, n. 6, p. 669–682, jun. 1995. ISSN 0167-9473. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016794739400020J>>.

FERREIRA, R. S. **A multiscalar, multicriteria approach for image segmentation**. Dissertação (Mestrado) — Pontifícia Universidade Católica, Rio de Janeiro, RJ, 2011. DOI: 10.17771/PUCRio.acad.18290.

FITCH, P. J. R. et al. An integrated and quantitative approach to petrophysical heterogeneity. **Marine and Petroleum Geology**, v. 63, p. 82–96, maio 2015. ISSN 0264-8172. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0264817215000434>>. Acesso em: 18 jan. 2016.

FRANKLIN, S. B. et al. Parallel Analysis: a method for determining significant principal components. **Journal of Vegetation Science**, v. 6, n. 1, p. 99–106, fev. 1995. ISSN 1654-1103. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.2307/3236261/abstract>>.

GAREY, M. V.; HARTMANN, M. T. Anurans of reserva natural Salto Morato, municipality of Guaraqueçaba, state of Paraná, southern Brazil. **Biota Neotropica**, v. 12, n. 4, p. 137–145, dez. 2012. ISSN 1676-0603. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1676-06032012000400015&lng=en&nrm=iso&tlng=pt>. Acesso em: 18 jan. 2016.

GAUCH, H. G.; WHITTAKER, R. H. Coenocline Simulation. **Ecology**, v. 53, n. 3, p. 446–451, 1972. ISSN 0012-9658. Disponível em: <<http://www.jstor.org/stable/1934231>>. Acesso em: 18 jan. 2016.

GEARY, R. C. The contiguity ratio and statistical mapping. **The Incorporated Statistician**, v. 5, n. 3, p. 115–146, nov. 1954. ISSN 1466-9404. Disponível em: <<http://www.jstor.org/stable/2986645>>. Acesso em: 18 jan. 2016.

GETIS, A.; ORD, J. K. The Analysis of Spatial Association by Use of Distance Statistics. **Geographical Analysis**, v. 24, n. 3, p. 189–206, set. 2010. ISSN 00167363. DOI: 10.1111/j.1538-4632.1992.tb00261.x.

GOODCHILD, M. F. **Spatial autocorrelation**. Norwich: Geo Books, 1986. ISBN 9780860942238.

GOWER, J. C. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. **Biometrika**, v. 53, n. 3/4, p. 325–338, dez. 1966. ISSN 0006-3444. Disponível em: <<http://www.jstor.org/stable/2333639>>. Acesso em: 18 jan. 2016.

GRINNELL, J. **The Niche-Relationships of the California Thrasher**. The Auk, 1917. Disponível em: <<http://archive.org/details/jstor-4072271>>. Acesso em: 18 jan. 2016.

GUÉNARD, G. et al. Multiscale codependence analysis: an integrated approach to analyze relationships across scales. **Ecology**, v. 91, n. 10, p. 2952–2964, 2010. Disponível em: <<http://www.esajournals.org/doi/abs/10.1890/09-0460.1>>. Acesso em: 18 jan. 2016.

HAINING, R. P. **Spatial data analysis in the social and environmental sciences**. Cambridge [England]; New York: Cambridge University Press, 1993. ISBN 978-0-521-38416-2 978-0-521-44866-6.

HALBRITTER, A. H. et al. How comparable are species distributions along elevational and latitudinal climate gradients? **Global Ecology and Biogeography**, v. 22, n. 11, p. 1228–1237, nov. 2013. ISSN 1466-8238. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/geb.12066/abstract>>. Acesso em: 27 mar. 2016.

HAPP, P. N. et al. Segmentação multiresolução: uma abordagem paralela para segmentação de imagens de alta resolução em arquiteturas de múltiplos núcleos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14. **Anais...** Natal, Brasil: INPE, 2009. p. 6935–6942.

HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. **Applied Statistics**, v. 28, n. 1, p. 100, 1979. ISSN 00359254. Disponível em: <<http://www.jstor.org/stable/10.2307/2346830?origin=crossref>>. Acesso em: 18 jan. 2016.

HAY, G. J.; CASTILLA, G. Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT). In: INTERNATIONAL CONFERENCE ON OBJECT-BASED IMAGE ANALYSIS (OBIA 2006), 1. **Proceedings of...** 2006. Disponível em: <http://www.isprs.org/proceedings/XXXVI/4-C42/Papers/01_Opening%20Session/OBIA2006_Hay_Castilla.pdf>. Acesso em: 07 jan. 2015.

_____. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: BLASCHKE, T. et al. (Ed.). **Object-Based Image Analysis**. [S.l.]: Springer Berlin Heidelberg, 2008, (Lecture Notes in Geoinformation and Cartography). p. 75–89. ISBN 978-3-540-77058-9.

HENGL, T. et al. **A practical guide to geostatistical mapping of environmental variables**. Luxembourg: Publications Office, 2007. ISBN 978-92-79-06904-8 92-79-06904-7.

HIEMSTRA, P. H. et al. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. **Computers & Geosciences**, v. 35, n. 8, p. 1711–1721, ago. 2009. ISSN 00983004. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0098300409000867>>. Acesso em: 18 jan. 2016.

HIJMAN, R. J. et al. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology**, v. 25, n. 15, p. 1965–1978, dez. 2005. ISSN 0899-8418, 1097-0088. DOI: 10.1002/joc.1276.

HUSSAIN, S.; MILLER, D. **The Economics of Ecosystems and Biodiversity (TEEB) for Agriculture & Food – Concept Note**. TEEB, 2014. Disponível em: <<http://www.teebweb.org/publication/the-economics-of-ecosystems-and-biodiversity-teeb-for-agriculture-food-concept-note/>>. Acesso em: 23 abr. 2016.

- HUTCHINSON, M. F. Concluding remarks. In: **Population studies: Animal ecology and demography**. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.: [s.n.], 1957. p. 415–427. Disponível em: <<http://link.springer.com/article/10.1007/BF02464429>>. Acesso em: 18 jan. 2016.
- JACKSON, D. A. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. **Ecology**, v. 74, n. 8, p. 2204–2214, 1993. ISSN 0012-9658. Disponível em: <<http://www.jstor.org/stable/1939574>>. Acesso em: 18 jan. 2016.
- JACKSON, J. E. **A user's guide to principal components**. [S.l.]: John Wiley & Sons, 2005.
- JEFFERY, M. I.; FIRESTONE, J.; BUBNA-LITIC, K. **Biodiversity Conservation, Law and Livelihoods**: Bridging the north-south divide: IUCN Academy of Environmental Law Research Studies. Cambridge: Cambridge University Press, 2008. ISBN 978-1-139-46912-8.
- KAGANAMI, H. G.; BEIJI, Z. Region-based segmentation versus edge detection. In: INTERNATIONAL CONFERENCE ON INTELLIGENT INFORMATION HIDING AND MULTIMEDIA SIGNAL PROCESSING IIH-MSP '09, 5. **Proceedings of...** Kyoto, Japan, 2009. p. 1217–1221.
- KAHN, C. H. **The art and thought of Heraclitus**: A new arrangement and translation of the fragments with literary and philosophical commentary. [S.l.]: Cambridge University Press, 1981. ISBN 978-0-521-28645-9.
- KATAYAMA, N. et al. Landscape heterogeneity–biodiversity relationship: effect of range size. **PLoS ONE**, v. 9, n. 3, p. e93359, mar. 2014. ISSN 1932-6203. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093359>>. Acesso em: 18 jan. 2016.
- KHOSLA, R. et al. Spatial variation and site-specific management zones. In: M.A. Oliver (Ed.). **Geostatistics with applications in earth sciences**. Dordrecht: Springer Netherlands, 2010. p. 195–219. ISBN 978-90-481-9132-1.
- KNEGT, H. J. D. et al. Spatial autocorrelation and the scaling of species–environment relationships. **Ecology**, v. 91, n. 8, p. 2455–2465, 2010. Disponível em: <<http://www.esajournals.org/doi/abs/10.1890/09-1359.1>>. Acesso em: 18 jan. 2016.
- KOURGLI, A.; BELHADJ-AISSA, A. Texture primitives description and segmentation using variography and mathematical morphology. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. **Conference...** Netherlands, 2004. v. 7, p. 6360–6365 vol.7.
- LANDIM, P. M. B. **Análise estatística de dados geológicos multivariados**. São Paulo: Oficina de Textos, 2011. ISBN 978-85-7975-032-8.
- LEGENDRE, P. Quantitative methods and biogeographic analysis. In: **Evolutionary biogeography of the marine algae of the North Atlantic**. Berlin: Springer, 1990, (NATO ASI series Ser. G, v. 22). p. 9–34. ISBN 3540517618 3540517618 0387517618.

LEGENDRE, P.; LEGENDRE, L. F. J. **Numerical Ecology**. New Jersey: Elsevier, 2012. ISBN 9780444538697.

LEHMAN, S. M. Effects of altitude on the conservation biogeography of lemurs in southeast madagascar. In: GROW, N. B.; GURSKY-DOYEN, S.; KRZTON, A. (Ed.). **High Altitude Primates**. New York: Springer, 2014, (Developments in Primatology: Progress and Prospects). p. 3–22. ISBN 978-1-4614-8174-4 978-1-4614-8175-1. DOI: 10.1007/978-1-4614-8175-1_1.

LEVINE, N. Spatial Autocorrelation Statistics. In: **CrimeStat**: a spatial statistics program for the analysis of crime incident locations (v. 4.0). Houston, TX: Ned Levine & Associate, 2010. Disponível em: <<http://beta.nij.gov/topics/technology/maps/Documents/crimestat-files/CrimeStat%20IV%20Chapter%205.pdf>>. Acesso em: 18 jan. 2016.

LIRA, P. K. et al. Land-use and land-cover change in Atlantic Forest landscapes. **Forest Ecology and Management**, v. 278, p. 80–89, ago. 2012. ISSN 03781127. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0378112712002630>>. Acesso em: 18 jan. 2016.

LOEHRLEIN, M. **Sustainable Landscaping**: principles and practices. [S.l.]: CRC Press, 2013. ISBN 978-1-4665-9321-3.

MACHADO, P. L. O. A.; BERNARDI, A. C. C.; SILVA, C. A. (Ed.). **Agricultura de precisão para o manejo da fertilidade do solo em sistema de plantio direto**. Rio de Janeiro, RJ: Embrapa Solos, 2004. ISBN 8-85864-13-3.

MAIA NETO, J. et al. **Plano operativo de prevenção e combate aos incêndios florestais do parque nacional da serra da canastra**. São Roque de Minas - MG, Brasil, 2007. 23 p. Disponível em: <http://www.ibama.gov.br/phocadownload/prevfogo/plano_operativo_parna_da_serra_da_canastra-2007.pdf>. Acesso em: 14 mar. 2016.

MARROT, P.; GARANT, D.; CHARMANTIER, A. Spatial autocorrelation in fitness affects the estimation of natural selection in the wild. **Methods Ecol Evol**, v. 6, n. 12, p. 1474–1483, dez. 2015. ISSN 2041-210X. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12448/abstract>>.

MARTHA, T. R. et al. Segment Optimization and Data-Driven Thresholding for Knowledge-Based Landslide Detection by Object-Based Image Analysis. **IEEE Transactions on Geoscience and Remote Sensing**, v. 49, n. 12, p. 4928–4943, dez. 2011. ISSN 0196-2892.

MATHERON, G. **The theory of regionalized variables and its applications**. [S.l.]: École nationale supérieure des mines, 1971.

MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**, v. 22, n. 1, p. 83–92, 2002.

MORAES, R. A. de; SAWAYA, R. J.; BARRELLA, W. Composição e diversidade de anfíbios anuros em dois ambientes de Mata Atlântica no Parque Estadual Carlos Botelho, São Paulo, sudeste do Brasil. **Biota Neotrop**, v. 7, n. 2, p. 1–10, 2007. Disponível em: <<http://www.scielo.br/pdf/bn/v7n2/a03v07n2>>. Acesso em: 18 jan. 2016.

- MORAN, P. A. P. Notes on Continuous Stochastic Phenomena. **Biometrika**, v. 37, n. 1/2, p. 17–23, jun. 1950. ISSN 0006-3444. Disponível em: <<http://www.jstor.org/stable/2332142>>. Acesso em: 27 mar. 2016.
- MUELLER, M.; SEGL, K.; KAUFMANN, H. Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. **Pattern Recognition**, v. 37, n. 8, p. 1619–1628, ago. 2004. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320304000548>>.
- NARVAES, P.; BERTOLUCI, J.; RODRIGUES, M. T. Composição, uso de hábitat e estações reprodutivas das espécies de anuros da floresta de restinga da Estação Ecológica Juréia-Itatins, sudeste do Brasil. **Biota Neotrop**, v. 9, n. 2, p. 1–7, 2009. Disponível em: <<http://www.scielo.br/pdf/bn/v9n2/a11v09n2>>. Acesso em: 27 mar. 2016.
- NAVULUR, K. **Multispectral Image Analysis Using the Object-Oriented Paradigm**. [S.l.]: CRC Press, 2006. ISBN 9781420043075.
- NERY, T. C. **Diversidade de anfíbios anuros em áreas antropizadas em Vitória da Conquista – BA**. Dissertação (Mestrado) — Universidade Estadual do Sudoeste da Bahia – UESB – Campus de Itapetinga, Vitória da Conquista – BA, 2014. Disponível em: <<http://www.uesb.br/ppgca/dissertacoes/2014/THIAGO-CASTRO.pdf>>. Acesso em: 27 mar. 2016.
- NUSSBAUM, S.; MENZ, G. eCognition Image Analysis Software. In: **Object-based image analysis and treaty verification**. Netherlands: Springer, 2008. p. 29–39. ISBN 978-1-4020-6960-4 978-1-4020-6961-1. DOI: 10.1007/978-1-4020-6961-1_3.
- OKSANEN, J. et al. **vegan**: Community ecology package. [s.n.], 2016. R package version 2.3-3. Disponível em: <<https://CRAN.R-project.org/package=vegan>>. Acesso em: 18 jan. 2016.
- OLIVER, M. (Ed.). **Geostatistical applications for precision agriculture**. Dordrecht, Netherlands: Springer, 2010. ISBN 978-90-481-9132-1 978-90-481-9133-8. Disponível em: <<http://link.springer.com/10.1007/978-90-481-9133-8>>. Acesso em: 13 jan. 2016.
- PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, p. 683–691, 2004.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, n. 5, p. 1633–1644, oct. 2007. ISSN 1607-7938. Disponível em: <<http://www.hydrol-earth-syst-sci.net/11/1633/2007/>>. Acesso em: 27 mar. 2016.
- PNUD - Programa das Nações Unidas para o Desenvolvimento. **Acompanhando a agenda 2030 para o desenvolvimento sustentável**: subsídios iniciais do Sistema das Nações Unidas no Brasil sobre a identificação de indicadores nacionais referentes aos objetivos de desenvolvimento sustentável. Brasília: PNUD, 2015. ISBN 978-85-88201-29-3. Disponível em: <http://www.pnud.org.br/Docs/Acompanhando_Agenda2030-Subsidios_iniciais-Brasil.pdf>. Acesso em: 25 abr. 2016.

PRADO, G. M.; POMBAL JR., J. P. Distribuição espacial e temporal dos anuros em um brejo da Reserva Biológica de Duas Bocas, sudeste do Brasil. **Arquivos do Museu Nacional**, v. 63, n. 4, p. 685–705, 2005.

R Core Team. **R**: a language and environment for statistical computing. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org/>>. Acesso em: 27 mar. 2016.

RAMIADANTSOA, T. et al. Large-Scale Habitat Corridors for Biodiversity Conservation: A Forest Corridor in Madagascar. **PLoS ONE**, v. 10, n. 7, p. e0132126, jul. 2015. DOI: 10.1371/journal.pone.0132126.

RIBEIRO, R. da S.; EGITO, G. T. B. T. do; HADDAD, C. F. B. Chave de identificação: anfíbios anuros da vertente de Jundiá da Serra do Japi, Estado de São Paulo. **Biota Neotropica**, v. 5, n. 2, p. 235–247, 2005. ISSN 1676-0603. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1676-06032005000300017&lng=en&nrm=iso&tlng=pt>. Acesso em: 27 mar. 2016.

RITCHIE, M. E. **Scale, heterogeneity, and the structure and diversity of ecological communities**. Princeton: Princeton University, 2009. ISBN 978-1-4008-3168-5.

ROE, J. H. et al. No place like home: an experimental comparison of reintroduction strategies using snakes. **Journal of Applied Ecology**, v. 47, n. 6, p. 1253–1261, dez. 2010. ISSN 1365-2664. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2664.2010.01886.x/abstract>>.

RYBICKI, J.; HANSKI, I. Species-area relationships and extinctions caused by habitat loss and fragmentation. **Ecology Letters**, v. 16, p. 27–38, maio 2013. ISSN 1461023X. Disponível em: <<http://doi.wiley.com/10.1111/ele.12065>>.

SABA, F.; VALADANZOUJ, M.; MOKHTARZADE, M. The optimization of multi resolution segmentation of remotely sensed data using genetic algorithm. In: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. **Proceedings of...** Tehran, Iran, 2013. Volume XL-1/W3, p. 345–349. Disponível em: <<http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-1-W3/345/2013/isprsarchives-XL-1-W3-345-2013.pdf>>. Acesso em: 31 out. 2016.

SANDERS, N. J.; RAHBEK, C. The patterns and causes of elevational diversity gradients. **Ecography**, v. 35, n. 1, p. 1–3, jan. 2012. ISSN 09067590. Disponível em: <<http://doi.wiley.com/10.1111/j.1600-0587.2011.07338.x>>. Acesso em: 27 mar. 2016.

SANTOS, R. T. **Um modelo de referência para o processo de definição de zonas de manejo em agricultura de precisão**. Tese (Doutorado) — Universidade de São Paulo, 2014. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3141/tde-26082015-154543/>>. Acesso em: 27 mar. 2016.

SAYRE, R. et al. **A new map of global ecological land units - an ecophysiographic stratification approach**: a special publication of the Association of American Geographers. Washington, DC: Association of American Geographers, 2014. ISBN 978-0-89291-276-6.

- SILVA-SOARES, T. et al. Anuran Amphibians from RPPN Campo Escoteiro Geraldo Hugo Nunes, Guapimirim Municipality, Rio de Janeiro, Southeastern Brazil. **Biota Neotropica**, v. 10, n. 2, p. 225–233, jun. 2010. ISSN 1676-0603. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1676-06032010000200025&lng=en&nrm=iso&tlng=pt>. Acesso em: 27 mar. 2016.
- SIQUEIRA, C. C. et al. Species composition and density estimates of the anurofauna of a site within the northernmost large Atlantic Forest remnant (Parque Estadual do Desengano) in the state of Rio de Janeiro, Brazil. **Biota Neotropica**, v. 11, n. 4, p. 131–137, dez. 2011. ISSN 1676-0603. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1676-06032011000400014&lng=en&nrm=iso&tlng=en>. Acesso em: 18 jan. 2016.
- SKOV, F. Potential plant distribution mapping based on climatic similarity. **Taxon**, v. 49, n. 3, p. 503–515, 2000. ISSN 0040-0262. DOI: 10.2307/1224346.
- SRIVASTAVA, R. M.; PARKER, H. M. Robust measures of spatial continuity. In: ARMSTRONG, D. M. (Ed.). **Geostatistics**. Netherlands: Springer, 1989, (Quantitative Geology and Geostatistics, 4). p. 295–308. ISBN 978-94-015-6846-3, 978-94-015-6844-9. Disponível em: <http://link.springer.com/chapter/10.1007/978-94-015-6844-9_22>. Acesso em: 27 mar. 2016.
- STEAD, D.; UNECE - United Nations Economic Commission for Europe. **Spatial planning**: key instrument for development and effective governance with special reference to countries in transition. Geneva, Switzerland: UNITED NATIONS, 2008. Disponível em: <https://www.unece.org/fileadmin/DAM/hlm/documents/Publications/spatial_planning.e.pdf>. Acesso em: 31 out. 2016.
- THORNDIKE, R. L. Who belongs in the family? **Psychometrika**, v. 18, n. 4, p. 267–276, dez. 1953. ISSN 0033-3123, 1860-0980. DOI: 10.1007/BF02289263.
- TILMAN, D. Forecasting agriculturally driven global environmental change. **Science**, v. 292, n. 5515, p. 281–284, abr. 2001. ISSN 00368075, 10959203. DOI: 10.1126/science.1057544.
- VARADHAN, R. et al. **features**: feature extraction for discretely-sampled functional data. [s.n.], 2011. R package version 2011.8-2. Disponível em: <<http://CRAN.R-project.org/package=features>>. Acesso em: 27 mar. 2016.
- VASCONCELOS, T. S. et al. Biogeographic distribution patterns and their correlates in the diverse frog Fauna of the Atlantic Forest Hotspot. **PLoS ONE**, v. 9, n. 8, p. e104130, ago. 2014. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0104130>>. Acesso em: 27 mar. 2016.
- VINCENT, L.; SOILLE, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 13, n. 6, p. 583–598, jun. 1991. ISSN 0162-8828.
- VITULE, J. R. S. et al. Brazil's drought: protect biodiversity. **Science**, v. 347, n. 6229, p. 1427–1428, mar. 2015. ISSN 0036-8075, 1095-9203. Disponível em: <<http://www.sciencemag.org/content/347/6229/1427.2>>. Acesso em: 27 mar. 2016.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for environmental scientists**. 2nd ed. Chichester: Wiley, 2007. (Statistics in practice). ISBN 978-0-470-02858-2.

WOODCOCK, C. E.; STRAHLER, A. H. The factor of scale in remote sensing. **Remote Sensing of Environment**, v. 21, n. 3, p. 311–332, abr. 1987. ISSN 0034-4257. DOI: 10.1016/0034-4257(87)90015-0.

WULDER, M. A.; NELSON, T. A.; SEEMANN, D. Using spatial pattern to quantify relationship between samples, surroundings, and populations. **Environmental Monitoring and Assessment**, v. 131, n. 1-3, p. 221–230, jun. 2007. ISSN 0167-6369, 1573-2959. DOI: 10.1007/s10661-006-9470-8. Disponível em: <<http://link.springer.com/10.1007/s10661-006-9470-8>>. Acesso em: 05 abr. 2016.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística. Conceitos e Aplicações**. São Paulo: Oficina de Textos, 2013. ISBN 9788579750779.

ZANE, L. et al. Within-field zoning using a region growing algorithm guided by geostatistical analysis. In: STAFFORD, J. V. (Ed.). **Precision agriculture '13**. Wageningen Academic Publishers, 2013. p. 313–319. ISBN 978-90-8686-778-3. DOI: 10.3920/978-90-8686-778-3_37. Disponível em: <http://link.springer.com/chapter/10.3920/978-90-8686-778-3_37>. Acesso em: 18 jan. 2016.

ZHANG, X. et al. Zone mapping application for precision-farming: a decision support tool for variable rate application. **Precision Agriculture**, v. 11, n. 2, p. 103–114, ago. 2009. ISSN 1385-2256, 1573-1618. DOI: 10.1007/s11119-009-9130-4. Disponível em: <<http://link.springer.com/article/10.1007/s11119-009-9130-4>>. Acesso em: 22 jan. 2016.

ZUUR, A. F.; IENO, E. N.; SMITH, G. M. **Analysing Ecological Data**. New York: Springer Science + Business Media, LLC, 2007. ISBN 9780387459721 0387459723.