



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21c/2018/08.14.18.06-TDI

AVALIAÇÃO DA QUALIDADE DOS DADOS PRODES: ESTIMATIVA E REGIONALIZAÇÃO DOS ERROS DE MAPEAMENTO

Luis Eduardo Pinheiro Maurano

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelos Drs. Maria Isabel Sobral
Escada, e Camilo Daleles Rennó,
aprovada em 24 de agosto de 2018.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34R/3RL377B>>

INPE
São José dos Campos
2018

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GBDIR)

Serviço de Informação e Documentação (SESID)

CEP 12.227-010

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/7348

E-mail: pubtc@inpe.br

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):****Presidente:**

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CGCPT)

Membros:

Dra. Carina Barros Mello - Coordenação de Laboratórios Associados (COCTE)

Dr. Alisson Dal Lago - Coordenação-Geral de Ciências Espaciais e Atmosféricas (CGCEA)

Dr. Evandro Albiach Branco - Centro de Ciência do Sistema Terrestre (COCST)

Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia e Tecnologia Espacial (CGETE)

Dr. Hermann Johann Heinrich Kux - Coordenação-Geral de Observação da Terra (CGOBT)

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação - (CPG)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SESID)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SESID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SESID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SESID)

EDITORAÇÃO ELETRÔNICA:

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SESID)

Murilo Luiz Silva Gino - Serviço de Informação e Documentação (SESID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21c/2018/08.14.18.06-TDI

AVALIAÇÃO DA QUALIDADE DOS DADOS PRODES: ESTIMATIVA E REGIONALIZAÇÃO DOS ERROS DE MAPEAMENTO

Luis Eduardo Pinheiro Maurano

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelos Drs. Maria Isabel Sobral
Escada, e Camilo Daleles Rennó,
aprovada em 24 de agosto de 2018.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34R/3RL377B>>

INPE
São José dos Campos
2018

Dados Internacionais de Catalogação na Publicação (CIP)

Maurano, Luis Eduardo Pinheiro.

M445a Avaliação da qualidade dos dados Prodes: estimativa e regionalização dos erros de mapeamento / Luis Eduardo Pinheiro Maurano. – São José dos Campos : INPE, 2018.

xx + 96 p. ; (sid.inpe.br/mtc-m21c/2018/08.14.18.06-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2018.

Orientadores : Drs. Maria Isabel Sobral Escada, e Camilo Daleles Rennó.

1. Mapeamento de desmatamento da Amazônia Brasileira.
2. Sensoriamento Remoto. 3. Avaliação de mapeamento.
4. Espacialização de erros. I.Título.

CDU 504.122:528.8



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).


This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): **Luis Eduardo Pinheiro Maurano**

Título: "AVALIAÇÃO DA QUALIDADE DOS DADOS PRODES: ESTIMATIVA E REGIONALIZAÇÃO DOS ERROS DE MAPEAMENTO"

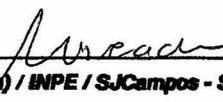
Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de **Mestre** em **Sensoriamento Remoto**

Dr. Antônio Miguel Vieira Monteiro

Pl: 


Presidente / INPE / SJCampos - SP
 Participação por Vídeo - Conferência
 Aprovado Reprovado

Dra. Maria Isabel Sobral Escada




Orientador(a) / INPE / SJCampos - SP
 Participação por Vídeo - Conferência
 Aprovado Reprovado

Dr. Camilo Daleles Rennó



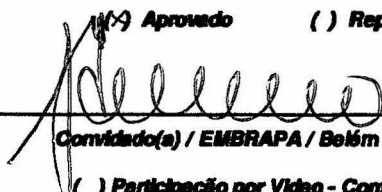
Orientador(a) / INPE / SJCampos - SP
 Participação por Vídeo - Conferência
 Aprovado Reprovado

Dr. Thales Sehn Körting



Membro da Banca / INPE / São José dos Campos - SP
 Participação por Vídeo - Conferência
 Aprovado Reprovado

Dr. Adriano Venturieri



Convidado(a) / EMBRAPA / Belém - PA
 Participação por Vídeo - Conferência
 Aprovado Reprovado

Este trabalho foi aprovado por:

maioria simples

unanimidade

São José dos Campos, 24 de agosto de 2018

“Os que se encantam com a prática sem a ciência são como os timoneiros que entram no navio sem timão nem bússola, nunca tendo certeza do seu destino”.

Leonardo da Vinci

*A minha esposa Luciana e filhos Lucas e Nathalia e aos netos Gabriel e
Isabela.*

AGRADECIMENTOS

Primeiramente agradeço aos meus orientadores - Dra. Maria Isabel Sobral Escada e Dr. Camilo Daleles Rennó, pelo apoio inestimável para a realização deste estudo, além da confiança, companheirismo a mim conferidos, bem como, aos membros da banca examinadora por terem aceitado o convite, dispondo de tempo e conhecimento para analisar este estudo.

Ao Instituto Nacional de Pesquisas Espaciais (INPE) que proporcionou o acesso ao conhecimento para esta realização através dos docentes envolvidos no curso de pós-graduação em Sensoriamento Remoto e todos os colegas da Divisão de Processamento de Imagens (DPI) que contribuíram e sempre incentivaram o desenvolvimento deste estudo.

Agradecimento especial aos colegas Paulo Alves, Maurício Trancho e Fernanda Baruel pelo trabalho extremamente importante de validação dos dados, além da Dra. Clotilde Ferri por ter dado suporte a esta etapa, todos servidores da Fundação de Ciências, Aplicações e Tecnologia Espaciais (Funcate). Agradeço também ao Raian Maretto pelo apoio no desenvolvimento dos programas em linguagem R, essenciais ao desenvolvimento deste estudo.

A Luciana e ao Lucas pelo amor, apoio, paciência e compreensão nos momentos de ausência.

Por fim, agradeço a todos os colegas da pós-graduação e, em especial a Lidiane Costa, pela amizade durante estes últimos dois anos.

RESUMO

O Instituto Nacional de Pesquisas Espaciais (INPE), no âmbito do Projeto de Monitoramento do Desmatamento na Amazônia por Satélites (PRODES), realiza desde 1988, o inventário anual de perda de área de cobertura florestal primária, através do mapeamento do desmatamento por corte raso na Amazônia Legal. O levantamento envolve grande extensão territorial, em que mais de 5 milhões de km² são mapeados anualmente, e as informações geradas são usadas pelo governo brasileiro para avaliação e estabelecimento de políticas públicas relativas ao controle do desmatamento. Apesar da série histórica do PRODES estar completando 30 anos, até hoje não se realizaram estudos para aferir e contabilizar os erros e incertezas encontrados no mapeamento para toda extensão da Amazônia Legal. A literatura especializada na área recomenda que levantamentos realizados com uso de dados de sensoriamento remoto venham sempre acompanhados de medidas de exatidão e/ou qualidade. Porém, os métodos utilizados para aferir estas medidas normalmente se restringem a quantificá-los de maneira global, sem regionalizá-los, nem espacializá-los. Em mapeamentos de larga escala territorial, a exemplo do executado pelo PRODES, informar índices globais de qualidade é importante, mas pode não ser suficiente se os dados não vierem acompanhados de indicadores que forneçam algum modelo de espacialização dos erros, levando em consideração os contextos regionais. Este estudo desenvolveu uma metodologia baseada em um modelo estatístico regionalizado para avaliar os dados gerados pelo PRODES para o ano de 2014. Apoiado em um ambiente espaço celular, foram extraídas métricas de ecologia de paisagem e com auxílio de técnicas de mineração de dados foi gerado um mapa com padrões de desmatamento espacializado em células de 50 x 50 km. Com base no mapa obtido foram sorteados aproximadamente 1800 pontos aleatórios para validação independentes. Um modelo de regressão logística multivariado foi desenvolvido e tres variáveis independentes relacionadas as métricas foram selecionadas para estimar o erro no mapeamento. Aplicado o modelo a toda base de dados foi possível espacializar a área de estudo em células com maior e menor probabilidade de se encontrar erro no mapeamento do PRODES, onde se verificou que 73% das células apresentam erro estimado menor que 1% enquanto somente 0,4% apresentam erro estimado maior que 10%. A distribuição do erro estimado por padrão de desmatamento também determinou que a grande maioria das células apresentasse erro concentrados na faixa menor que 1% independente do padrão. Esta abordagem de regionalização de erros por modelo de regressão logística se mostrou adequada e os resultados obtidos conferem ao mapeamento do PRODES de 2014 excelente nível de qualidade e confiabilidade encorajando sua replicação para outros mapeamentos semelhantes.

Palavras-chave: mapeamento de desmatamento na Amazônia Brasileira, sensoriamento remoto, avaliação de mapeamento e espacialização de erros.

DATA QUALITY ASSESSMENT: ESTIMATES AND REGIONALIZATION OF MAPPING ERRORS

ABSTRACT

The National Institute for Space Research (INPE) has carried out since 1998, in the scope of the Amazon Deforestation Satellite Monitoring (PRODES), the annual inventory of primary forest coverage through the mapping of clear cut deforestation in the legal Amazon. This mapping is done through the usage of the Landsat satellite images or through images of sensors with similar characteristics, which can calculate the annual deforestation rates in the Legal Amazon. The survey involves great territorial extension in which more than 5 million square kilometers are annually mapped. The information generated is used by the Brazilian government for assessment and public policy establishments concerning illegal deforestation. Even though the historic PRODES historical series is completing 30 years, there have been no studies conducted to assess and account for the errors and uncertainties found in the mapping of all extension of the Legal Amazon. The specialized literature recommends that the surveys performed with the use of remote sensing be always followed by measurements of accuracy and/or quality. However, the methods used to evaluate these measurements are often restricted to quantifying them in a global approach without reorganizing them or spatializing them. On mappings of a large territorial scale, having as an example the one executed by PRODES, informing quality global rates is important but cannot be sufficient if not accompanied by rates that provide some spatialization error models, considering the regional contexts and spatial attributes. This work developed a consistent methodology based on a regionalized statistical model to evaluate the data generated by PRODES for the year 2014. Based on a cells 50 x 50 km resolution map, landscape ecology metrics were extracted and using data mining techniques, a deforestation patterns map was created. Based on this map, around 1,800 points were randomly selected for validation process by three independent experts. A multivariate logistic regression model was developed and three independent variables associated to the landscape ecology metrics were selected to explain the event. Applying the model to the database, it was possible to spatialize the data in cells with a higher and lower probability of finding an error in the PRODES mapping, finding that 78% of the cells presented an estimated error of less than 1% while only 0.3% presented error estimated greater than 10%. The estimated error distribution by deforestation pattern also determined that the great majority of the cells present concentrated errors in the range less than 1% independent of the pattern. This errors regionalization approach using logistic regression model was adequate and the results obtained give to the mapping of the PRODES 2014 excellent quality level and they encourage their use to other similar mappings.

Keywords: deforestation mapping, remote sensing and mapping assessment.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 2.1 - Taxas PRODES no período 1988 a 2017.	14
Figura 2.2 - Mapa final gerado pelo PRODES 2014.....	15
Figura 2.3 - Acurácia X Precisão.....	17
Figura 2.4 - Imagem e mapa de exatidão obtida da classificação.....	20
Figura 2.5 - Exemplo de classificação por árvore de decisão.	22
Figura 2.6 - Curva de Regressão Logística Univariada.....	26
Figura 2.7 - Mapa da probabilidade de ocorrência de queimadas.....	29
Figura 2.8 - Arquitetura simplificada do sistema.....	33
Figura 3.1 - Mapa da área de estudo correspondente à Amazônia Legal, contendo mosaico de imagens do satélite <i>Landsat</i> TM ano 2000 cobrindo os nove estados da Amazônia Legal.....	35
Figura 3.2 - Fluxograma resumido.	37
Figura 3.3 - Grade celular de 50 x 50 km.	41
Figura 3.4 - Exemplo de métricas de paisagem calculada para a célula realçada em vermelho.....	44
Figura 3.5 - Padrões de desmatamento utilizados no treinamento.	46
Figura 3.6 - Distribuição espacial das amostras de treinamento para classificação.....	47
Figura 3.7 - Pontos amostrais para validação sorteados na célula.	50
Figura 3.8 - Interface da ferramenta de validação.....	51
Figura 4.1 - Mapa de padrões de desmatamento em células de 50 x 50 km. ...	57
Figura 4.2 - Mapa com comparação entre as 100 classificações dos padrões de desmatamento.....	58
Figura 4.3 - Distribuição das 120 células sorteadas aleatoriamente.	59
Figura 4.4 - Diagrama <i>boxplot</i> para cada padrão de desmatamento.	60
Figura 4.5 - Célula padrão Geométrico Regular.....	62
Figura 4.6 - Distribuição dos polígonos PRODES por faixa de tamanho.	63

Figura 4.7 - Distribuição dos polígonos PRODES por quantidade e área média dos polígonos.	64
Figura 4.8 - Célula padrão Consolidado.	65
Figura 4.9 - Mapa da probabilidade de erros estimados por célula.	70
Figura 4.10 - Localização das 20 células com erro estimado maior que 5%.	71
Figura 4.11 - Proporção de células por faixa de erro estimado.	72
Figura 4.12 - Localização das células analisadas individualmente.	74
Figura 4.14 - Célula padrão Multidirecional.	76
Figura 4.15 - Célula padrão Geométrico Regular (exemplo 1).	77
Figura 4.16 - Célula padrão Geométrico Regular (exemplo 2).	78
Figura 4.17 - Célula padrão Difuso.	79
Figura 4.18 - Célula padrão Linear.	80
Figura 4.19 - Célula padrão Espinha de Peixe.	81
Figura 4.10 - Célula padrão Multidirecional.	82

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 3.1 - Principais dados utilizados neste estudo.....	36
Tabela 3.2 - Descrição dos padrões de desmatamento.....	38
Tabela 3.3 - Descrição das métricas calculadas.	42
Tabela 3.4 - Descrição das variáveis adicionadas ao estudo.....	53
Tabela 4.1 - Resumo dos erros da classificação.....	55
Tabela 4.2 - Matriz de confusão da classificação dos padrões de desmatamento.	56
Tabela 4.3 - Conformidade entre classificações por faixa.....	58
Tabela 4.4 - Análise dos erros por padrão de desmatamento.....	61
Tabela 4.5 - Modelo de regressão logística selecionado pelo método <i>stepwise forward</i>	67
Tabela 4.6 - Modelo de regressão logística pelo método <i>stepwise backward</i> ..	67
Tabela 4.7 - Modelo de regressão logística com 1 variável independente selecionado por busca exaustiva.	68
Tabela 4.8 - Modelo de regressão logística com 2 variáveis independentes selecionado por busca exaustiva.	68
Tabela 4.9 - Modelo de regressão logística com 3 variáveis independentes selecionado por busca exaustiva.	68
Tabela 4.10 - Modelo de regressão logística com 4 variáveis independentes selecionado por busca exaustiva.	69
Tabela 4.11 - Distribuição dos erros estimados por faixa de probabilidade.	73

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO.....	1
1.1 OBJETIVOS.....	7
2 FUNDAMENTAÇÃO TEÓRICA.....	11
2.1 PRODES: Histórico e relevância do monitoramento do desmatamento da Amazônia Legal	11
2.2 Métodos de avaliação de mapeamentos temáticos	15
2.3 Uso de Métricas de paisagem para mapeamento de padrões de desmatamento	20
2.4 Técnicas de Mineração de dados	22
2.5 Modelos de Regressão	24
2.6 Método para avaliação de mapeamento temático	29
2.7 Ferramenta <i>Web-Gis</i> para avaliação	31
3 METODOLOGIA	35
3.1 Área de estudo	35
3.2 Relação dos dados utilizados	36
3.3 Procedimentos metodológicos.....	36
3.3.1 Fluxograma geral da metodologia.....	36
3.3.2 Análise de métricas de paisagem e padrões de desmatamento .	38
3.3.3 Mineração de Dados	41
3.3.4 Amostras de treinamento para classificação.....	45
3.3.5 Classificação	47
3.3.6 Sorteio do painel amostral para validação	48
3.3.7 Validação dos pontos amostrais	50
3.4 Modelagem estatística.....	52
3.5 Espacialização do modelo de regressão	54
4 RESULTADOS	55
4.1 Mapa de padrões de desmatamento	55
4.2 Amostragem	59

4.3	Análises preliminares sobre os pontos validados	60
4.4	Preparação das variáveis para o modelo estatístico	65
4.5	Modelagem estatística multivariada.....	65
4.5.1	Modelo de regressão logístico	66
4.5.2	Espacialização do modelo estatístico.....	69
4.5.3	Análise das células por tipologia	73
4.5.4	Resumo das análises do modelo	82
5	CONSIDERAÇÕES FINAIS.....	83
	REFERÊNCIAS BIBLIOGRÁFICAS	86
	ANEXO A – CÓDIGO FONTE DO PROGRAMA R C5.0	96

1 INTRODUÇÃO

Amazônia Legal Brasileira é um conceito político-administrativo estabelecido originalmente pela Lei 1.806 (BRASIL, 1953), com o intuito de definir a área de atuação da Superintendência do Plano de Valorização Econômica da Amazônia (SPVEA), órgão federal substituído em 1966 pela Superintendência para o Desenvolvimento da Amazônia (SUDAM). É uma região que engloba nove estados brasileiros, pertencentes à bacia amazônica, sendo eles: Mato Grosso, Pará, Acre, Amazonas, Roraima, Rondônia, Amapá, Tocantins e Maranhão (neste caso só a parte a oeste do meridiano 44°), perfazendo uma superfície de aproximadamente 5,2 milhões de km² e correspondente a cerca de 60% do território brasileiro. A Amazônia Legal engloba parte de três, de um total de seis biomas continentais brasileiros (IBGE, 2004), sendo 84% correspondente ao bioma Amazônia, 15% ao Cerrado e 1% ao Pantanal. Apesar de sua grande extensão territorial, a região tem pouco mais de 24 milhões de habitantes, ou seja, 12,4% da população nacional e a menor densidade demográfica do país com cerca de quatro habitantes por km² (IBGE, 2017). Neste estudo esta região será referenciada como Amazônia Legal.

Esta área abriga a floresta amazônica - maior floresta tropical contínua do planeta e também a bacia amazônica, importante bacia hidrográfica com cerca de um quinto do volume total de água doce do planeta e um significativo patrimônio mineral ainda não totalmente mensurado (MMA, 2017). Além da sua rica biodiversidade, a floresta amazônica fornece importantes serviços ecossistêmicos. No Brasil, por exemplo, a floresta amazônica contém mais carbono armazenado que o total global de emissões de CO₂ pela queima de combustível fóssil em uma década (OBORN et al. 2011), desempenhando assim, um papel importante no mercado global de carbono (CHAMBERS et al. 2001; LOARIE et al. 2009; LE QUERE et al. 2009). Adicionalmente, a vegetação atua como uma eficiente bomba, reciclando a água sobre toda a extensão da floresta, exercendo um relevante papel no ciclo hidrológico que contribui para a regulação do clima regional (SPRACKLEN et al. 2012; WERTH; AVISSAR 2002). Deste modo, a floresta amazônica desempenha um importante papel no

clima local e de outras regiões, influenciando a produtividade primária e agrícola de diversos ecossistemas, a produção de energia hidroelétrica, além de processos biológicos, relacionados com a disseminação de doenças e zoonoses. A relevância geográfica e ecológica da floresta amazônica, bem como as ameaças de mudanças e suas respectivas consequências, são motivadores do interesse da sociedade em relação ao conhecimento da dinâmica do desmatamento e da degradação florestal na região (INPE, 2008).

As mudanças na cobertura florestal na Amazônia Legal decorrem do complexo processo de ocupação humana na região, que foi incipiente até o final da década de 50. A fase inicial do planejamento regional (1930-1960) correspondeu à implantação do “Estado Novo” por Getúlio Vargas, e foi muito mais discursiva do que ativa. As estratégias geopolíticas do governo que visavam à integração desse espaço regional com o restante do país, bem como a necessidade de unificar o mercado nacional, associada ao avanço da industrialização, que desencadearam a abertura da região. No governo de Juscelino Kubitschek, calcado nas estratégias “Energia e Transporte” e “Cinquenta Anos em Cinco”, ações efetivas afetaram a região pela implantação das rodovias (1958) Belém-Brasília e Brasília-Acre, dois grandes eixos que contornam as laterais da floresta. A partir desses eixos acentuou-se a migração que já se efetuava em direção a Amazônia Legal, crescendo a população regional de 1 para 5 milhões entre 1950-1960, e de modo acelerado a partir de então (BECKER, 2004).

Entre 1950 e 1960 as frentes de ocupação se intensificaram devido às primeiras medidas do Estado para o povoamento da região, a criação da Superintendência de Valorização Econômica da Amazônia (SPVEA) e a abertura da rodovia Belém-Brasília. As políticas de ocupação procuraram então combinar os empreendimentos de exploração econômica com estratégias geopolíticas. (BECKER, 2005). Mas foi somente entre 1966 e 1985 que se deu o planejamento regional efetivo da região. O Estado tomou para si a iniciativa de um novo e ordenado ciclo de ocupação da Amazônia Legal, apoiado em um projeto geopolítico que visava uma rápida modernização da sociedade e do território nacionais. A partir do final dos anos 60 e início dos anos 70, a ocupação da Amazônia Legal tornou-se prioridade nacional e o governo federal passou a

viabilizar e subsidiar a ocupação de terras para expansão pioneira, sob o slogan nacionalista de “integrar para não entregar”. Data desta época o Programa de Integração Nacional (PIN) (BRASIL, 1970), que inicia a colonização oficial da região prevendo a reserva de uma faixa de terra de 100 km ao longo de cada rodovia federal a serem implantadas nas áreas de vazio demográfico e ocupadas pelos chamados excedentes populacionais do nordeste liberados pela grande seca de 1969/1970, constituindo, desta forma, uma rede hierarquizada de núcleos urbanos (agrovilas) como base de sua organização (MAGALHÃES, 1990; BECKER, 1990).

A ocupação da região também foi percebida como prioritária, em face da possibilidade de nela se desenvolverem focos revolucionários. A tendência à consolidação do povoamento pode ser observada pelo avanço econômico significativo de algumas regiões e pelo uso intensivo da tecnologia na agroindústria especialmente no cerrado, particularmente no Mato Grosso. Ocorrem mudanças na pecuária, principalmente na porção Sudeste do Pará e no Mato Grosso, com melhorias em respeito às pastagens, aos rebanhos e à indústria de couro e de leite, mudanças estas bastante significativas em termos econômicos. A Amazônia Legal do século XXI não pode mais ser vista como mera fronteira de expansão de forças exógenas nacionais ou internacionais, pois se tornou uma região no sistema espacial nacional, com estrutura produtiva própria e múltiplos projetos de diferentes atores (BECKER, 2005).

Atualmente a produção agropecuária da região tem contribuição importante no agronegócio brasileiro e, conseqüentemente, no Produto Interno Bruto (PIB) regional e nacional (IBGE, 2015). Segundo os dados mais recentes produzidos no âmbito do projeto TerraClass (ALMEIDA et al. 2016) do Instituto Nacional de Pesquisas Espaciais (INPE) em 2014, dos aproximadamente 764.000 km² já apontados pelo PRODES como desmatados na Amazônia Legal, 63% eram ocupados por pastagens e 6% por agricultura (INPE, 2017). Entre 1990 a 2016, a área colhida de soja na região aumentou de 1,6 milhão para 10,6 milhões de hectares, um crescimento de 562%. No mesmo período, a área colhida de milho cresceu de 670 mil hectares para 4,3 milhões de hectares (IBGE, 2016). Com relação ao rebanho bovino, segundo dados do último censo agropecuário

do IBGE, a produção saltou de sete milhões para 70 milhões de cabeças entre anos de 1975 e 2006 – apontando uma variação de 900%, enquanto a área empregada nas pastagens passou de 20,3 milhões de hectares em 1975 para 61,6 milhões hectares em 2006 - variação de 203%. Nota-se esta mesma variação quando se analisa a taxa de lotação das pastagens que aumentou de 0,30 cabeças/hectares em 1975 para 0,91 cabeças/hectares em 2006 (IBGE, 2006; VALENTIN; ANDRADE, 2009).

Em contraponto a este cenário, para Vieira et al. (2014) historicamente a Amazônia Legal vem sendo alvo de ações visando integrar o território e levar à região o desenvolvimento econômico, como a construção de rodovias, ferrovias e hidroelétricas, sem contemplar a questão ambiental e social na mesma medida. A estrutura espacial da Amazônia Legal de hoje está relacionada com o período de sua ocupação, traduzida pela forte desigualdade na apropriação dos recursos naturais para produção agrícola e pecuária. As distorções nas formas do uso da terra levaram a uma exclusão social das diversas categorias socioambientais pelo setor produtivo mais capitalizado. Os autores identificaram a necessidade de desenvolvimento de estratégias que garantam o envolvimento dos atores sociais locais e seu comprometimento com a sustentabilidade socioeconômica e ambiental da região, estruturando o segmento produtivo rural com níveis aceitáveis de sustentabilidade ecológico-econômica, caso se queira diminuir o impacto ambiental.

A implementação de ações e estratégias geopolíticas para ocupação humana na Amazônia Legal teve como consequência o aumento de impactos sociais, como o conflito por terras, e ambientais, como o desmatamento, responsável por grande parte das alterações da paisagem na região. Devido às altas taxas de desmatamento, estimadas desde o final da década de 80 pelo INPE, e a preocupação em escala global com o impacto do desmatamento na emissão de gases de efeito estufa, tem-se início a criação de políticas públicas e de ações para a contenção e controle do desmatamento.

No final da década de 70 têm início, no INPE, os primeiros estudos visando demonstrar a viabilidade do uso da tecnologia do sensoriamento remoto orbital

no mapeamento e quantificação do desmatamento na região amazônica (TARDIN et al. 1978; TARDIN et al. 1979). Neste período, a experiência acumulada nestes estudos permitiu ao INPE responder a uma demanda da SUDAM que necessitava de indicadores sobre a eficácia da política de ocupação da Amazônia Legal induzida por incentivos fiscais concedidos a região (VALERIANO et al. 2012). Porém, apenas em 1988 inicia-se de forma sistemática o cálculo das estimativas anuais de taxa de desmatamento por corte raso, pelo Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite (PRODES), entendendo corte raso como áreas onde houve supressão total da floresta primária, com exposição quase total do solo, independente da futura utilização da área. Concebido inicialmente para ser somente um inventário anual de perda de floresta primária, o PRODES evoluiu e a partir de 2002 passou a disponibilizar de forma pública na Internet, além das taxas de desmatamento, dados espaciais digitais em vários formatos, dando maior transparência a esses dados (INPE, 2017).

O PRODES utiliza imagens de satélites de sensoriamento remoto para observação da terra da classe *Landsat* (20 a 30 metros de resolução espacial e taxa de revisita de 16 dias) numa combinação que busca minimizar o problema da cobertura de nuvens e garantir critérios de interoperabilidade entre sensores de diferentes satélites. As imagens do sensor *Thematic Mapper* (TM), do satélite americano *Landsat-5*, foram, historicamente as mais utilizadas pelo projeto, sendo substituídas pelos dados do sensor *Operational Land Imager* (OLI) a bordo do satélite *Landsat-8* após 2013. As imagens do satélite CBERS-2, 2B e 4 do programa sino-brasileiro de sensoriamento remoto também foram bastante utilizadas, além das imagens do satélite indiano *Resourcesat 1* e 2 e do satélite inglês UK-DMC2. Com essas imagens, a área mínima mapeada pelo PRODES é de 6,25 hectares (INPE, 2017).

Atualmente o PRODES emprega como base tecnológica o sistema de informações geográficas TerraAmazon. Este sistema é construído baseado na biblioteca de classes e funções de sistema de informação geográfica TerraLib - para desenvolvimento de aplicações geográficas desenvolvidas pelo INPE e seus parceiros. Esta biblioteca esta disponível na internet na forma de código

aberto permitindo um ambiente colaborativo para o desenvolvimento de várias ferramentas de SIG.

Desde 2004 o PRODES está inserido no Plano de Ação para a Prevenção e Controle do Desmatamento na Amazônia Legal (PPCDAm), coordenado pelo Ministério do Meio Ambiente (MMA) (MMA, 2017), sendo um importante instrumento para o planejamento de políticas e ações de controle e combate ao desmatamento. Os dados produzidos pelo PRODES também tem sido utilizado com sucesso nos acordos voluntários de cadeias produtivas do agronegócio com o compromisso de combater o desmatamento na Amazônia Legal, tais como a Moratória da Soja (RUDORFF et al. 2011) e o Termo de Ajustamento de Conduta da pecuária (TAC da Carne) (NEPSTAD et al. 2014). Além de sua importância para políticas públicas, seus dados têm sido utilizados como referência nas negociações internacionais de mudanças climáticas, como a Convenção Quadro das Nações Unidas sobre a Mudança Climática (UNFCCC), e também nos modelos de emissão de carbono reportados nos Relatórios de Inventário Nacional de Emissões de Gases de Efeito Estufa da UNFCCC. Também em apoio ao PPCDAm e em atendimento às necessidades dos órgãos ambientais que têm atividades ligadas à fiscalização e ao combate ao desmatamento, o INPE desenvolveu e mantém em operação desde 2004 um Sistema de Detecção de Desmatamento em Tempo Real (DETER), que tem como objetivo emitir alertas diários de desmatamento e de alterações na cobertura florestal com imagens do sensor MODIS (Terra), WFI (CBERS) e AWiFS (*Resourcesat*) (DINIZ et al. 2015; FINER et al. 2018).

Apesar da série histórica do PRODES ter mais de 25 anos e ser amplamente utilizada no meio científico, nas convenções de mudança climática, nos modelos de emissão de carbono e nos programas de combate ao desmatamento, até hoje não se realizaram estudos de larga escala para aferir e contabilizar os erros e incertezas encontrados neste mapeamento. O mapeamento não contempla uma metodologia de aferição de erros inerentes ao processo de produção dos seus dados. Os erros podem ser oriundos de diversas fontes, como imprecisão na definição das classes e na delimitação dos alvos e podem ser introduzidos facilmente no processo de mapeamento

temático com imagens de sensoriamento remoto. A incorporação de medidas para o rastreamento de erros possibilita gerar informações mais exatas e que conferem maior credibilidade ao mapeamento proposto. Mapas oriundos de dados de sensoriamento remoto que não tem uma estimativa de exatidão são de pouca utilidade para realizar inferências estatísticas (McROBERTS, 2011). Por essa razão, medidas de qualidade em mapeamentos temáticos obtidos com imagens de sensoriamento remoto devem ser encorajadas e realizadas para que se possibilite determinar os níveis de confiança destes mapeamentos.

Visando suprir essa informação para o sistema de monitoramento da cobertura florestal do INPE, foi desenvolvida uma metodologia para estimar o nível de confiança do mapeamento produzido pelo PRODES para a Amazônia Legal, observando os locais onde ocorrem os erros, bem como os fatores relacionados a eles. Como exemplo, foi utilizado o dado de desmatamento do ano de 2014 e um modelo estatístico multivariado que possibilitou apontar os principais fatores e regiões potenciais de ocorrência de erros. Espera-se com os resultados produzidos neste estudo, agregar informações de incerteza aos dados de desmatamento, além de fornecer informações à equipe operacional que realiza o mapeamento do PRODES, sobre os fatores que influenciam a qualidade do mapeamento, bem como, indicar as potenciais regiões de ocorrência de erros, de forma a preveni-los, melhorando a confiabilidade dos mapas gerados.

1.1 Objetivos

O objetivo principal deste estudo foi estabelecer uma metodologia consistente e replicável que permita calcular estimativas confiáveis dos níveis de qualidade do mapeamento produzido pelo PRODES para a Amazônia Legal para o ano de 2014. Resumidamente, esse estudo foi conduzido a partir de duas perguntas principais:

1. Onde a probabilidade de se verificar erros é maior/menor no mapeamento do PRODES?

2. Quais fatores contribuem para o aumento do erro no mapeamento do PRODES?

Para responder a essas perguntas, partiu-se da suposição de que o padrão de desmatamento influencia na acurácia do mapeamento do desmatamento. Desta forma, áreas que apresentam padrões mais complexos e heterogêneos em relação ao tamanho e a forma das áreas desmatadas, como as do padrão espinha de peixe, relacionado principalmente com os projetos de assentamentos do INCRA, tenderiam a apresentar maiores probabilidades de erro do que as áreas em que predominam polígonos de desmatamento grandes, uniformes e geométricos, que representam as grandes fazendas da Amazônia Legal. Sendo delineados quatro objetivos específicos:

1. Regionalizar a área de estudo baseado em padrões de desmatamento representados em células de 50 X 50 km, a partir de uma tipologia de padrões de desmatamento estabelecida por Saito et al. (2011) com técnicas de mineração de dados;
2. Estabelecer um esquema amostral para avaliação regionalizada da confiabilidade do desmatamento na Amazônia Legal, considerando o mapa de padrões de desmatamento e levantamentos pré-existentes;
3. Desenvolver um modelo estatístico multivariado que possibilite explicar as principais fontes de erro do mapeamento do desmatamento de 2014;
4. Utilizar o modelo estatístico ajustado para gerar uma superfície contínua que aponte regiões com maiores e menores potenciais de erro para a região da Amazônia Legal.

A abordagem é baseada no estabelecimento de um modelo estatístico multivariado envolvendo regressão logística e análise espacial visando espacializar/regionalizar os erros e buscando explicar a relação com suas possíveis fontes. Fatores como classes de tamanho e forma dos polígonos, grau de fragmentação das áreas no entorno do desmatamento, entre outras, foram estimados e analisados. Para isso, uma abordagem que utiliza o espaço celular como unidade de área de análise, bem como métricas de paisagem, foi empregada para reconhecimento de padrões de desmatamento encontrados nos

dados do PRODES de 2014. Ao se estabelecer um modelo operacional confiável de espacialização dos erros, este poderia ser reaplicado em mapeamentos futuros, ou mesmo nos pretéritos, permitindo assim associar os dados produzidos pelo PRODES a níveis de qualidade, garantindo parâmetros de confiabilidade mínimos para os mapas gerados.

Como já mencionado, um dos benefícios derivados deste estudo é a possibilidade de dar subsídios à equipe operacional que elabora anualmente o mapeamento do PRODES, indicando áreas potenciais de ocorrência de erros, de modo que a equipe possa destinar especial atenção a esses locais, seja através de treinamento específico ao grupo de intérpretes para caracterização do desmatamento nestas regiões, seja na maior atenção do grupo de auditores dada a essas áreas quando o controle de qualidade do mapeamento é realizado. Este estudo também pode trazer benefícios adicionais significativos para os órgãos do governo brasileiro que tem como função a fiscalização, prevenção e combate ao desmatamento, tanto no âmbito federal, como por exemplo, o Instituto Brasileiro do Meio Ambiente e Recursos Naturais Renováveis (IBAMA) e/ou Instituto Chico Mendes de Conservação da Biodiversidade (ICMbio), como no âmbito estadual, como os Órgãos Estaduais do Meio Ambiente (OEMAs). Os agentes da fiscalização dispõem não apenas da informação sobre a ocorrência do desmatamento, mas também da informação, nos locais indicados, das áreas mais suscetíveis a erros de mapeamento do PRODES no planejamento e execução das suas políticas fiscalizatórias, possibilitando priorizar áreas que apresentam maiores chances de acerto, o que implica na redução de custos operacionais relativos ao deslocamento de aeronaves e equipes de campo em locais em que a probabilidade de erro é alta.

Em adição, os modelos de emissões de carbono por mudança na cobertura da terra e que se baseiam nos dados gerados pelo PRODES, com o acréscimo dessa informação, poderão ser calibrados de forma mais adequada e serem mais informativos com relação as suas incertezas, ao incluírem as informações sobre os potenciais erros espacializados no mapeamento.

2 FUNDAMENTAÇÃO TEÓRICA

Nas próximas seções serão apresentados e discutidos o histórico do sistema de monitoramento do desmatamento do INPE e sua relevância, bem como métodos e técnicas de avaliação da exatidão em mapeamento temático com uso de dados de sensoriamento remoto orbital. Em seguida, são apresentadas técnicas de espacialização de eventos pontuais e as métricas de paisagem utilizadas no mapeamento de padrões de desmatamento. No final deste capítulo, é apresentada uma seção sobre modelos de regressão com ênfase em regressão logística.

2.1 PRODES: Histórico e relevância do monitoramento do desmatamento da Amazônia Legal

A década de 70 data o início dos primeiros levantamentos do desmatamento na Amazônia Legal realizados por Tardin et al. (1978) e Tardin et al. (1979), neste último os autores mostraram ser viável o mapeamento do desmatamento com uso de imagens do satélite *Landsat*. No estudo foram analisadas 31 cenas do sensor *Multispectral Scanner System* (MSS) adquiridas em 1977 e 1978 que envolviam partes dos estados do Mato Grosso, Pará e Amazonas, que totalizavam uma área de 552.000 km². O total de desmatamento calculado à época foi de 41.000 km² e mencionado pelos autores como uma ameaça ao equilíbrio ecológico da região.

Os levantamentos não tiveram continuidade e durante a década de 80 a velocidade do desmatamento na Amazônia Legal passou a ser vista como um motivo de preocupação pelo governo federal devido à relevância deste processo no ciclo de carbono global tema de interesse crescente pela comunidade internacional. Para responder a esta pressão internacional e, em preparação para a Conferência ECO-92, em 1988 o governo brasileiro solicitou ao INPE desenvolver e operar um sistema de monitoramento anual do desmatamento da Amazônia Legal (VALERIANO et al. 2012). Desde então o INPE realiza o inventário anual de perda de área de cobertura florestal primária na Amazônia Legal e estima sua taxa anual de desmatamento, através do

mapeamento do desmatamento por corte raso com uso de imagens de satélites da classe *Landsat* (multiespectrais, com resolução de 20 a 60 m e taxa de revisita ao menos quinzenal).

A taxa anual de desmatamento tem sido um parâmetro fundamental do governo federal e/ou governos estaduais para a proposição de políticas públicas de gestão de terras na região e também para a avaliação da efetividade de suas implementações. Inúmeras publicações científicas utilizam os dados do PRODES e a comunidade acadêmica considera este sistema uma eficiente tecnologia espacial aplicada à medição e ao controle do desmatamento (KINTISCH, 2007). Foi com base nos dados do PRODES que o governo brasileiro submeteu com sucesso em 2014 a primeira proposta de Nível de Referência Florestal (FREL) para a UNFCCC para fins de remuneração por redução de emissões de gases de efeito estufa, por desmatamento e degradação florestal evitado (REDD) (MMA, 2017a). O PRODES também é referência para os dados de atividade de desmatamento na Amazônia Legal reportados nos Relatórios de Inventário Nacional de Emissões de Gases de Efeito Estufa que o governo brasileiro submeteu à UNFCCC cumprindo o compromisso firmado nesta Convenção.

Outro fator importante relacionado aos dados PRODES é o seu uso na implementação de acordos voluntários para controle das cadeias produtivas do agronegócio com o compromisso de combater o desmatamento na Amazônia Legal, bem como, nos acordos intergovernamentais a exemplo da Conferência das Nações Unidas sobre Mudanças Climáticas (COP 21). Desta maneira, este mapeamento contribui para a governança e desenvolvimento de políticas públicas estratégicas que buscam o desenvolvimento e a produção sustentável na região.

O histórico da confecção dos mapas do PRODES pode ser dividido em três fases distintas. O período entre 1988 e 2000, que correspondeu ao chamado PRODES Analógico, no qual eram utilizadas aproximadamente 220 imagens do satélite *Landsat-5 TM* por ano, impressas na composição colorida associando a banda 5 à cor vermelha (*Red*), a banda 4 à cor verde (*Green*) e banda 3 à cor

azul (*Blue*) em papel fotográfico na escala 1:250.000 e interpretadas visualmente e manualmente por uma equipe especializada. Sobre estas imagens eram sobrepostos overlays que recebiam o desenho do polígono associado ao desmatamento. Posteriormente estes overlays eram digitalizados em um sistema de informações geográficas para o cálculo da área e da taxa de desmatamento correspondente.

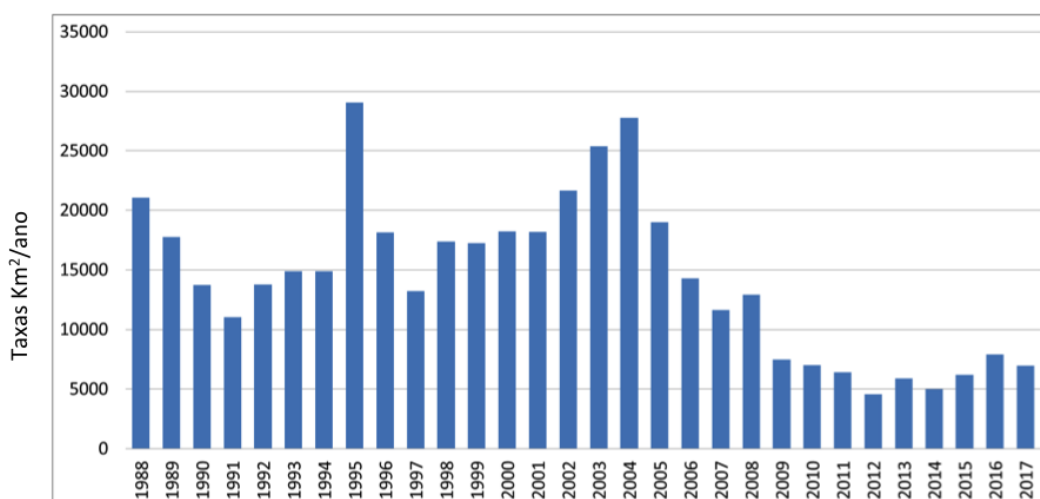
O período entre 2000 e 2005 ficou conhecido como PRODES Digital, onde o mapeamento passou a ser realizado utilizando o sistema SPRING (CÂMARA et al., 1996) e processamentos semi-automatizados (SHIMABUKURO et al., 1998) de classificação e edição. Além disto, em 2002 ocorreu a disponibilização de todos os dados na Internet, estabelecendo um marco, pois tornou o processo de geração e divulgação do dado transparente, permitindo que sejam avaliados tanto pela comunidade científica, quanto por qualquer outro setor da sociedade. Desde 2004, o PRODES está inserido como ação do Ministério de Ciência, Tecnologia, Inovação e Comunicações (MCTIC) no Grupo Permanente de Trabalho Interministerial (GPTI) para a redução dos índices de desmatamento da Amazônia legal. O GTPI é parte do Plano de Ação para a Prevenção e Controle do Desmatamento na Amazônia Legal (PPCDAm), atualmente coordenado pelo Ministério do Meio Ambiente (MMA) (MMA, 2017b).

Em 2005 o PRODES começou a ser realizado utilizando uma plataforma multiusuário denominada de TerraAmazon, desenvolvida sobre a biblioteca de classes e funções TerraLib (CÂMARA et al., 2008). O TerraAmazon é um sistema de informações geográficas projetado para ser um editor multiusuário de dados geográficos vetoriais. Possui ferramentas de classificação de uso e cobertura do solo, assim como operações espaciais entre dados vetoriais. Isto permite que se façam análises de transição, entre outras aplicações (TERRAAMAZON, 2017). Nesta nova plataforma, o processo voltou a ser realizado somente por meio de fotointerpretação diretamente na tela do computador. A adoção do critério de fotointerpretação foi realizada para garantir maior exatidão da classificação e velocidade na classificação das

imagens, o que possibilitou que a divulgação dos dados ocorresse ainda durante o ano da detecção.

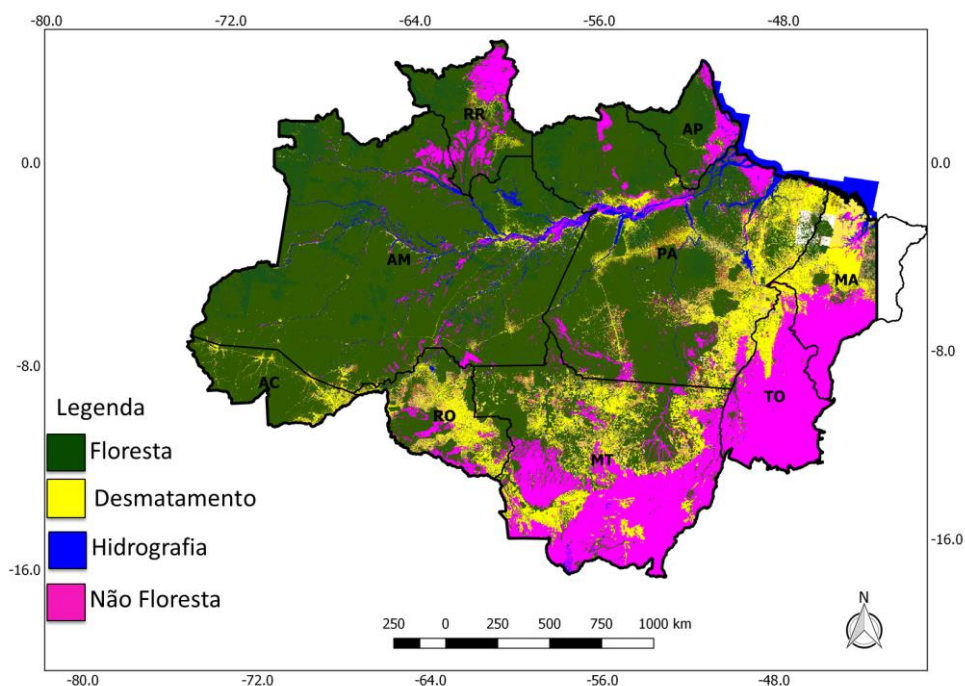
Durante as três fases acima, duas premissas foram mantidas integralmente a fim de garantir manutenção de uma série histórica compatível e comparável ao longo de todos os anos: a) área mínima de mapeamento igual ou superior a 6,25 hectares e; b) mapeamento exclusivamente de desmatamento tipo corte raso, ou seja, supressão completa da cobertura florestal, sem incluir derrubadas parciais da floresta resultantes de degradação florestal, queimadas e extração seletiva de madeira. Destaca-se também que a metodologia do PRODES utiliza o conceito de máscara acumulativa, agregando o desmatamento mapeado nos anos anteriores em uma base integrada contendo todas as áreas de desmatamento por corte raso já detectadas. A máscara do PRODES é usada para eliminar a possibilidade de que desmatamentos antigos já detectados sejam identificados e mapeados novamente (INPE, 2008). A Figura 2.1 apresenta a evolução da série histórica das taxas de desmatamento produzidas pelo PRODES desde 1988 e a Figura 2.2 apresenta o mapa do desmatamento consolidado gerado pelo PRODES em 2014 para toda Amazônia Legal.

Figura 2.1 - Taxas PRODES no período 1988 a 2017.



Fonte: INPE (2017).

Figura 2.2 - Mapa final gerado pelo PRODES 2014.



Fonte: Produção do autor.

2.2 Métodos de avaliação de mapeamentos temáticos

Para lidar com problemas de recursos naturais, produtos gerados a partir de imagens de satélite tornaram-se uma fonte bem aceita de informações contribuindo para a construção de mapas temáticos de cobertura terrestre. A utilidade desses mapas depende de sua exatidão e da ausência de erros sistemáticos (McROBERTS, 2011). O sensoriamento remoto é uma fonte potencial para obtenção de mapas temáticos, como aqueles que descrevem a cobertura da terra, pois fornece uma representação similar à do mapa da superfície da Terra, bem como, disponível em uma variedade de escalas espaciais e temporais. O mapeamento temático a partir de dados de detecção remota normalmente é baseado em uma classificação ou interpretação de imagem de satélite onde a imagem classificada resultante pode ser tratada como um mapa temático que representa a cobertura da terra da região (FOODY, 2001).

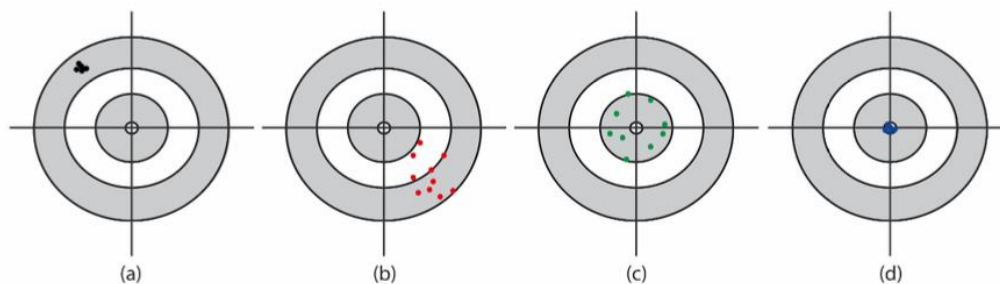
Uma preocupação fundamental é que os mapas de cobertura terrestre derivados são muitas vezes julgados como de qualidade insuficiente para aplicações operacionais. Este julgamento baseia-se tipicamente em uma avaliação do mapa de cobertura de terra derivado do processo de classificação em relação a um determinado conjunto de dados de referência. Os desacordos entre os dois conjuntos de dados são tipicamente interpretados como erros no mapa temático derivado dos dados de sensoriamento remoto (CONGALTON, 1991; SMEDES, 1975). No entanto, existem muitas incertezas associadas ao significado e à interpretação da qualidade do mapa que a tornam uma variável difícil de considerar objetivamente e que limitam substancialmente a capacidade de avaliar o grau em que o potencial do sensoriamento remoto como fonte de dados de cobertura de terra está sendo realizado (FOODY, 2001).

Jensen (2009) aponta que erros são normalmente introduzidos no processo de mapeamento temático utilizando sensoriamento remoto e destaca quatro ações para sua redução: a) registro dos passos operacionais utilizados, b) documentação dos erros geométricos e temáticos, c) melhoria na legenda e d) uso de métodos de avaliação de exatidão. Card (1982) adverte que usuários não deveriam usar ou considerar um mapa sem alguma estimativa de erro associada a ele. Se os erros e/ou a exatidão não foram avaliados e relatados, independentemente do motivo, os mapas podem ter pouca ou nenhuma utilidade para inferência científica (McROBERTS, 2011).

Para a FAO (2016), a incerteza inerente a mapeamentos temáticos, pode ser resumida como a falta de conhecimento do valor verdadeiro e está relacionada a dois problemas: precisão e acurácia. A acurácia é uma medida relativa da exatidão de uma estimativa e explica erros sistemáticos também referidos como viés. Portanto, uma estimativa de acurácia não superestima ou subestima sistematicamente o valor verdadeiro. A acurácia do mapa pode ser quantificada criando uma matriz de confusão, que compara a classificação do mapa a ser avaliado com uma classificação de referência. A precisão está relacionada ao erro aleatório que pode ser quantificado por um intervalo de

confiança. A Figura 2.3 exemplifica a diferença básica entre estes dois conceitos.

Figura 2.3 - Acurácia X Precisão.



(a) sem acurácia, mas preciso; (b) sem acurácia e impreciso; (c) acurado, mas impreciso; e (d) preciso e acurado.

Fonte: FAO (2016).

Segundo Foody (2001), a matriz de confusão está atualmente no centro da literatura de avaliação de precisão, onde uma simples tabulação cruzada entre a classe mapeada em relação à observada fornece uma base para avaliação de precisão. Congalton (1991) também aponta que uma das técnicas mais utilizadas na realização de avaliação de exatidão da classificação de dados de sensoriamento remoto é o emprego da matriz de erro, também conhecida como matriz de confusão. A matriz de confusão é uma matriz quadrada de números definidos em linhas e colunas que expressam o número de unidades da amostra (pixels, grupos de pixels ou polígonos) que foi atribuído a uma classe em particular, confrontada com classes de referência, geralmente verificadas em campo. Normalmente as colunas representam os dados de referência, enquanto as linhas representam a classificação gerada a partir dos dados de sensoriamento remoto (SUAREZ, 2012).

A partir da matriz de confusão são estimados índices globais de exatidão, que também podem ser expressos em erros de omissão e inclusão. A exatidão das categorias individuais pode ser obtida de forma similar, dividindo o número de

amostras corretas em dada categoria, pelo número total de amostras da coluna correspondente. Quando o número total de amostras corretamente classificadas de uma categoria é dividido pelo número total de amostras desta categoria, proveniente dos dados de referência (total da coluna), esta medida de exatidão estima a probabilidade de uma amostra ser corretamente classificada nesta categoria e é denominada de exatidão do produtor (seu valor complementar corresponde ao erro de omissão). Por outro lado, se o número total de amostras corretamente classificadas em uma categoria é dividido pelo número total de amostras classificadas nessa categoria, então este resultado representa a exatidão do consumidor, sendo o seu complementar uma medida de erro de inclusão (SUAREZ, 2012).

Segundo Congalton (2001), a matriz de erro permite o uso de técnicas de estatística analítica como as técnicas discretas multivariadas e tem sido utilizada na realização de testes estatísticos para estimar a exatidão de dados obtidos com produtos de sensoriamento remoto. O índice Kappa é um exemplo de um método muito utilizado para avaliar a concordância entre a verdade terrestre e o mapa temático e expressa à medida da diferença entre a concordância dos dados de referência e a classificação (FERREIRA, 2007).

Apesar de algumas restrições e críticas, a avaliação da qualidade de mapeamentos temáticos obtidos através de imagens de sensoriamento remoto, baseando-se em índices Kappa, tem sido amplamente utilizada e aceita pela comunidade da área nas últimas décadas. Pontius e Millones (2011) citam que o uso do índice Kappa se tornou parte da cultura do sensoriamento remoto e de outras áreas da ciência, mas apontam várias limitações e fazem severas críticas ao seu uso para a avaliação da exatidão de mapeamentos obtidos com imagens de satélite. Eles examinaram cinco diferentes índices Kappa explorando suas propriedades matemáticas e ilustrando suas limitações. Para os autores, o índice Kappa é complicado para calcular, difícil de entender, inútil para interpretar e decretaram a “morte do Kappa”, por entenderem que estes índices não são adequados para as aplicações práticas em sensoriamento remoto. Nos casos estudados por eles, o Kappa forneceu informações redundantes ou enganosas para a tomada de decisões práticas. Como

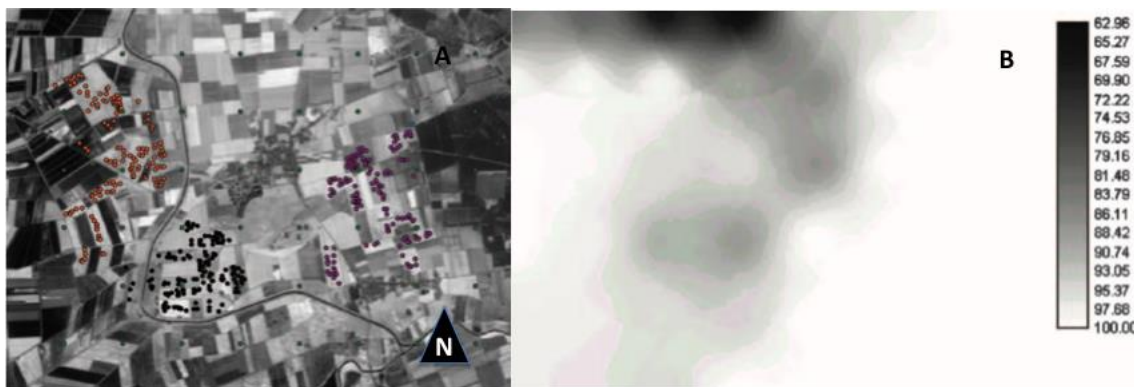
alternativa, os autores recomendam resumir a matriz de confusão em dois componentes mutuamente exclusivos mais simples, denominados diferença de quantidade e diferença de alocação. Maiores detalhes sobre estes índices podem ser encontrados em Pontius e Millones, (2011) e Pontius e Santacruz (2014).

Além dos índices citados nesta seção, existem outras formas e métodos que são usados para medir a qualidade de um mapeamento feito a partir de dados de sensoriamento remoto, porém, quase todos fornecem índices globais de exatidão, sem regionalizá-los ou espacializá-los. Fonte e Gonçalves (2015) afirmam que, como a matriz de confusão é construída para todo o mapa, seus índices se aplicam à área como um todo e, embora possam haver diferentes níveis de exatidão em diferentes regiões do mapa, essa variabilidade não pode ser capturada com esta metodologia. Foody (2005) indica que a exatidão em mapeamentos derivados de dados de sensoriamento remoto são medidas globais e fornecem informação simplificada para ser aplicada uniformemente em toda região representada. O autor propôs a construção de matrizes de confusão para regiões geograficamente delimitadas, ou seja, calculadas para partes da imagem, em vez de apenas uma matriz de confusão para toda a imagem. Esta abordagem foi testada em uma classificação para áreas de culturas agrícolas localizadas no sudeste da Inglaterra utilizando-se imagens do sensor de alta resolução *High Resolution Visible* (HRV) a bordo do satélite francês SPOT. A exatidão global da classificação foi estimada em 84%, mas variou localmente de 53,33% a 100%.

A Figura 2.4(A) apresenta a área de teste utilizada por Foody (2005) com os 48 pontos utilizados para validação da classificação. Já a Figura 2.4(B) apresenta em um mapa de superfície a variação espacial da exatidão calculada localmente, interpolada pelo método do inverso do quadrado da distância. Com base nesta análise, o autor conclui que grande parte da imagem classificada pode ser vista como satisfazendo o nível de exatidão alcançado, no caso 84%, mas sua interpretação estrita exigiria que toda a classificação de imagem fosse rejeitada, uma vez que o limiar requerido, estabelecido em 85%, não foi alcançado. Cabe destacar que a parte norte da imagem a ser classificada não

recebeu quantidade suficiente de pontos amostrais para o teste de classificação, fato que pode explicar a baixa exatidão do mapa nesta área.

Figura 2.4 - Imagem e mapa de exatidão obtida da classificação.



A) Imagem com 1000 pontos amostrais para teste da classificação separados em três locais. B) Mapa interpolado especializando a exatidão da classificação em %.

Fonte: Foody (2005).

Com o conhecimento prévio de onde são esperados diferentes níveis de exatidão é possível estabelecer matrizes de confusão geograficamente restritas, que fornecem informações sobre a distribuição espacial da exatidão de classificação (FONTE; GONÇALVES, 2015). Em mapeamentos de larga escala territorial, a exemplo do executado pelo PRODES, prover índices globais de exatidão são importantes e sua construção deve sempre ser almejada, mas estes índices podem não ser suficientes se não vierem acompanhados de indicadores que forneçam algum modelo de espacialização geográfica da exatidão, levando em consideração características locais e/ou padrões de desmatamento.

2.3 Uso de métricas de paisagem para mapeamento de padrões de desmatamento

A quantificação da estrutura da paisagem, por meio de métricas da paisagem, é pré-requisito para a avaliação do funcionamento e das mudanças que

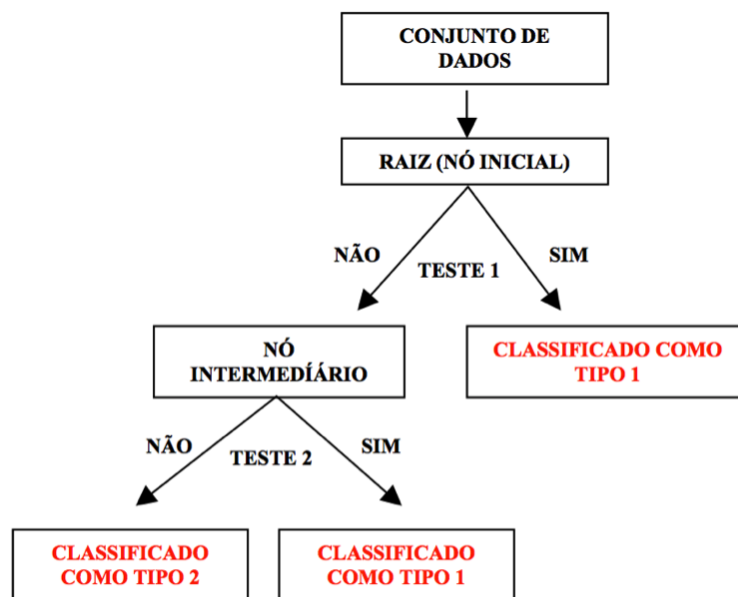
ocorrem nessa unidade de análise. As métricas são úteis em várias situações: permitem descrever e mensurar os elementos que compõem a estrutura da paisagem, comparar paisagens, identificar mudanças temporais, avaliar cenários propostos durante o planejamento e relacionar padrões a funções ecológicas, entre outros. Quando analisadas em função do seu significado ecológico, podem fornecer informações importantes ao planejamento, à conservação e à preservação dos recursos naturais. Devido à grande aplicabilidade e importância, uma variedade de métricas tem sido desenvolvida, tais como: área, de densidade, tamanho e variabilidade das manchas, bordas, formas, área nuclear, de vizinho mais próximo, diversidade, de contágio e espalhamento (MACGARIGAL, 1995; RUDOLPHO et al. 2013). Para a análise da paisagem utilizando os dados de desmatamento pode-se trabalhar com objetos individuais, representados por cada um dos polígonos de desmatamento (SILVA et al. 2008) ou com células onde cada uma é representada por um conjunto de polígonos de desmatamento.

Na análise baseada em células a área de estudo é subdividida em pequenas regiões regulares, em que, cada célula representa uma porção da paisagem. Assim, cada célula da grade é associada a um padrão de desmatamento, onde cada padrão é descrito por um conjunto de métricas da paisagem (SAITO et al., 2011). A tipologia estabelecida pelos autores baseia-se em Geist e Lambin (2001) e foi ampliada e adaptada para a floresta Amazônica, sendo composta por oito classes e é apresentada no Capítulo 2. Silva et al (2008) utilizaram métricas da paisagem para analisar padrões de polígonos de desmatamento em duas regiões da Amazônia Legal a partir da construção de tipologias de padrões de desmatamento associado a diferentes tipos de atores (pequenos, médios e grandes estabelecimentos rurais) e técnicas de mineração de dados para a classificação desses polígonos. Essa técnica será descrita na seção seguinte e alguns trabalhos que procuraram associar padrões de uso e cobertura da terra com diferentes processos são utilizados para exemplificar seu uso.

2.4 Técnicas de Mineração de dados

Normalmente as técnicas de mineração de dados utilizam o conceito de árvore de decisão. Classificação por árvore de decisão divide um conjunto de dados em subconjuntos menores onde cada nó representa uma decisão. Com base em valores característicos as árvores de decisão classificam as instâncias. Cada nó representa um recurso em uma instância em uma árvore de decisão que deve ser classificada e cada ramo representa um valor. Cada nó interno de uma árvore de decisão contém um teste, cujo resultado é usado para decidir qual ramificação seguir desse nó. A classificação das instâncias é iniciada a partir do nó raiz e ordenada de acordo com os valores característicos (QUINLAN, 1993; PANDYA; PANDYA, 2015).

Figura 2.5 - Exemplo de classificação por árvore de decisão.



Fonte: Tsuchiya (2002).

A versão C5.0 é uma nova geração de *Machine Learning Algorithms* (MLAs) baseado em árvores de decisão. O algoritmo C5.0 é uma atualização da versão anterior – C4.5 e tem várias vantagens importantes sobre esta última. As regras geradas são mais precisas e o tempo usado para gerá-las é menor. Na

versão C5.0 novas técnicas foram introduzidas: a) Boosting: várias árvores de decisão são geradas e combinadas para melhorar as previsões; b) custos variáveis de erros na classificação: torna possível evitar erros; c) capacidade de lidar com novos atributos: datas, horários, timestamps, atributos ordenados discretos; d) os valores podem ser marcados como ausentes ou não aplicáveis para casos particulares e e) suporta amostragem e validação cruzada (BUJLOW et al. 2012). Uma descrição detalhada da versão C5.0 pode ser encontrada em (QUINLAN, 2017).

Saito (2010) empregou este tipo de abordagem utilizando como unidade de análise um ambiente de espaço celular. No estudo foi desenvolvida uma tipologia baseada nos dados de desmatamento gerados pelo PRODES, em que os padrões de desmatamento foram relacionados com diferentes formas de ocupação humana e usos da terra. Foram utilizadas métricas da paisagem considerando o conjunto de polígonos presente nas células e técnicas de mineração de dados para classificação das células por meio de um algoritmo de árvore decisão, o qual possibilitou classificar os padrões de desmatamento associados aos diferentes tipos e trajetórias de padrões de ocupação humana. Silva et al. (2005) propuseram uma metodologia para definição de uma tipologia de padrões de desmatamento a partir da coleta de um conjunto de amostras de referência, utilização de métricas da ecologia da paisagem para avaliar as propriedades de forma e arranjo espacial dos polígonos de desmatamento e técnicas de mineração de dados utilizando um algoritmo de classificação por árvore de decisão. Já Gavlak et al. (2011) analisaram as relações entre a dinâmica da cobertura da terra em regiões com diferentes estágios de ocupação da fronteira agropecuária no Distrito Florestal Sustentável da BR-163 com base nos dados de desmatamento PRODES para anos 1997 a 2007 e vegetação secundária referente anos 2000 e 2008 integrados em ambiente espaço celular. A identificação dos padrões de desmatamento representados por células foi realizada por meio de um conjunto de métricas da ecologia da paisagem e posterior classificação por árvore de decisão C 4.5.

2.5 Modelos de Regressão

Modelos de regressão são modelos matemáticos que relacionam o comportamento de uma variável Y com outra X através de uma função. Quando esta função é do tipo $Y = f(X) = \beta_0 + \beta_1 X$, tem-se o modelo de regressão linear simples, onde β_0 e β_1 são respectivamente o coeficiente linear e angular da reta que representa esta função. A variável X é a variável independente da equação, enquanto Y é a variável dependente. O modelo de regressão é chamado de simples quando envolve apenas uma variável independente. O modelo de regressão é multivariado quando envolve uma relação com duas ou mais variáveis independentes, isto é, quando o comportamento de Y é explicado por mais de uma variável independente (NETER et al., 1996). Como esta relação nem sempre é perfeita, admite-se que há um erro associado a cada valor estimado pela função ajustada obtida a partir de um conjunto de amostras. Num caso multivariado, onde m variáveis independentes são consideradas, $Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_m + e$ o último termo e representa o erro ou desvio dos valores estimados pela equação em relação aos verdadeiros valores de Y . Neste caso, b_k representam os coeficientes ajustados a partir da amostra de modo que a soma quadrática de todos os erros seja mínima.

Tipicamente, quando se faz uma análise de regressão, procura-se alcançar dois objetivos: a) encontrar um bom ajuste entre os valores preditos pelo modelo e os valores observados da variável dependente; b) descobrir quais das variáveis dependentes contribuem (ou explicam) de forma significativa este relacionamento. Para tanto, pressupõe-se que os erros sejam independentes e não correlacionados com a variável dependente, possuam distribuição normal com média zero e variância constante. Além disso, é comum desconsiderar que as variáveis independentes sejam variáveis aleatórias (NETER et al., 1996).

Existem inúmeras funções não lineares que podem ser utilizadas para descrever o relacionamento entre duas ou mais variáveis, como por exemplo, o modelo polinomial, logarítmico, exponencial e potencial. É importante ressaltar

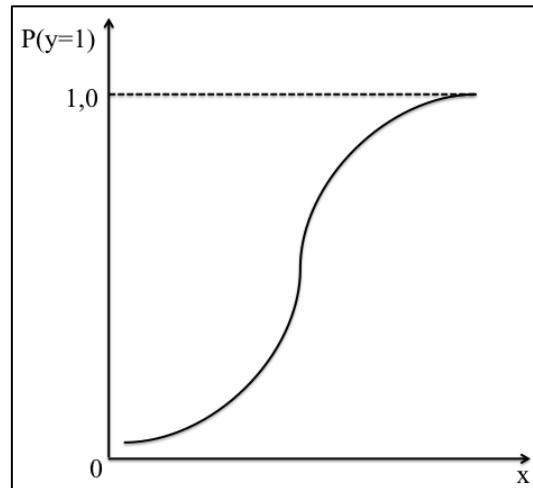
que as pressuposições sobre a distribuição dos erros refletem diretamente sobre o comportamento da variável dependente Y . Por exemplo, considerando-se um modelo linear multivariado, se os erros são normalmente distribuídos, então Y também o será. Quando a variável dependente não pode ser considerada normalmente distribuída então esta abordagem não pode ser aplicada.

A técnica de regressão logística foi desenvolvida por volta de 1960 em resposta ao desafio de realizar previsões ou explicar a ocorrência de determinado fenômeno quando a variável dependente fosse de natureza binária (CORRAR et al, 2014). A regressão logística faz parte de uma categoria de modelos estatísticos denominados modelos lineares generalizados. Esta ampla classe de modelos inclui regressão ordinária e análise de variância (ANOVA), bem como, estatísticas multivariadas tais como análise de covariância (ANCOVA) e modelos de regressão loglinear. A regressão logística permite prever um resultado discreto, como a associação de grupo, de um conjunto de variáveis que podem ser contínuas, discretas, dicotômicas, ou uma mistura de qualquer destas. Geralmente a variável dependente ou variável resposta é dicotômica ou binária, tal como presença/ausência ou sucesso/falha (CONNOR, 2017). Hosmer e Lemeshow (1989) indicam que a forma tradicional da regressão logística consiste em um modelo que relaciona um conjunto de m variáveis independentes X_1, X_2, \dots, X_m a uma variável dependente Y que assume apenas dois estados, 0 ou 1, e afirmam que este modelo tornou-se um método padrão de análise de regressão de dados dicotômicos. Quando a variável dependente Y é binária, tomando valores 0 e 1 com probabilidades p e $1 - p$ respectivamente, Y é uma variável aleatória Bernoulli. Neste caso, o modelo de regressão logístico simples pode ser definido por:

$$P(Y = 1) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (2.1)$$

onde β_0 e β_1 são parâmetros do modelo e podem ser estimados por diversos métodos numéricos. A Figura 2.6 representa a chamada curva da regressão logística (curva S), que descreve a relação entre a probabilidade associada à ocorrência de determinado evento, $P(Y = 1)$ e uma variável independente.

Figura 2.6 - Curva de Regressão Logística Univariada.



Fonte: Produção do autor.

O modelo apresentado na Equação 2.1 pode ser também estendido usando-se duas ou mais variáveis independentes. Neste caso, o modelo de regressão logístico múltiplo pode ser definido por:

$$P(Y = 1) = \frac{e^{\beta_0 + \sum_{i=1}^m \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^m \beta_i X_i}} \quad (2.2)$$

onde m representa o número de variáveis independentes do modelo.

O uso dos modelos de regressão logística depende direta ou indiretamente das estimativas dos seus coeficientes. Porém, a presença de multicolinearidade pode ocasionar problemas no ajuste do modelo por causar impactos nas estimativas dos parâmetros do modelo. O problema da multicolinearidade existe quando há uma dependência linear exata ou aproximada entre as

covariáveis do modelo. Dependendo do nível de associação entre as variáveis independentes, a estimação dos parâmetros pode ficar imprecisa. A detecção de multicolinearidade nem sempre é fácil e em geral recorre-se à análise do fator de inflação da variância *Variance Inflation Factor* (VIF):

$$VIF_k = \frac{1}{1 - r_k^2} \quad (2.3)$$

onde r_k^2 é o coeficiente de determinação obtido pela regressão entre a variável independente X_k e as demais variáveis independentes. Se $VIF_k > 10$, então X_k têm forte colinearidade com as demais variáveis independentes (RIBEIRO, 2015; HOSMER; LEMESHOW, 1989).

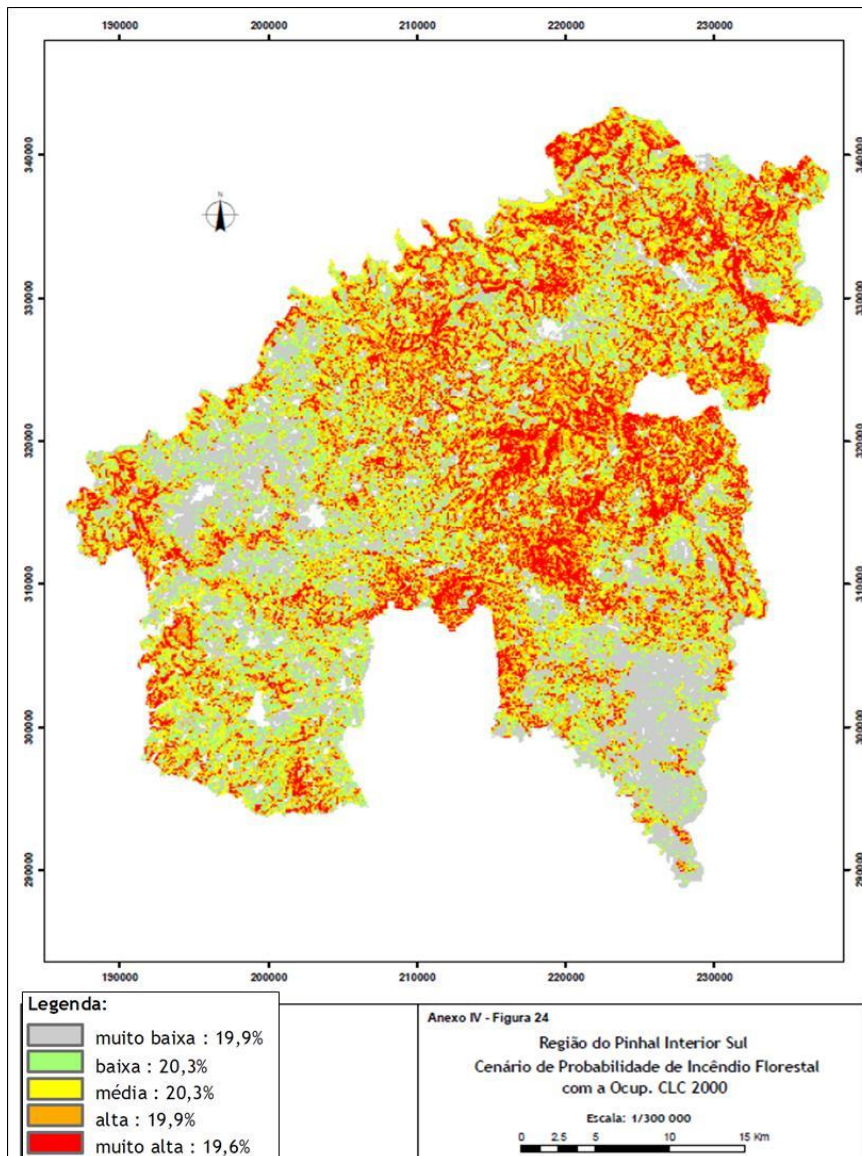
A seleção das variáveis independentes que irão compor o modelo final pode ser feita basicamente por dois métodos: busca exaustiva e *stepwise* (passo a passo). Na busca exaustiva, todas as combinações de variáveis independentes são testadas, desde um modelo sem nenhuma variável selecionada até o modelo completo que considera todas as variáveis independentes simultaneamente. A escolha do melhor modelo é feita com base em índices que indiquem aquele modelo que melhor se ajusta aos dados analisados, como por exemplo, o coeficiente de determinação ajustado. Este método tem um custo computacional muito elevado tornando-se inviável nos casos em que há muitas variáveis independentes envolvidas na seleção. Por outro lado, o método *stepwise* se baseia na construção de um modelo através da inclusão (*stepwise forward*) ou eliminação (*stepwise backward*) de variáveis independentes uma a uma, testando a mudança do poder explicativo de cada modelo. O ganho ou a perda do poder explicativo pode ser avaliado através do teste F que compara os modelos completo e reduzido. Esta avaliação também pode ser feita através do índice *Akaike's Information Criterion* (AIC).

$$AIC = 2k + n \cdot \log\left(\frac{SQE}{n}\right) \quad (2.4)$$

onde k representa o número de parâmetros do modelo, n o tamanho da amostra e SQE é a soma dos quadrados dos erros. Observe que este índice é uma combinação entre uma medida de ajuste do modelo (SQE) e uma medida de simplicidade do modelo (dado pelo número de parâmetros k). Quanto menor for o valor AIC, melhor o modelo.

Um exemplo prático do uso deste tipo de abordagem pode ser encontrado em Saturnino (2012), que a partir de um modelo de regressão logística multivariado calculou a probabilidade de ocorrência de incêndios florestais na região de Pinhal Interior Sul (PIS), centro de Portugal. Neste caso as variáveis selecionadas foram: ocupação do solo, declividade do relevo, exposição, altimetria, distância à rede viária, densidade de rede viária, distancia aos aglomerados populacionais e visibilidade dos postos de vigia. Segundo o autor a técnica da regressão logística empregada mostrou-se perfeitamente adequada ao desenvolvimento de modelos de probabilidade de ocorrência e propagação de incêndios do tipo estrutural, tendo como base o histórico de incêndios, parâmetros topográficos, uso do solo e fatores de origem humana. Como a região é caracterizada por ocupação essencialmente florestal e relevo bastante acidentado, a variável referente à ocupação do solo, à declividade e à distância à rede viária foram os parâmetros com maior contribuição na equação final, ficando a densidade de rede viária e a visibilidade dos postos de vigia com fatores de menor peso. O resultado do modelo espacializado é exibido na Figura 2.7.

Figura 2.7 - Mapa da probabilidade de ocorrência de queimadas.



Fonte: Saturnino (2012).

2.6 Método para avaliação de mapeamento temático

Como citado na seção 2.3 é desejável que mapeamentos temáticos voltados a estudos de cobertura florestal feitos através da aplicação de técnicas de sensoriamento remoto venham acompanhados de índices de validação, agregando informações de qualidade ao produto. O princípio da avaliação de precisão é baseado na comparação do mapa classificado com dados de referência de qualidade superior, coletados através de uma abordagem

baseada em amostras. Os dados de referência de melhor qualidade podem também ser obtidos através de dados recolhidos em campo, mas como isso é caro e trabalhoso, ele é mais comumente obtido através de imagens de satélite de melhor resolução espacial do que os dados que foram utilizados para criar o mapa original (FAO, 2016).

Realizar processos de aferição de erros em levantamentos de larga escala territorial não é tarefa simples nem barata. Aferições em campo de mapeamentos de grande extensão são prejudicadas, em parte, pela dificuldade de acesso aos locais a serem visitados e pelos custos financeiros, operacionais e logísticos necessários, além de demandarem tempo elevado de planejamento e execução. Nos últimos 10 anos, pesquisadores têm estudado métodos alternativos para avaliar a exatidão em mapeamentos de larga escala, quase todos baseados em soluções que utilizam a internet como plataforma operacional, desenvolvendo ferramentas que integram bases de imagens de alta resolução disponíveis no ambiente *Google Maps/Earth*, *Bing Maps* (Microsoft) ou *Leaflet* e séries temporais extraídas dos produtos do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), além de outros dados auxiliares. Baseando-se nestes estudos, Adami et al. (2012) conceberam e desenvolveram uma ferramenta que combinava em um mesmo ambiente gráfico as imagens disponíveis no ambiente *Google Maps*, imagens *Landsat* e séries temporais extraídas do sensor MODIS, para mensurar erros e validar os mapeamentos gerados no âmbito do Projeto Monitoramento da Cana-de-Açúcar via Imagens de Satélite – Canasat (INPE, 2016), determinando seus níveis de exatidão.

Em mapeamentos do porte do PRODES, mesmo executados por profissionais experientes e respeitando um rigor metodológico na sua confecção, erros são esperados e encontrados no resultado final. Assim, uma tarefa necessária é quantificá-los e informá-los de maneira transparente para garantir o fornecimento de parâmetros de confiabilidade mínimos para os mapas gerados. Em virtude da escala territorial envolvida neste projeto, missões de campo para aferição da qualidade da informação gerada são inviáveis por ser dispendiosas financeiramente, pois dependem de logística complexa para acesso as áreas

desmatadas e também demandam tempo elevado no planejamento e execução.

2.7 Ferramenta *Web-Gis* para avaliação

A ferramenta de avaliação inicialmente concebida por Adami et al. (2012) foi atualizada por Maurano e Adami (2017) e adaptada para ser utilizada neste estudo. Esta ferramenta consiste em submeter pontos sorteados aleatoriamente com base no mapeamento a ser validado a um processo de avaliação feito por um ou mais especialistas independentes, que realizam a classificação dos pontos sem ter conhecimento prévio da sua classe no mapeamento original. O dado avaliado por este(s) especialista(s) será considerado como o dado de referência e posteriormente comparado com o mapa base. Seu funcionamento é todo baseado em uma página de internet construída para facilitar esta tarefa. Esta ferramenta apoia-se em quatro tecnologias básicas:

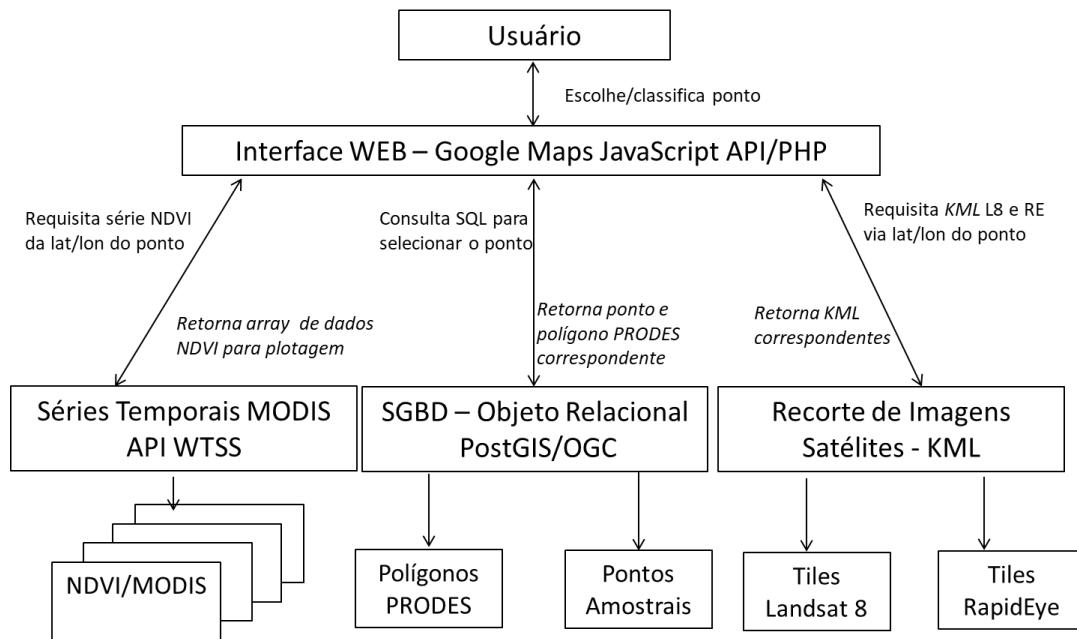
- a) API (Interface de Programação de Aplicativos) *JavaScript* do *Google Maps, Bing Maps ou Open Layers*: Conjunto de funções escritas em linguagem *JavaScript* que possibilitam customizar aplicações utilizando o ambiente de visualização de imagens e mapas, incorporando-os a uma página internet específica;
- b) Sistema Gerenciador de Banco de Dados Objeto-Relacional (SGBD-OR): Gerenciador de banco de dados com extensão espacial para armazenar os dados vetoriais contendo os polígonos do PRODES e os pontos amostrais da validação em padrões estabelecidos pelo *Open Geospatial Consortium (OGC)*;
- c) *Web Time Series Service (WTSS)*: API *Java Script* para extração e plotagem de séries temporais geradas a partir das coleções de imagens do sensor MODIS dos produtos MOD09Q1 e MOD13Q1. Esta API foi desenvolvida pelo grupo da Divisão de Processamento de Imagens (DPI) do INPE e permite acesso as camadas armazenadas em banco de

dados espaço-temporal do tipo matriz multidimensional, como o SciDB, ou em um sistema de arquivos estruturados para consulta (QUEIROZ et al. 2015);

- d) GDAL - Biblioteca de Abstrações de Dados Geo-Espaciais: Utilizada para realizar os recortes prévios das imagens *Landsat8/OLI* e *RapidEye*, armazenando-os em formato KML (*Keyhole Markup Language*);
- e) Linguagem PHP (*Hypertext Preprocessor*) é uma linguagem interpretada livre, usada originalmente para o desenvolvimento de aplicações presentes e atuantes no lado do servidor, capaz de gerar conteúdo dinâmico na Internet.

A interface web, construída sobre a API do *Google Maps* ou similares utilizando linguagem HTML, PHP e *Java Script*, concentra as principais funções da ferramenta. Através dela o analista-avaliador se comunica com o sistema gerenciador de banco de dados, selecionando o ponto a ser validado. Ela também recupera e exibe a série temporal *Normalized Difference Vegetation Index* (NDVI) do MODIS correspondente ao ponto escolhido, via API WTSS. Adicionalmente, a interface coordena a exibição dos recortes das imagens *Landsat 8/OLI* e *RapidEye*, armazenados em formato KML, que servem de auxílio para o processo de validação. A arquitetura básica da ferramenta é apresentada na Figura 2.8. Maiores detalhes sobre esta ferramenta de validação podem ser encontrados em Maurano e Adami (2017) e Adami et al. (2012).

Figura 2.8 - Arquitetura simplificada do sistema.



Fonte: Produção do autor.

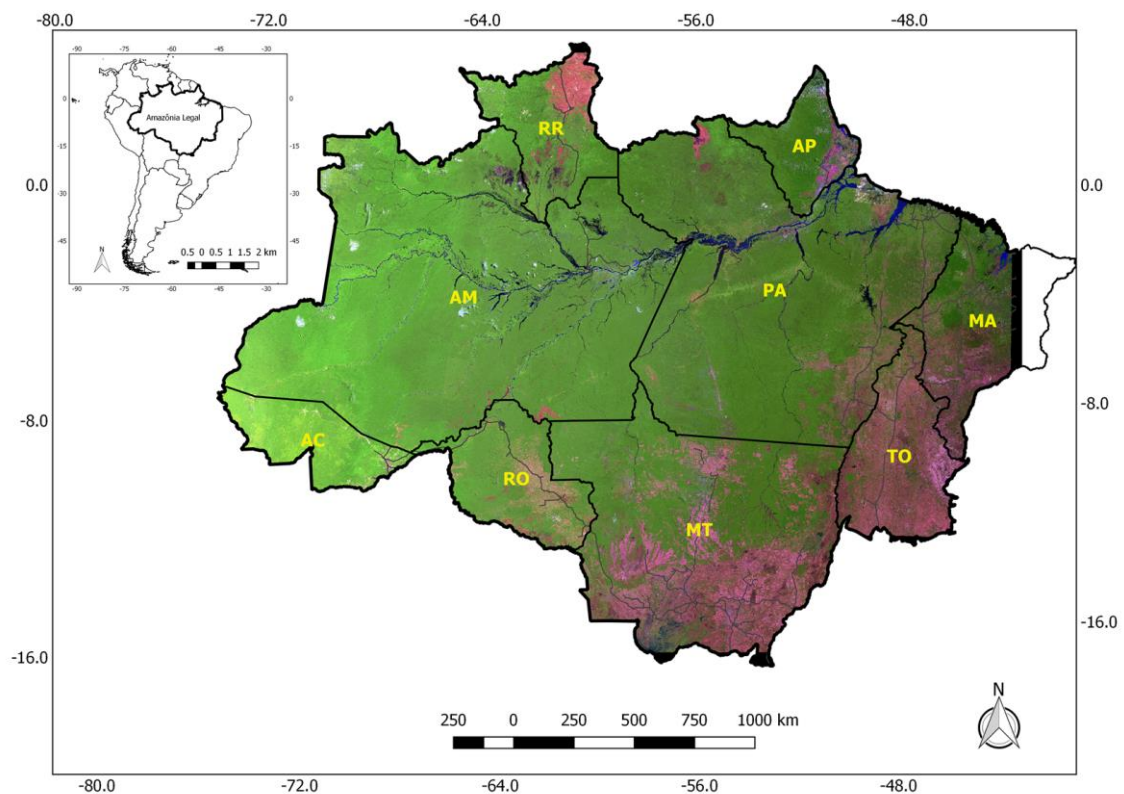
3 METODOLOGIA

Neste Capítulo é apresentada a área de estudo, os materiais e a metodologia utilizada no desenvolvimento deste estudo.

3.1 Área de estudo

A área de estudo escolhida para este estudo foi a que se denomina Amazônia Legal. Trata-se de uma região que engloba nove estados brasileiros, pertencentes à bacia amazônica, sendo eles: Mato Grosso, Pará, Acre, Amazonas, Roraima, Rondônia, Amapá, Tocantins e Maranhão (neste caso só a parte a oeste do meridiano 44°) perfazendo uma superfície de aproximadamente cinco milhões de km² correspondente a cerca de 60% do território brasileiro. A Figura 3.1 apresenta a área de estudo selecionada.

Figura 3.1 - Mapa da área de estudo correspondente à Amazônia Legal, contendo mosaico de imagens do satélite *Landsat* TM ano 2000 cobrindo os nove estados da Amazônia Legal.



Fonte: Produção do autor.

3.2 Relação dos dados utilizados

Os dados utilizados neste estudo estão descritos na Tabela 3.1.

Tabela 3.1 - Principais dados utilizados neste estudo.

Dados	Tipo	Fonte
Máscara com desmatamento acumulado até 2014	Dado vetorial em forma de polígonos	PRODES
Incremento dos desmatamentos de 2014	Dado vetorial em forma de polígono	INPE
Recortes de imagens <i>RapidEye</i> 2014 (utilizados na validação)	Dado matricial em composição colorida recortados em <i>tiles</i> de 50 x 50 km para as 120 células utilizadas na validação	Acervo INPE
Recortes de imagens <i>Landsat</i> anos 2013 e 2014 (utilizados na validação)	Dado matricial em composição colorida recortados em <i>tiles</i> de 50 x 50 km para as 120 células utilizadas na validação	PRODES INPE

Fonte: Produção do autor

3.3 Procedimentos metodológicos

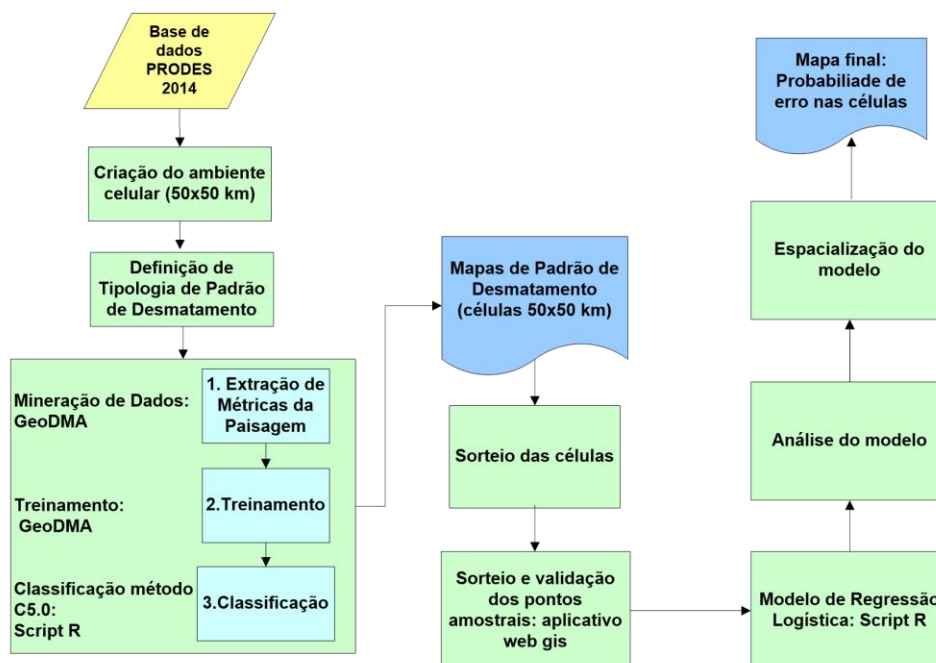
Como já mencionado no Capítulo 1, o objetivo principal deste estudo é o desenvolvimento de uma metodologia consistente que permita indicar variáveis que influenciam os erros, baseadas nos padrões de desmatamento e espacializar os erros nos dados gerados no âmbito do projeto PRODES, através de construção de modelos estatísticos multivariados. Essa metodologia é descrita a seguir.

3.3.1 Fluxograma geral da metodologia

A Figura 3.2 esquematiza de forma resumida as etapas constantes na metodologia a ser empregada neste estudo. Essas etapas consistem primeiramente na preparação dos dados do PRODES para sua representação

no espaço celular, extração das métricas de paisagem e posterior classificação dos padrões de desmatamento, de acordo com uma tipologia estabelecida por Saito et al. (2010). A classificação dos padrões de desmatamento tem como objetivo compartimentar a área de estudo em regiões homogêneas considerando os padrões espaciais de desmatamento. Posteriormente, com base nos padrões obtidos no total de células classificadas, um subconjunto foi sorteado aleatoriamente nos estratos representados pelas classes obtidas no mapeamento. O processo de validação, baseado em pontos amostrais sorteados no conjunto de células selecionadas, foi realizado por três intérpretes no mesmo ambiente de validação descrito de forma detalhada na Seção 2.8 deste documento. Na etapa seguinte, um modelo de regressão logística foi obtido para explicar as principais fontes de erro, considerando variáveis relacionadas com os níveis de fragmentação da paisagem, a partir de métricas de paisagem extraídas para cada célula. Essas etapas são detalhadas nas seções seguintes.

Figura 3.2 - Fluxograma resumido.


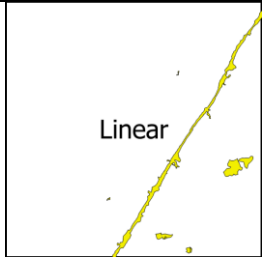


Fonte: Produção do autor.

3.3.2 Análise de métricas de paisagem e padrões de desmatamento


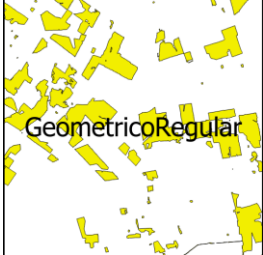
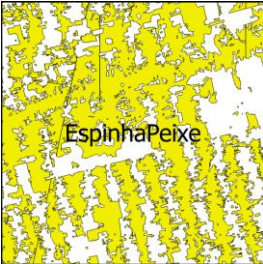


Para análise de padrões de desmatamento foram utilizados dados do PRODES agregando-se todos dados de desmatamento acumulado até 2014 em uma base única. Fatores como área, perímetro e forma do polígono de desmatamento, a fragmentação florestal, a proximidade do desmatamento em relação às áreas de borda de floresta e áreas de desmatamento consolidado, entre outros, podem ter influência direta na interpretação das imagens e no correto mapeamento dos polígonos de desmatamento pelo PRODES. Para a contextualização dos polígonos de desmatamento foram definidas regiões baseadas na análise e classificação de padrões de desmatamento na Amazônia Legal descritos na literatura, como, por exemplo, em Escada (2003) e em Ewers e Laurance (2006). Saito et al. (2011) estabeleceu uma tipologia semelhante, baseada em células de 50 X 50 km, com seis principais categorias de padrões de desmatamento encontrados nos dados do PRODES, conforme descrito na Tabela 3.2.

Tabela 3.2 - Descrição dos padrões de desmatamento.

Tipologia	Descrição	Exemplo
Difuso	Composto por manchas pequenas, isoladas; Forma variada, irregular, baixa a média densidade, distribuição uniforme. Representa áreas em estágios iniciais de ocupação.	
Linear	Composto por manchas médias e grandes, manchas de forma alongada disposta ao longo das vias de acesso, baixa densidade. Representa ocupação inicial ao longo de rodovias.	

continua

Tabela 3.2 – Conclusão.

Tipologia	Descrição	Exemplo
Multidirecional desordenado	Composto por manchas pequenas, médias e grandes, com forma variada, irregular e complexa. Média/alta densidade. Representa estágio intermediário de ocupação.	 <p>Multidirecional</p>
Geométrico Regular	Composto predominantemente por manchas médias a grandes e isoladas, forma geométrica regular, baixa a média densidade. Representa formas de ocupação inicial e/ou intermediária de áreas de grandes fazendas.	 <p>GeometricoRegular</p>
Espinha de Peixe	Composto predominantemente por manchas grandes alongadas e lineares com ramificações semelhantes à vértebra de peixe; Média densidade. Padrão associado às formas de ocupação do INCRA, de projetos de assentamentos, cujos lotes são ocupados por pequenos produtores rurais, em estágios iniciais e intermediários de ocupação.	 <p>EspinhaPeixe</p>
Consolidado	Composto por manchas grandes e contínuas de desmatamento, forma variada, densidade baixa e áreas pequenas de remanescentes florestais, manchas de áreas desmatadas compactas. Representa estágios avançados de ocupação.	 <p>Consolidado</p>
Floresta	Composto predominantemente por floresta primária no interior da célula e ausência quase total de manchas de desmatamento.	 <p>Floresta</p>

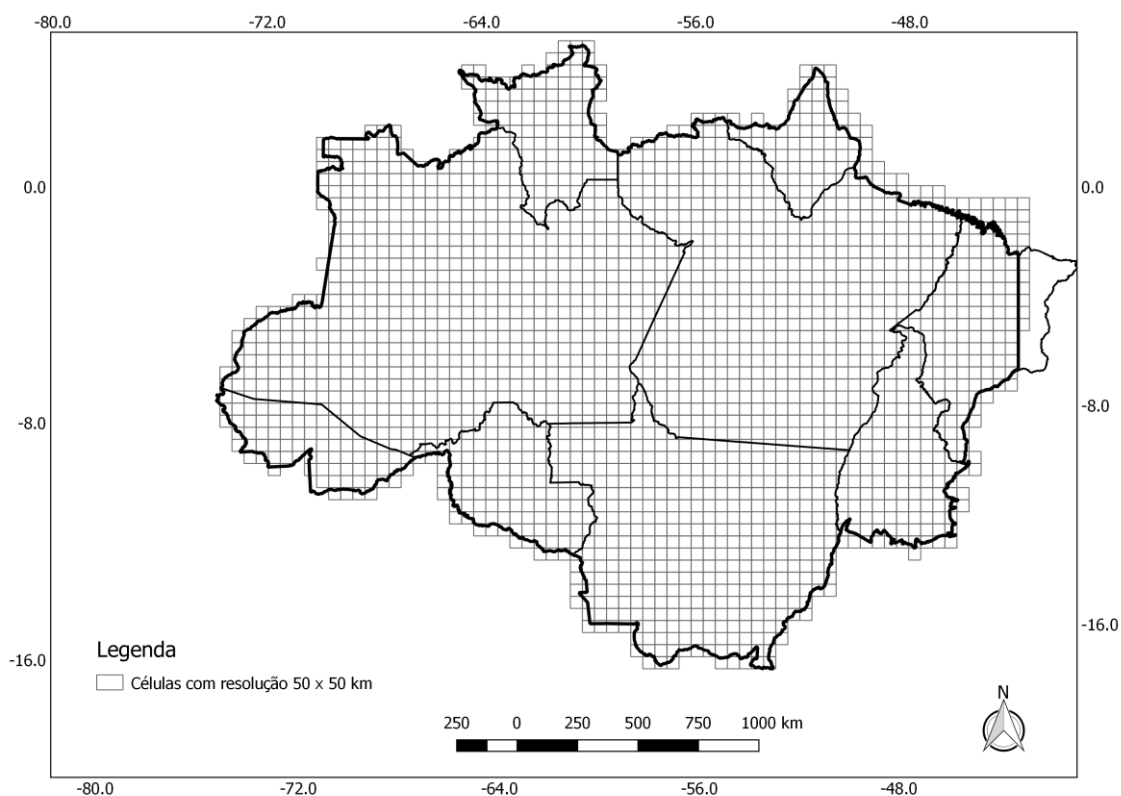
Fonte: Adaptado de Saito (2010).

Na análise baseada em células, a área de estudo é subdividida em pequenas regiões regulares. No espaço celular, os polígonos são agregados em unidades maiores (células) e cada uma dessas unidades passa a representar uma porção da paisagem. Assim, cada célula desta grade é associada a um padrão de desmatamento, onde cada padrão é descrito por um conjunto de métricas da paisagem.

A extração das métricas de paisagem no espaço celular foi realizada no *Geographic Data Mining Analyst* (GeoDMA) que é um sistema que integra ferramentas de análise de métricas de paisagens, manipulação de recursos multitemporais e técnicas de mineração de dados (KORTING et al., 2013). O GeoDMA foi implementado na Linguagem C++, baseado na biblioteca TerraLib (INPE, 2017) e funciona como um *plugin* para o aplicativo do TerraView versão 5 (INPE, 2017), que possibilita manipular e visualizar dados armazenados em bancos de dados geográficos. Os dados de entrada do GeoDMA podem ser imagens, objetos provenientes de segmentação, mapas vetoriais e grades celulares. Para a execução deste estudo optou-se por utilizar os dados de desmatamento gerados pelo PRODES em formato vetorial padrão *shapefile* (ESRI).

Um problema crítico nesta abordagem está associado à definição da resolução da grade de células, que normalmente é feita de forma empírica. No caso deste estudo, foi utilizado um plano de células com resolução de 50 x 50 km. Este valor foi definido com base no trabalho de Saito et al. (2011) que analisou a sensibilidade de algumas métricas da paisagem, na escala regional da Amazônia Legal, em relação à variação do tamanho das células associadas aos padrões de desmatamento encontrados nos dados do PRODES, observando uma maior estabilidade das métricas no tamanho de 50 X 50 km. Para envolver toda a área de estudo foram criadas de 2.191 células. A Figura 3.3 mostra o espaço celular com resolução de 50 x 50 km criado para acomodar o processamento dos dados deste estudo. Uma vez definido o tamanho das células e obtida a grade celular, a etapa seguinte consistiu na extração de métricas de paisagem e na classificação das células com técnicas de mineração de dados como apresentado na seção a seguir.

Figura 3.3 - Grade celular de 50 x 50 km.



Fonte: Produção do autor.

3.3.3 Mineração de Dados

O processo de mineração dos dados foi realizado como o uso do aplicativo GeoDMA, com a abordagem celular e constituiu quatro fases distintas:

- 1) extração de métricas da paisagem para cada célula, baseado no mapa de desmatamento acumulado do PRODES até 2014;
- 2) seleção de amostras de treinamento e teste;
- 3) classificação das células, utilizando um algoritmo de árvore de decisão C5.0 onde são definidos os limiares das métricas que distinguem as classes e;
- 4) avaliação da classificação obtida comparando-a com as amostras independentes.

Com base neste no espaço celular de resolução de 50 x 50 km, foram calculadas 21 métricas da paisagem descritas na Tabela 3.3 cuja formulação se baseia no trabalho de McGarigal et al.(1995).

Tabela 3.3 - Descrição das métricas calculadas.

Variável (Tipo)	Descrição	Fórmula	Intervalo (Unidade)
CA (A)	Medida da composição da paisagem representada pela soma de áreas de todos os fragmentos de determinada classe.	$CA = \sum_{j=1}^n a_j$	≥ 0 (ha)
PLAND (A)	Porcentagem da área desmatada.	$PLAND = \frac{\sum_{j=1}^n a_j}{A} \times 100$	[0,100] (%)
PD (A)	Número de manchas dividido pela área total.	$PD = \frac{n}{A} \times 10000 \times 100$	≥ 0 (nº/100ha)
NP (A)	Número de manchas de um determinado tipo dentro da célula.	$NP = n$	≥ 0
MPS (T)	Média do tamanho da mancha.	$MPS = \frac{\sum_{j=1}^n a_j}{n} \times 10^{-4}$	≥ 0 (ha)
PSSD (T)	Desvio padrão das áreas das manchas.	$PSSD = \sqrt{\frac{\sum_{j=1}^n (a_j - MPS)^2}{n}} \times 10^{-4}$	≥ 0 (ha)
LSI (F)	Medida de complexidade de forma das manchas.	$LSI = \frac{\sum_{j=1}^n e_j}{2 \times \sqrt{\pi} \times A}$	≥ 0
MSI (F)	Média da razão do perímetro pela raiz quadrada da área dividido pelo número de manchas.	$MSI = \frac{\sum_{j=1}^n \frac{p_j}{2 \times \sqrt{\pi} \times a_j}}{n}$	≥ 1
AWMSI (F)	Razão entre o perímetro e a raiz quadrada da área, ponderado pela área da mancha	$AWMSI = \sum_{j=1}^n \left[\frac{p_j}{2 \times \sqrt{\pi} \times a_j} \times \frac{a_j}{\sum_{j=1}^n a_j} \right]$	≥ 1
MPFD (F)	Dimensão fractal média das manchas.	$MPFD = \frac{\sum_{j=1}^n \frac{2 \times \ln p_j}{\ln a_j}}{n}$	

continua

Tabela 3.3 – Continuação.

Variável (Tipo)	Descrição	Fórmula	Intervalo (Unidade)
AWMPFD (F)	Dimensão fractal da mancha média ponderada pela área das manchas na paisagem.	$AWMPFD = \sum_{j=1}^n \left[\frac{2 \times \ln p_j}{\ln a_j} \times \frac{a_j}{\sum_{j=1}^n a_{ij}} \right]$	
TE (B)	Soma dos perímetros das manchas.	$TE = \sum_{j=0}^n e_j$	≥ 0 (m)
ED (B)	Quantidade de borda em relação à área de paisagem.	$ED = \frac{\sum_{j=1}^n e_j}{A} \times 10^4$	≥ 0 (m/ha)
MPAR (F)	Razão do perímetro médio pela área de cada mancha dividido pelo número de manchas na paisagem.	$MPAR = \frac{\sum_{j=1}^n \frac{p_j}{a_j}}{n}$	≥ 0 (m ⁻¹)
PSCOV (T)	Coefficiente de variação do tamanho das manchas.	$PSCOV = \frac{PSSD}{MPS} \times 100$	≥ 0
TABO (A)	Área total do maior objeto que intercepta a paisagem.		(ha)
PR (T)	Número de diferentes tipos de classes presentes dentro do limite da paisagem.	$PR = m$	≥ 0
IJI (F)	Índice de interspersão e justaposição. A interspersão observada sobre a máxima interspersão possível para o número determinado de tipos de manchas (n>3).	$IJI = \frac{-\sum_{j=1}^n \left(\frac{e_j}{\sum_{k=1}^n e_k} \right) \times \ln \left(\frac{e_j}{\sum_{k=1}^n e_k} \right)}{\ln(m-1)} \times 100$	[0,100] (%)
PRD (A)	Número de diferentes tipos de manchas presentes no limite da paisagem dividido pela área total da paisagem.	$PRD = \frac{m}{A} \times 10000 \times 100$	≥ 0 (nr/100 ha)

continua

Tabela 3.3 – Conclusão.

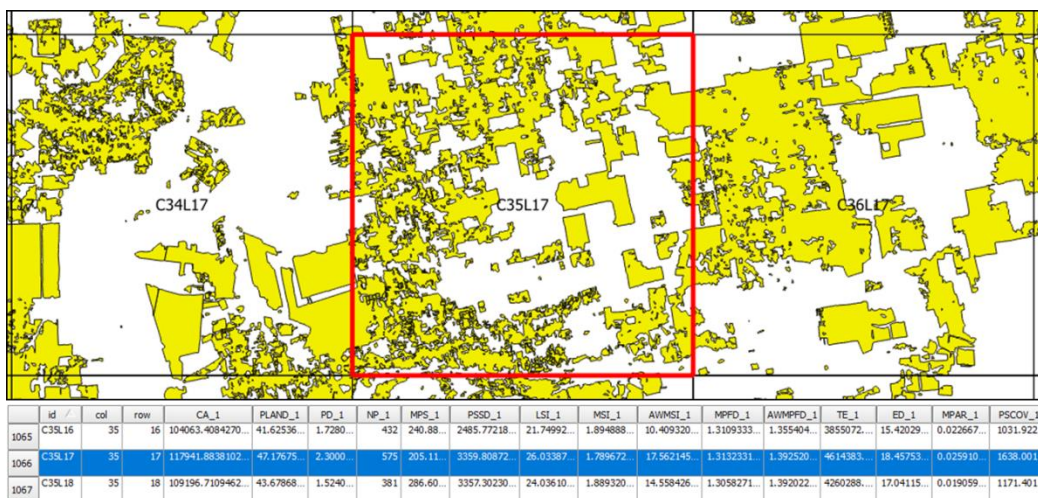
Variável (Tipo)	Descrição	Fórmula	Intervalo (Unidade)
SHDI (T)	Índice de Diversidade de Shannon.	$SHDI = \sum_{i=0}^m P_i \times \ln P_i$	$0 \leq SHDI \leq 0$
SIDI (T)	Índice de Diversidade de Simpson.	$SIDI = 1 - \sum_{i=0}^m P_i^2$	$0 \leq SIDI \leq 0$

Tipos dos índices: A: índice de área; B: índice de borda; F: índice de forma e T: índice de tamanho.

Fonte: Produção do autor.

Esta etapa foi realizada com uso do aplicativo GeoDMA, sendo os valores das métricas acima calculadas para cada célula e armazenados como novos campos na tabela de atributo do ambiente celular, como mostra a Figura 3.4. As formulações matemáticas utilizado para cálculo das métricas acima podem ser obtidas McGarigal et al.(1995).

Figura 3.4 - Exemplo de métricas de paisagem calculada para a célula realçada em vermelho.



Fonte: Produção do autor.

3.3.4 Amostras de treinamento para classificação

Calculadas as métricas para todas as células da área de estudo, a etapa seguinte consistiu em coletar as amostras de treinamento para posterior classificação visando obter o mapa de padrão de desmatamento regionalizado em células de 50 x 50 km. Esta operação foi feita manualmente selecionando um conjunto de células representativas de cada padrão de desmatamento e descrito na Tabela 3. A Figura 3.5 mostra alguns exemplos de padrões coletados para cada classe da tipologia.

Figura 3.5 - Padrões de desmatamento utilizados no treinamento.



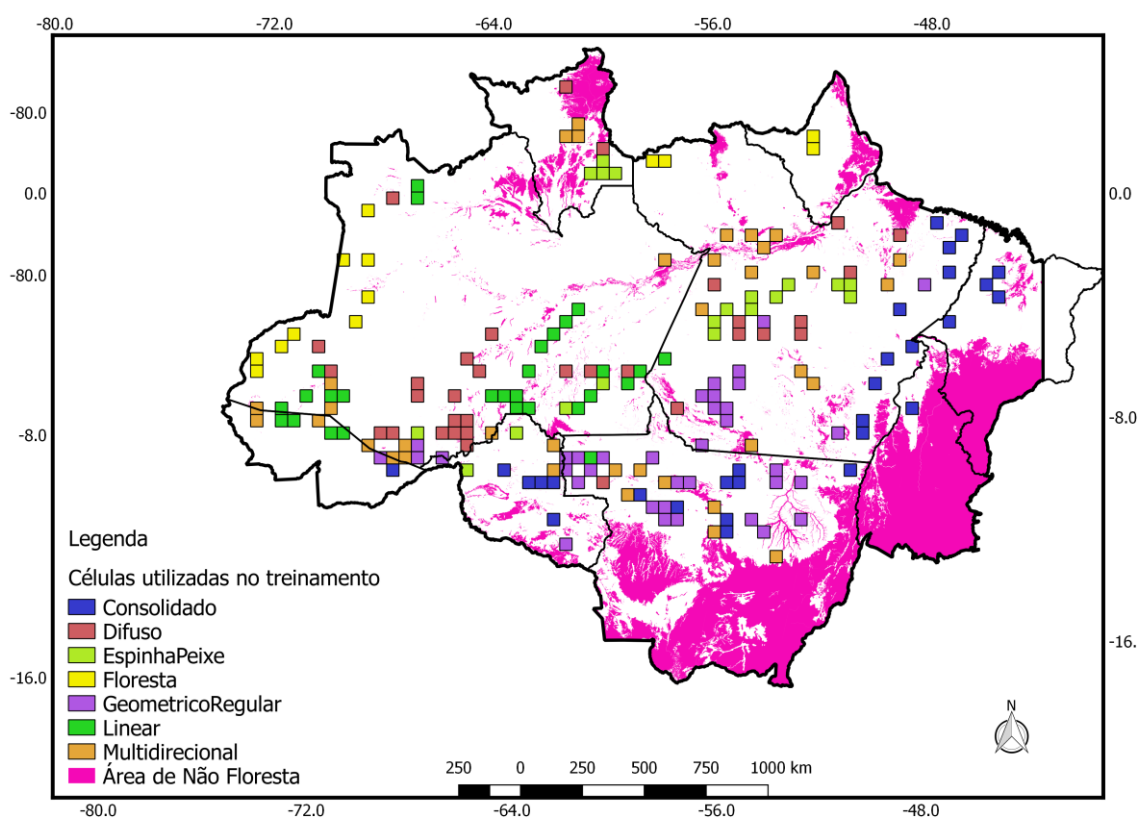
Fonte: Produção do autor.

A Figura 3.6 mostra a distribuição espacial das amostras coletadas no processo de treinamento. As áreas classificadas como Não Floresta¹ no mapeamento do PRODES, por se tratar de vegetação não florestal, segundo classificação do IBGE, não foram consideradas no momento da escolha das

¹No mapeamento executado pelo PRODES denomina-se “não floresta” qualquer tipo de vegetação que, pela definição do projeto Radam Brasil/IBGE, não seja do tipo: Ombrófila Densa, Floresta Ombrófila Aberta, Floresta Estacional Decidual, Áreas de Formação Pioneira de Influência Fluvial (Vegetação Aluvial), Vegetação Lenhosa Oligotrófica dos Pântanos e das Acumulações Arenosas (Campinarana) e áreas de Tensão Ecológica (Contato floresta/cerrado), com predominância de Fisionomia Florestal.

células de treinamento. Para cada padrão de desmatamento foi escolhido um conjunto de células visando compor as amostras de treinamento, que totalizaram 194 células coletadas visualmente de forma a contemplar todos os tipos de padrões envolvidos, sendo: a) Multidirecional: 36 amostras; b) Geométrico Regular: 36 amostras; c) Consolidado: 29 amostras; d) Difuso: 29 amostras; d) Espinha de Peixe: 19 amostras; e) Linear: 28 amostras e f) Floresta: 13 amostras.

Figura 3.6 - Distribuição espacial das amostras de treinamento para classificação.



Fonte: Produção do autor.

3.3.5 Classificação

Como a versão 5 do GeoDMA utilizada neste estudo não contemplava o novo algoritmo de classificação supervisionado baseado no conceito de árvore de decisão versão C5.0, foi necessário utilizar um programa escrito na linguagem R (R CORE TEAM, 2017) para realização desta etapa, cujo código fonte

encontra-se no Anexo I deste documento. Os detalhes da classificação por árvore de decisão e os resultados obtidos estão descritos nas Seções 2.4 e 4.1 respectivamente.

O método de classificação por árvore de decisão para geração posterior do mapa de padrão de desmatamento utilizando o algoritmo C5.0 consiste em gerar uma árvore de decisão extraída dos dados de amostras de treinamento e aplicá-la a toda base de dados, classificando-a segundo seus parâmetros. Porém uma pergunta pertinente seria “como garantir que uma única árvore de decisão gere um mapa de padrão de desmatamento que melhor represente os padrões destes desmatamentos verificados na Amazônia Legal?”.

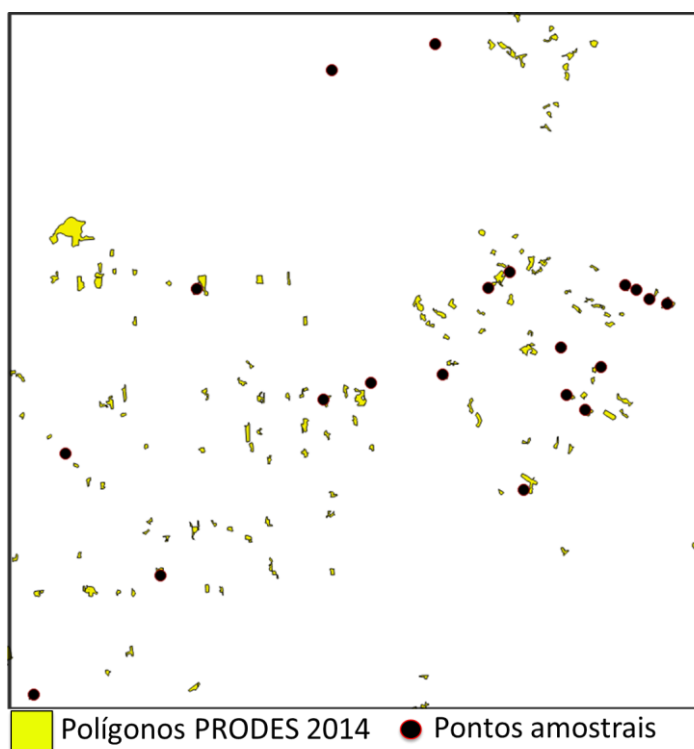
Para tratar esta questão, a solução encontrada foi não utilizar somente uma única árvore e sim gerar um conjunto de 100 árvores de decisão independentes, obtendo dessa maneira 100 mapas classificados. Para isso foi desenvolvido um programa escrito no pacote estatístico R que executou esta tarefa de forma automática, este programa foi construído de forma a utilizar 75% das amostras de treinamento para serem aplicadas na classificação propriamente dita e 25% das amostras para serem utilizadas como teste da classificação.

3.3.6 Sorteio do painel amostral para validação

A metodologia utilizada neste estudo está baseada na avaliação dos dados gerados pelo PRODES utilizando a abordagem de interpretação independente de pontos amostrais sorteados sobre a base de dados, conforme descrito na Seção 2.6. A solução empregada foi gerar um painel amostral de pontos através de sorteio aleatório estratificado utilizando o ambiente espaço-celular de resolução de 50 x 50 km como referência e valendo-se do mapa de padrão de desmatamento obtido com a classificação das métricas de paisagem descrita nas Seções 3.3.4 e 3.3.5. Inicialmente foi sorteado o total de 120 células, sendo 20 células para cada um dos padrões de desmatamento definidos para este estudo. O sorteio independente e aleatório das 120 células foi realizado em um programa escrito na linguagem R.

Para cada conjunto de células sorteadas em cada tipo de padrão de desmatamento, foi gerado um diagrama de *boxplot* comparando o comportamento do desmatamento acumulado até 2014, encontrado nas 20 células sorteadas para cada padrão de desmatamento com o restante da base de dados. *Boxplot* é uma ferramenta gráfica que representa a variação de dados observados de uma variável numérica por meio de quartis, identificando onde estão localizados 50% dos valores mais prováveis, a mediana e os valores extremos. Com base em cada célula selecionada pelo sorteio, foram gerados pontos amostrais que ficaram restritos ao limite geográfico desta célula. Empiricamente estabeleceu-se o sorteio de 15 pontos sobre área de desmatamento PRODES de 2014 para cada célula, obtendo-se um total aproximado de 1800 pontos amostrais que foram posteriormente classificados por três especialistas independentes com uso da ferramenta de validação. Ressalta-se que este valor de 15 pontos por célula nem sempre foi alcançado, pois em algumas das células selecionadas não havia polígonos de desmatamento para servir como base do sorteio. A Figura 3.7 mostra um exemplo uma célula com os respectivos pontos amostrais para validação sorteados no interior de cada uma delas.

Figura 3.7 - Pontos amostrais para validação sorteados na célula.

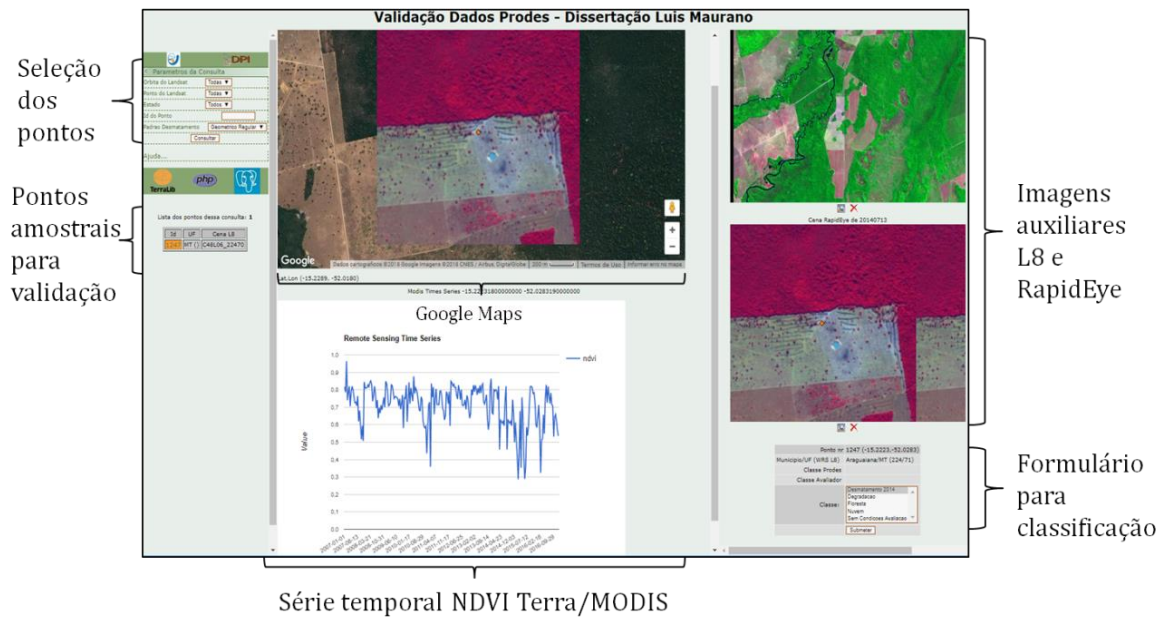


Fonte: Produção do autor.

3.3.7 Validação dos pontos amostrais

Para classificação dos pontos foi utilizada a ferramenta de validação desenvolvida e descrita na Seção 2.6. Neste caso, a validação foi executada por três intérpretes independentes alocados especificamente para esta operação. A interface principal da ferramenta de validação utilizada no processo de classificação dos pontos amostrais sorteados nas células está apresentada na Figura 3.8.

Figura 3.8 - Interface da ferramenta de validação.



Fonte: Produção do autor.

A operação do aplicativo é bastante simples. Todo o processo inicia-se com a escolha do ponto a ser classificado que é mostrado em uma lista no menu esquerdo da interface. Escolhendo-se o ponto, este é visualizado de forma centralizada na tela do *Google Maps*, onde também é plotado o gráfico da série temporal do NDVI extraída do produto MODIS/MOD13Q1. Na interface à direita, encontram-se os recortes de imagens de satélite correspondendo ao entorno do ponto sorteado, sendo que dois destes recortes foram extraídos das imagens *Landsat 8/OLI* utilizadas no PRODES para os anos de 2013 e 2014 e um terceiro extraído das imagens de alta resolução do satélite *RapidEye* de 2014. Logo abaixo deste último recorte está disponível um formulário, no qual os intérpretes definem a qual classe pertence o ponto selecionado. Ao finalizar o processo de classificação de cada ponto, a classe a ele associada, além do tempo gasto na sua avaliação são armazenados na tabela alfanumérica do sistema gerenciador de banco de dados. Como a abordagem previa avaliação independente realizada por três interpretes, para definição de uma classe final única validada foi utilizada da classe referente ao valor da moda entre as três avaliações.

3.4 Modelagem estatística

A fim de identificar os fatores condicionantes da ocorrência de erro de mapeamento PRODES para o ano de 2014, foi adotada neste estudo a abordagem baseada na regressão logística. Todo processamento para construção do modelo foi feito utilizando-se a linguagem R, através do uso das bibliotecas *Companion to Applied Regression* - CAR 2.1-4 e MASS 7.3-45. A estimação do modelo logístico é feita através da função *glm*, utilizando a função de ligação *logit* considerando-se que a variável dependente Y representa a probabilidade (p) de erro de detecção de desmatamento em n pontos avaliados independentemente no conjunto de células selecionadas. Esta probabilidade foi estimada pela proporção de erro de detecção de desmatamento, ou seja, a proporção dos pontos avaliados que não foram classificados como desmatamento pela maioria dos avaliadores em cada célula selecionada. Devido à ausência de desmatamentos no ano de 2014 em algumas células e da impossibilidade de validação de alguns pontos pelos avaliadores, algumas células foram descartadas desta análise.

Como variáveis independentes, foram utilizadas aquelas definidas na Tabela 3.4 utilizadas durante o processo de mineração de dados para classificação dos padrões de desmatamento. Além destas, foram calculadas, para cada célula, as variáveis constantes na Tabela 5 que representam proporções de cada classe mapeada pelo PRODES.

Tabela 3.4 - Descrição das variáveis adicionadas ao estudo.

Variável	Descrição
P_Floresta	Porcentagem de floresta em relação à área da célula
P_Prodes2014	Porcentagem de desmatamento do PRODES 2014 em relação à área da célula
P_Prodes2014Masc	Porcentagem de desmatamento do PRODES 2014 em relação ao PRODES acumulado até 2013 (máscara 2013)
P_Prodes2014Flor	Porcentagem de desmatamento do PRODES 2014 em relação à área de floresta

Fonte: Produção do autor.

Inicialmente foi analisada a colinearidade entre as variáveis independentes, eliminando-se aquelas que apresentaram valores de VIF (Equação 2.3) muito altos. Este procedimento garantiu que o processo de seleção não fosse prejudicado pela presença destas variáveis. Para a construção do modelo de regressão, foi utilizada a função *stepAIC* do R. Inicialmente foram gerados dois modelos: modelo com todas as variáveis independentes e o modelo nulo, sem nenhuma variável independente. Foram testadas as abordagens *forward*, *backward* e *both*, todas usando como critério de ganho de poder explicativo do modelo o índice *AIC* (Equação 2.4). É importante ressaltar que os modelos selecionados por essas abordagens não garantem que todas as variáveis selecionadas sejam significativas. Também foram avaliados os modelos por busca exaustiva até que o acréscimo de variáveis no modelo não trouxesse ganhos significativos, para isso também foi utilizado um programa escrito em linguagem R. Todos os modelos selecionados foram avaliados de modo a garantir que todas as variáveis independentes presentes sejam significativas e que tenham baixo valor de *VIF*.

3.5 Espacialização do modelo de regressão

As variáveis descritas na seção anterior integram o modelo de regressão logística multivariado onde a equação de ajuste gerada pela regressão é aplicada a cada célula de 50 x 50 km. Com isso foi possível obter um mapa resultante da aplicação do modelo de regressão para todas as células, indicando regiões com maior ou menor probabilidade de erro com valores variando entre 0 e 1 (ou entre 0 e 100%).

4 RESULTADOS

Neste Capítulo são descritos os resultados obtidos no estudo, organizados em 5 seções. Na Seção 4.1 é apresentado o mapa de padrões de desmatamento obtido baseado no ambiente espaço-celular de 50 x 50 km gerado após a classificação dos padrões de desmatamento. A Seção 4.2 discorre sobre a amostragem para confecção do painel amostral de validação; a Seção 4.3 exhibe as análises preliminares realizadas a partir dos pontos validados; a Seção 4.4 apresenta as variáveis utilizadas no modelo e o resultado de sua espacialização e finalmente na Seção 4.5 apresenta-se o modelo de probabilidade de erros gerado, juntamente com sua espacialização.

4.1 Mapa de padrões de desmatamento

Para cada uma das 100 classificações foi possível obter o erro estimado a partir das células utilizadas na classificação (75%) e das células separadas para teste (25%). Uma síntese destes erros estimados é apresentada na Tabela 4.1.

Tabela 4.1 - Resumo dos erros da classificação.

Resumo dos erros	%
Menor erro de classificação	6,00
Menor erro de teste	14,00
Maior erro de classificação	16,20
Maior erro de teste	44,40
Média erro de classificação	10,65
Média erro de teste	29,75
Desvio padrão erro de classificação	2,13
Desvio padrão erro de teste	6,06

Fonte: Produção do autor.

Para geração do mapa final com os padrões de desmatamento foram utilizados valores que representassem a moda, ou seja, a classe mais frequente entre as 100 classificações obtidas para cada célula de 50 x 50 km. A partir do mapa

classificado foi possível confeccionar a matriz de confusão e respectivos índices de acurácia do produtor e usuário (Tabela 4.2), comparando os dados da classificação obtida contra um conjunto de 113 células selecionadas visualmente de forma que todos os padrões de desmatamento envolvidos tivessem representação. As células sorteadas foram classificadas manualmente e este conjunto foi considerado como dado de referência. Os valores verificados permitem atestar a alta qualidade do mapa final produzido. As confusões verificadas entre o padrão Multidirecional, principalmente com o padrão Geométrico Regular, são aceitáveis, pois em alguns casos uma mesma célula pode apresentar uma grande heterogeneidade de polígonos, o que dificulta a identificação do padrão de desmatamento ocasionado confusão durante a classificação.

Tabela 4.2 - Matriz de confusão da classificação dos padrões de desmatamento.

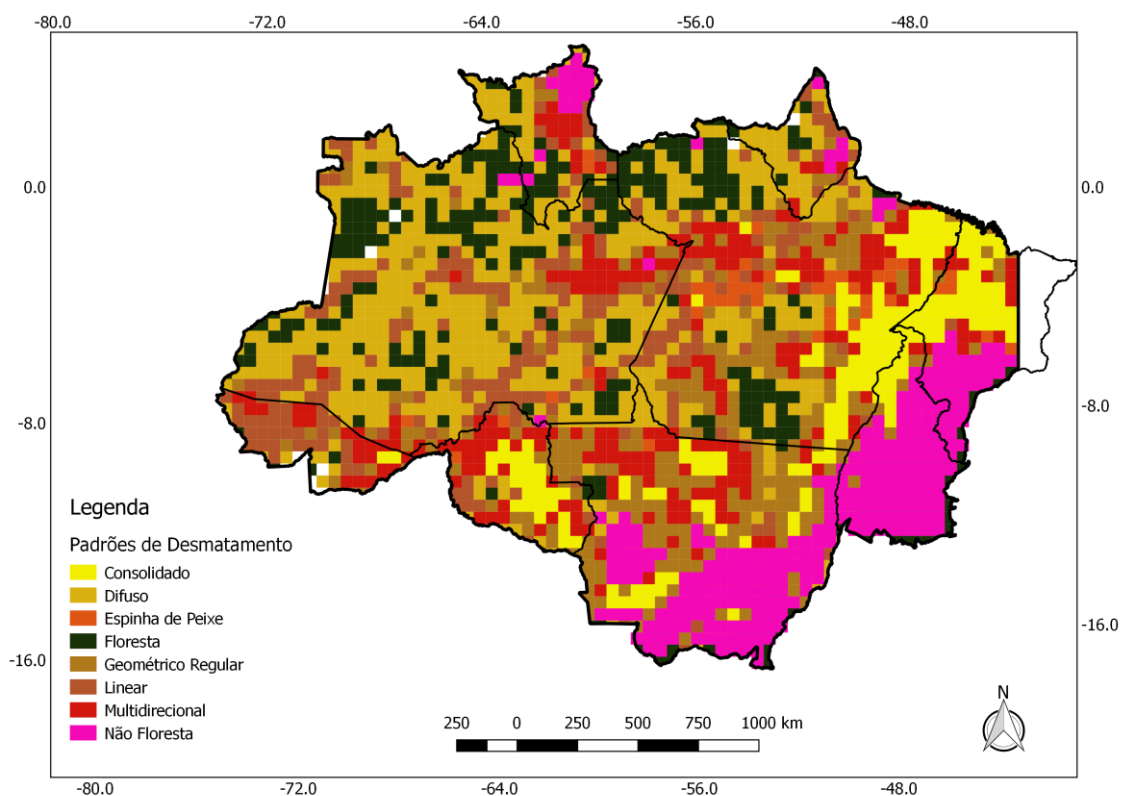
		Referência							N	AU (%)
		Cons	Dif	EspP	Flo	GeomR	Lin	Mult		
Classificação	Cons	13							13	100
	Dif		20						20	100
	EspP			9					9	100
	Flo				9				9	100
	GeomR					21	1	2	24	87
	Lin		1				11	1	13	84
	Mult	1		1		4		19	25	76
	N	14	21	10	9	25	12	22	113	
AP (%)		93	100	90	100	84	91	92		

Cons: Consolidado; Dif: Difuso; EspP: Espinha de Peixe; Flo: Floresta; GeomR: Geométrico Regular; Lin: Linear; Mult: Multidirecional; N: número total de amostras, AP: Acurácia Produtor; AU: Acurácia Usuário.

Fonte: Produção do autor.

O mapa final com os padrões de desmatamento obtido pelo método de classificação por algoritmo de árvore de decisão é apresentado na Figura 4.1.

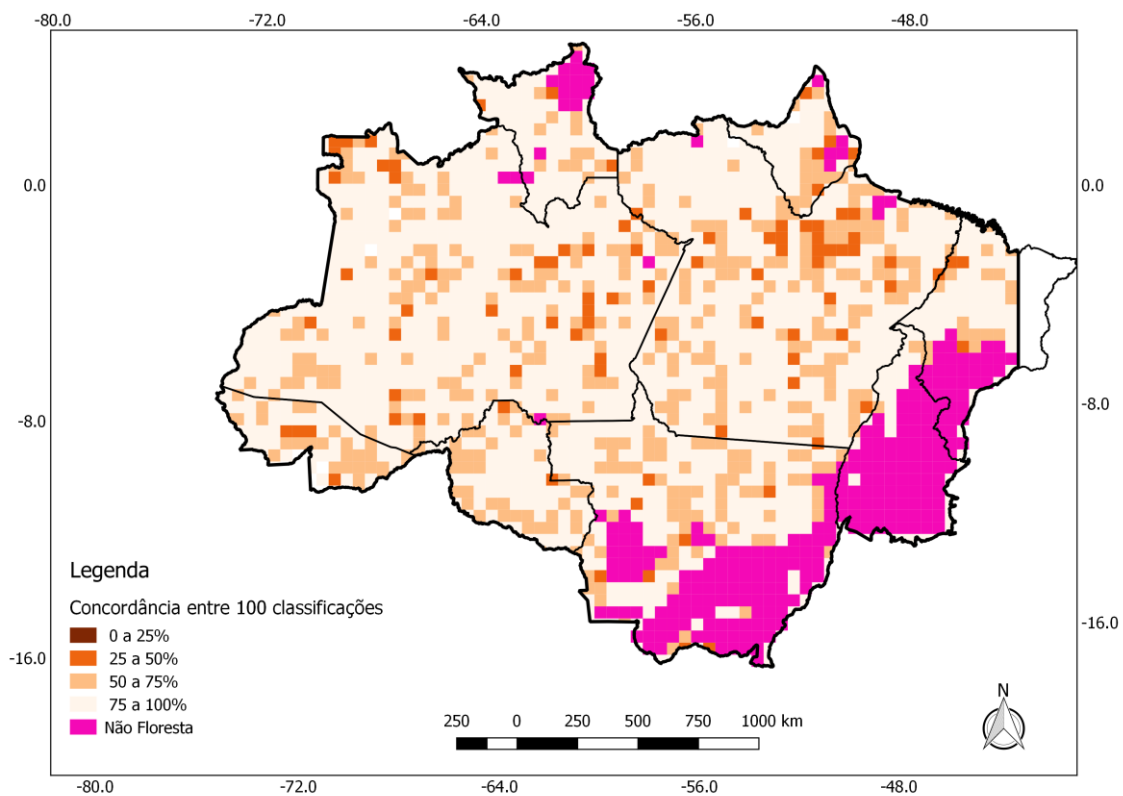
Figura 4.1 - Mapa de padrões de desmatamento em células de 50 x 50 km.



Fonte: Produção do autor.

Outra análise realizada para aferir a qualidade do mapa final obtido foi avaliar, com base nos dados gerados nas 100 classificações, a proporção da classe majoritária para cada célula (valor da moda), neste caso foi somada a quantidade de classes atribuída em cada classificação individual cujo valor era igual ao da moda entre estas classificações. Exemplo: Na célula de ID C35L14, a moda entre as classificações apontou o padrão Consolidado, analisando-se os dados individuais para esta célula nas 100 classificações isoladas, verifica-se que 80 delas foram classificadas como Consolidado, 18 como Geométrico Regular e duas como Multidirecional. Ou seja, a classe majoritária se repetiu em 80 vezes nos resultados das classificações isoladas. O mapa apresentado na Figura 4.2, espacializa esta comparação das células em faixas de 0 a 25%, 25 a 50%, 50 a 75% e 75 a 100%, onde é possível verificar que a grande maioria das células houve concordância superior a 75%.

Figura 4.2 - Mapa com comparação entre as 100 classificações dos padrões de desmatamento.



Fonte: Produção do autor.

Numericamente este resultado é apresentado na Tabela 4.3, onde nota-se que 73% das células estão na faixa de 75 a 100% de conformidade.

Tabela 4.3 - Conformidade entre classificações por faixa.

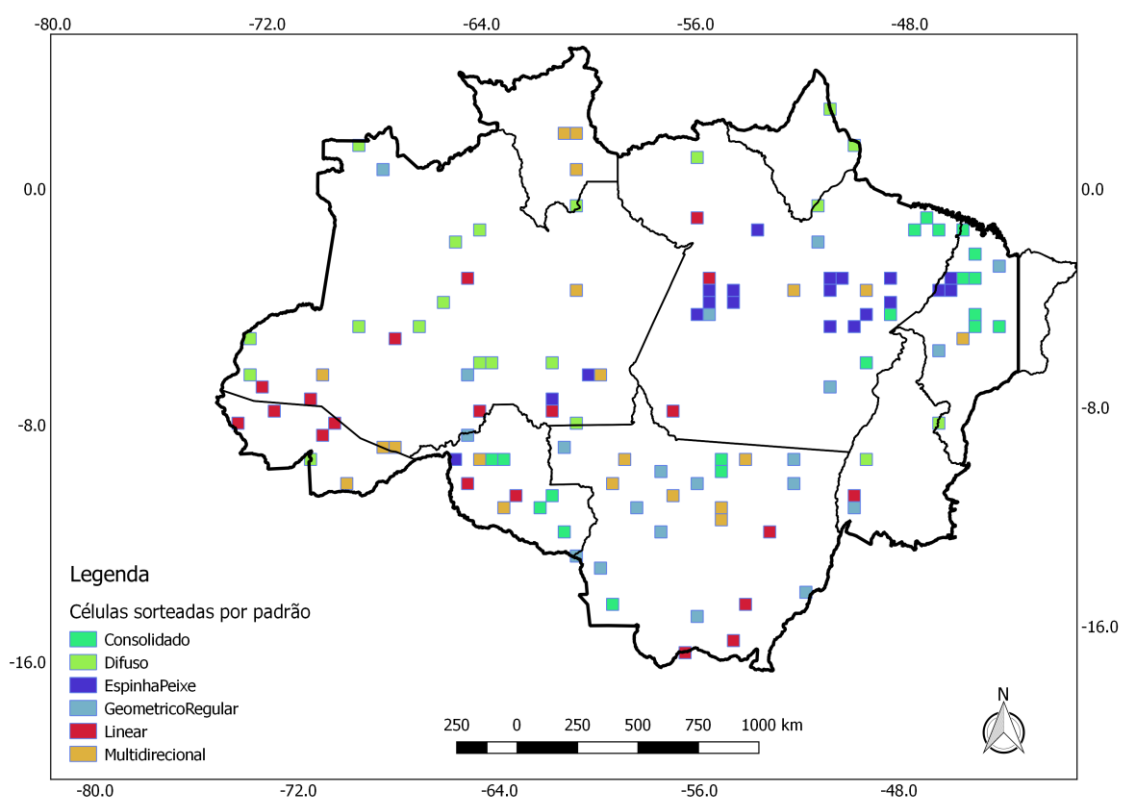
Faixas	Qtde. Células	%
0 a 25%	0	0,0%
25 a 50%	93	4,2%
50 a 75%	502	22,9%
75 a 100%	1596	72,8%
Total	2191	100,0%

Fonte: Produção do autor.

4.2 Amostragem

Para confecção do painel amostral dos pontos utilizados na validação foi realizado um sorteio aleatório de 120 células sendo 20 para cada padrão de desmatamento identificado no mapa de tipologia gerado. A Figura 4.3 mostra a distribuição espacial das células sorteadas com seu respectivo padrão de desmatamento.

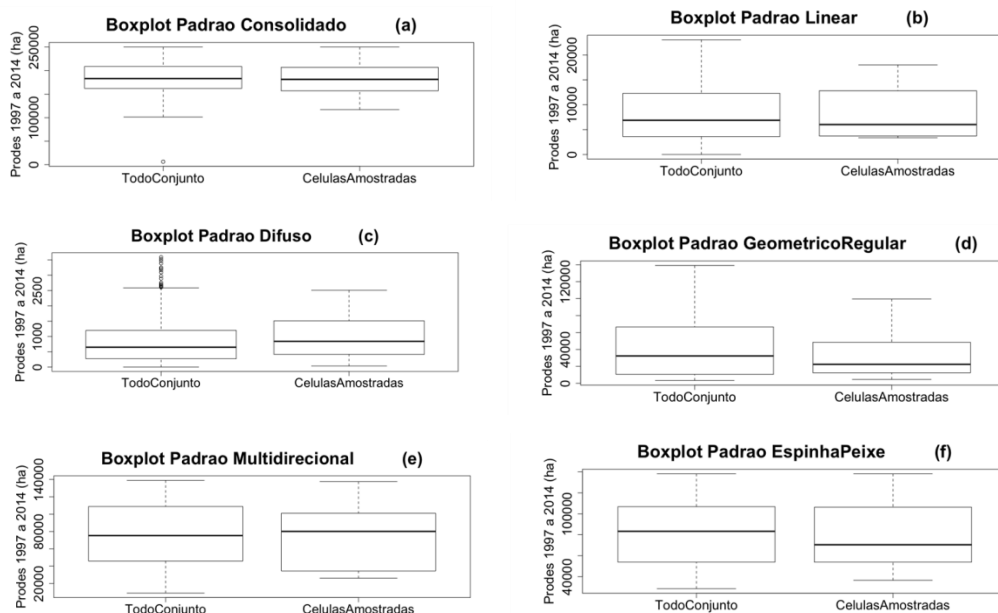
Figura 4.3 - Distribuição das 120 células sorteadas aleatoriamente.



Fonte: Produção do autor.

Analisando o total de desmatamento do PRODES acumulado em cada célula até 2014, a Figura 4.4 (a) a (f) mostra os *boxplots* para cada padrão de desmatamento utilizado no sorteio das células comparando o comportamento das células sorteadas em relação ao restante das células.

Figura 4.4 - Diagrama *boxplot* para cada padrão de desmatamento.



Fonte: Produção do autor.

É possível verificar que para todos os padrões de desmatamento, as células amostradas apresentam comportamento similar, quando analisado o total desmatamento até 2014, em relação às demais as células. Ou seja, as amostras apresentam comportamento semelhante (em termos de área desmatada) em relação ao restante do conjunto celular dos padrões, evidenciando que a amostragem é representativa dos padrões e apresenta homogeneidade.

4.3 Análises preliminares sobre os pontos validados

Com base nos pontos amostrais validados foi possível calcular, para cada padrão de desmatamento, índices de erros verificados entre a classe original do mapeamento PRODES de cada ponto amostral com a classe majoritária atribuída pelos auditores no processo de validação. A Tabela 4.4 apresenta os resultados desta análise. Destaca-se que dentre os 1374 pontos válidos avaliados, em 92,1% houve concordância total entre as três classificações feitas pelos auditores, enquanto em 7,9% houve uma classificação discordante.

Tabela 4.4 - Análise dos erros por padrão de desmatamento.

Padrão de desmatamento	Nº		% erro
	Pontos válidos	Qtde. Erros	
Difuso	104	0	0,00
Linear	241	1	0,41
GeométricoRegular	242	6	2,48
Multidirecional	285	2	0,70
Consolidado	244	6	2,46
EspinhaPeixe	258	3	1,16
Totais	1374	18	1,13

Fonte: Produção do autor.

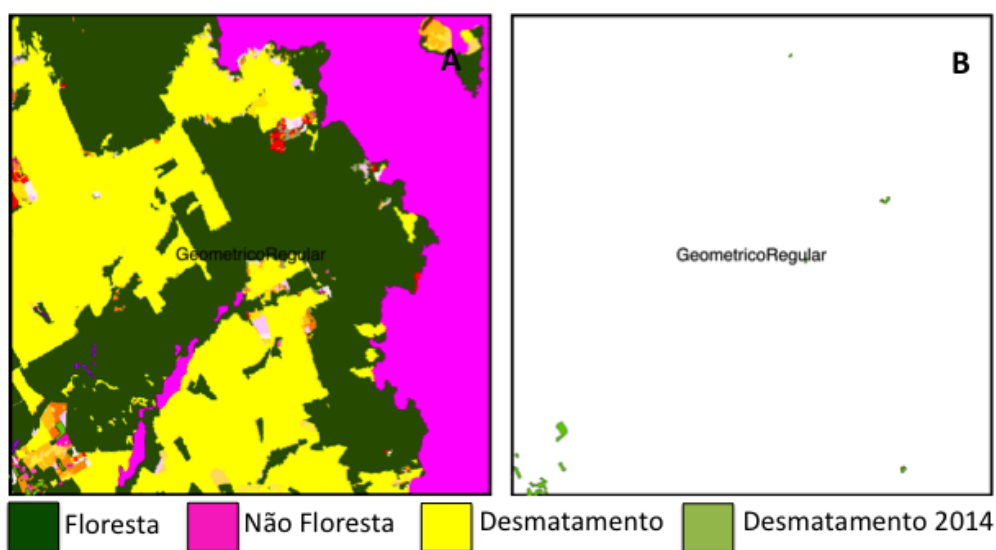
A metodologia de construção do painel amostral dos pontos para validação não foi adequada para confeccionar índices globais de precisão do mapeamento PRODES, já que só foram sorteados pontos com base nos polígonos de desmatamento de 2014, visando estimar possíveis erros de inclusão. Com isso não foi considerado analisar eventuais erros de omissão, pois não foram amostrados pontos sobre a classe de floresta. De qualquer forma, com base nos dados da Tabela 4.4 é possível afirmar que a precisão do PRODES, se analisado somente os erros de mapeamento, é bastante alta, chegando a 98,87% sendo de somente 1,13% seu erro de inclusão.

Alguns percentuais de erros relatados na Tabela 4.4 merecem ser analisados com mais detalhe, pois apresentam valores que não condizem com as suposições estabelecidas neste estudo (ver Seção 1.2), de que padrões que apresentam menor complexidade, apresentariam menores probabilidades de erro, como é o caso do padrão Geométrico Regular. Isso não se verificou no ano de análise. Essa tabela mostra que o padrão Geométrico Regular apresenta uma porcentagem de erro muito próxima ao padrão Consolidado, que é mais heterogêneo e apresenta polígonos com formas complexas e variadas. Dois exemplos foram selecionados para ilustrar os casos em que o

erro obtido na validação foi superior ao esperado. Figura 4.5 mostra uma célula classificada como padrão Geométrico Regular, que apresentou 13% de erro, bem acima do esperado.

Célula ID C31L08 – padrão Geométrico Regular: Nesta célula foram validados 15 pontos que apresentaram um erro de 13% (Figura 4.5 A). O padrão Geométrico Regular verificado mostra o desmatamento acumulado até 2013 cujos polígonos de desmatamento apresentam um padrão bem característico, predominando aqueles com forma regular e de tamanho grande. Porém, os polígonos de desmatamento mapeados em 2014 apresentam áreas pequenas (média de 29 ha – Figura 4.5 B), na região de borda das áreas previamente desmatadas. Esses polígonos de desmatamento, por serem pequenos, por apresentarem formas mais complexas e por estarem muito próximos de áreas previamente desmatadas, podem ter apresentado um grau maior de dificuldade em seu mapeamento. Juntamente com a tendência de queda do desmatamento apresentada no período de 2004 até os dias atuais, tem-se observado uma tendência de redução dos polígonos de desmatamento, o que pode ter implicado no aumento dos erros em padrões como o Geométrico Regular.

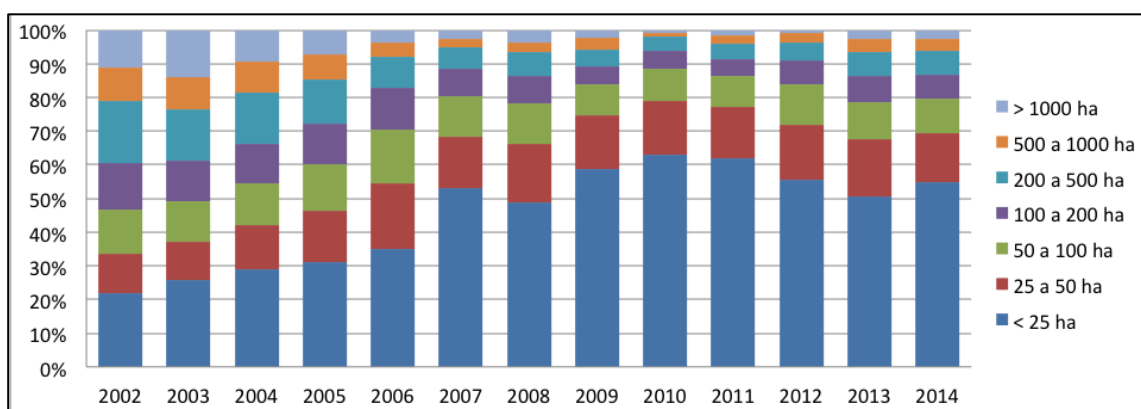
Figura 4.5 - Célula padrão Geométrico Regular.



Fonte: Produção do autor.

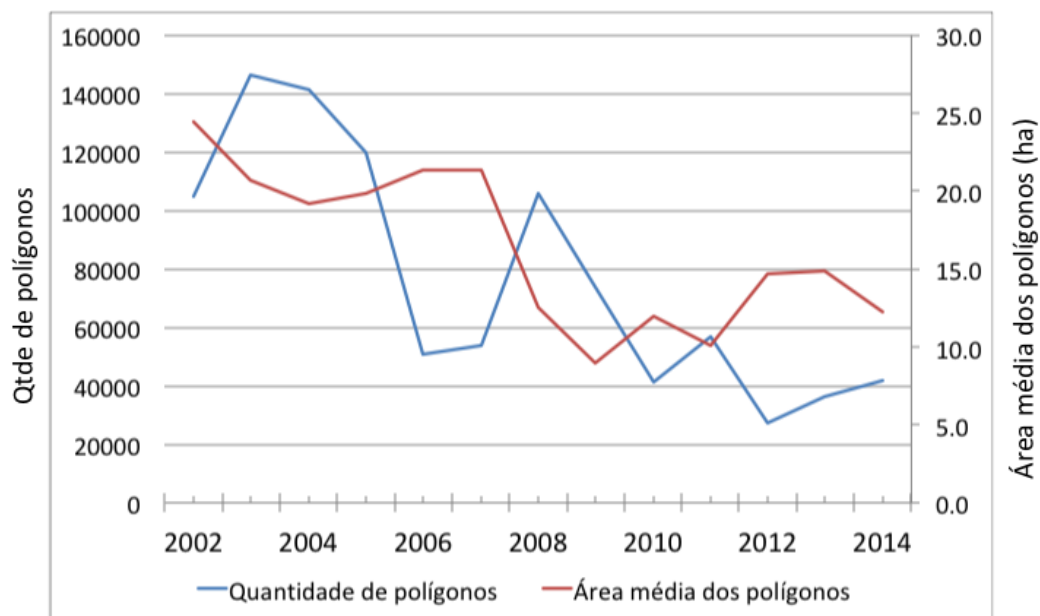
Assim como no exemplo apresentado na Figura 4.5, a presença de grande proporção de pequenos polígonos de desmatamento no ano de 2014 para a área de estudo como um todo, pode ter provocado o aumento dos erros associados ao padrão Geométrico. A Figura 4.6 representa a contribuição dos polígonos de desmatamento por faixas de tamanho para o período de 2002 a 2014. É possível observar neste gráfico que a contribuição dos polígonos pequenos, menores que 25 ha, aumentam ao longo da série temporal. Em 2002, esta faixa de tamanho representava 22% do total dos polígonos mapeados pelos PRODES, enquanto que em 2014 este percentual chegou a 57%. Com base nos mesmos dados é possível afirmar que o desmatamento na Amazônia Legal tem se tornado um evento pouco frequente em função da diminuição da quantidade de polígonos detectados a cada ano. Em 2002, foram mapeados 104.681 polígonos (média de 24,5 ha/polígono), enquanto que em 2014, foram mapeados 41.729 polígonos (média de 12,3 ha/polígono), apresentando uma redução expressiva de 40%, conforme apresentado na Figura 4.7. Presume-se que esses polígonos pequenos induzam a uma maior probabilidade de erro, uma vez que sua detecção é dificultada pelo seu tamanho pequeno, se aproximando dos limites de detecção com a resolução espacial dos sensores utilizados no mapeamento.

Figura 4.6 - Distribuição dos polígonos PRODES por faixa de tamanho.



Fonte: Produção do autor.

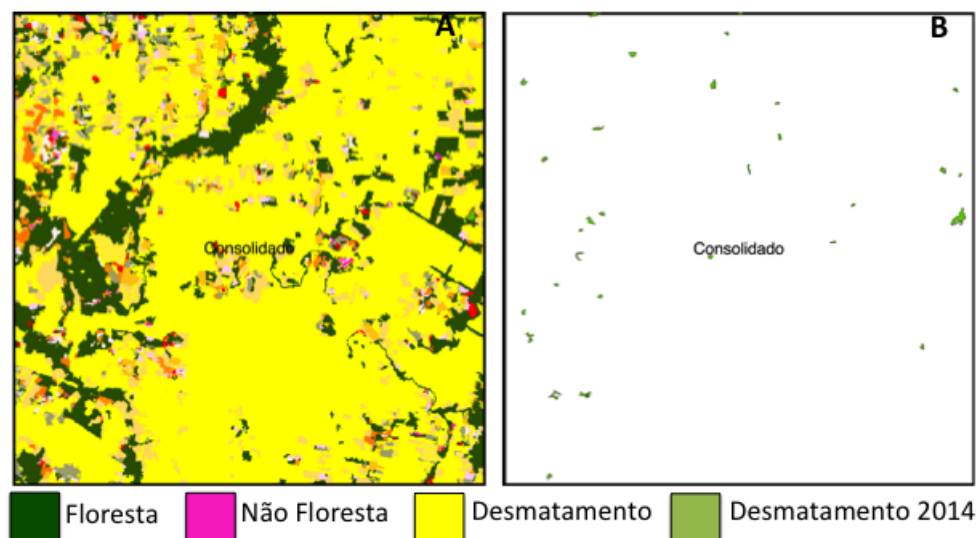
Figura 4.7 - Distribuição dos polígonos PRODES por quantidade e área média dos polígonos.



Fonte: Produção do autor.

Célula ID C41L17 – padrão Consolidado: Nesta célula foram validados 16 polígonos que apresentaram um erro de 12,5% (Figura 4.8 A). O padrão Consolidado verificado nesta célula mostra uma região com desmatamento acumulado até 2013 bem característico, com pouca floresta remanescente. Os 39 polígonos de desmatamento mapeados em 2014 nesta célula apresentam área pequena (média de 14 ha – Figura 4.8 B) e encontravam-se localizados em região de borda das áreas previamente desmatadas. Estes fatores podem implicar no aumento da quantidade de erros, por representarem mudanças sutis, com um grau maior de dificuldade de detecção.

Figura 4.8 - Célula padrão Consolidado.



Fonte: Produção do autor.

4.4 Preparação das variáveis para o modelo estatístico

O objetivo principal deste estudo foi elaborar um modelo estatístico multivariado ajustado para gerar uma superfície contínua que aponte regiões com maiores e menores probabilidades de erro de detecção de desmatamento para a região da Amazônia Legal. A variável dependente Y que se pretende analisar representa a proporção de erro de classificação estimado para cada célula. Nas 120 células sorteadas para a validação mapeamento, a estimativa de erro foi obtida avaliando a proporção de pontos sorteados em cada célula onde houve discordância entre a classificação do PRODES e a moda das três classes indicadas pelos auditores independentes para cada padrão de desmatamento. Células com menos de três pontos avaliados foram descartadas, assim como células se encontravam em área de domínio da classe Não Floresta, pois estas são regiões não mapeadas pelo projeto.

4.5 Modelagem estatística multivariada

Os procedimentos de modelagem têm como objetivo o desenvolvimento de modelo estatístico multivariado para obtenção da variável que possa representar para cada célula de 50 x 50 km, a probabilidade de verificação de

erros no mapeamento do PRODES 2014 com base no conjunto de variáveis independentes confeccionadas neste estudo.

4.5.1 Modelo de regressão logístico

Antes de iniciar a construção do modelo de regressão logístico, avaliou-se a colinearidade presente entre as variáveis independentes através do índice *VIF*. Iterativamente, foram descartadas aquelas variáveis que apresentaram *VIF* superior a 100, uma a uma, recalculando-se os valores a cada retirada. Optou-se por um valor bastante elevado para que não fossem descartadas prematuramente muitas variáveis. Nesta primeira etapa, foram descartadas as variáveis P_Floresta (*VIF* = 616,45) e P_Prodes2014Masc (*VIF* = 118,54).

Em seguida, foram obtidos dois modelos a partir do método *stepwise* usando as abordagens *forward* e *backward*, cujos resultados são apresentados respectivamente nas Tabelas 4.5 e 4.6. A abordagem *both* também foi testada, apresentando resultado idêntico à abordagem *backward*. É possível observar que a abordagem *forward* resultou num modelo bastante simplificado com apenas uma variável independente (P_Prodes2014Flor) significativa a 5%. Este modelo apresentou um valor *AIC* de 98,07, pouco inferior ao modelo nulo (sem nenhuma variável independente) cujo *AIC* foi de 101,19. Já a abordagem *backward* resultou num modelo bastante complexo, com 13 variáveis independentes e *AIC* igual a 92,14, valor bem menor do que o primeiro modelo. No entanto, é possível observar que este segundo modelo apresenta muitos coeficientes com baixa significância e variáveis com alto valor de *VIF*, o que não justifica seu uso para estimação da probabilidade erro. O menor valor de *AIC* deste modelo em relação ao primeiro indica que possivelmente é possível obter modelos que melhor representem a probabilidade de erro. Dessa forma, procedeu-se a busca exaustiva para encontrar este melhor modelo.

Tabela 4.5 - Modelo de regressão logística selecionado pelo método *stepwise forward*.

	Estimativa	Valor p
Intercepto	-4,71	< 2e-16
P_Prodes2014Flor	37,36	0,0081

Fonte: Produção do autor.

Tabela 4.6 - Modelo de regressão logística pelo método *stepwise backward*.

	Estimativa	Valor p	VIF
Intercepto	-125,10	0,1478	
P_Prodes2014	257,50	0,0032	5,22
PLAND	0,21	0,0012	45,53
PD	2,39	0,0062	27,93
MPS	-0,0029	0,1544	14,05
LSI	-0,54	0,0144	53,69
AWMSI	0,52	0,0300	45,26
MPFD	95,07	0,1539	9,33
MPAR	-155,70	0,0554	10,13
PSCOV	-0,0062	0,0245	42,07
IJI	0,0051	0,0809	7,29
TABO	-2,7E-06	0,0075	5,31
PRD	-3085,00	0,1370	5,91
SHEI	3,21	0,0249	3,03

Fonte: Produção do autor.

As Tabelas 4.7, 4.8, 4.9 e 4.10 apresentam os modelos obtidos por busca exaustiva considerando-se 1, 2, 3 ou 4 variáveis independentes. Na busca exaustiva, todas as combinações de variáveis foram testadas. É importante

ressaltar que todos os modelos testados que apresentassem qualquer uma das variáveis independentes com *VIF* maior que 4, foram descartados evitando-se assim que modelos não apropriados fossem escolhidos durante este processo. Os valores *AIC* para cada uma dos modelos selecionados foi respectivamente: 98,07, 96,34, 94,11 e 92,78.

Tabela 4.7 - Modelo de regressão logística com 1 variável independente selecionado por busca exaustiva.

	Estimativa	Valor p
Intercepto	-4,71	< 2e-16
P_Prodes2014Flor	37,36	0,0081

Fonte: Produção do autor.

Tabela 4.8 - Modelo de regressão logística com 2 variáveis independentes selecionado por busca exaustiva.

	Estimativa	Valor p	<i>VIF</i>
Intercepto	-4,98	< 2e-16	
PLAND	257,50	0,0039	2,73
TABO	3,21	0,0401	2,73

Fonte: Produção do autor.

Tabela 4.9 - Modelo de regressão logística com 3 variáveis independentes selecionado por busca exaustiva.

	Estimativa	Valor p	<i>VIF</i>
Intercepto	-5,20	< 2e-16	
P_Prodes2014Flor	38,21	0,0215	1,28
PLAND	3,94e-2	0,0087	2,82
TABO	-1,88e-6	0,0267	3,15

Fonte: Produção do autor.

Tabela 4.10 - Modelo de regressão logística com 4 variáveis independentes selecionado por busca exaustiva.

	Estimativa	Valor p	VIF
Intercepto	-4,40	< 4e-9	
P_Prodes2014Flor	33,31	0,0396	1,33
PLAND	4,82e-2	0,0031	2,92
MPAR	-33,38	0,1421	1,15
TABO	-1,92e-6	0,0212	3,13

Fonte: Produção do autor.

É importante notar que o modelo com quatro variáveis selecionadas (Tabela 4.10), apesar de possuir valor *AIC* menor, apresentou o coeficiente relativo à variável MPAR não significativo (Valor p superior a 5%). Desse modo, determinou-se que o melhor modelo foi aquele com três variáveis independentes (Tabela 4.9), ou seja, o modelo que melhor estimou a probabilidade de erro foi:

$$P(\text{erro}) = \frac{e^{-5,20+38,21 \cdot P_Prodes14Flor+3,94 \cdot PLAND/100-1,88 \cdot TABO/1000000}}{1 + e^{-5,20+38,21 \cdot P_Prodes14Flor+3,94 \cdot PLAND/100-1,88 \cdot TABO/1000000}} \quad (4.1)$$

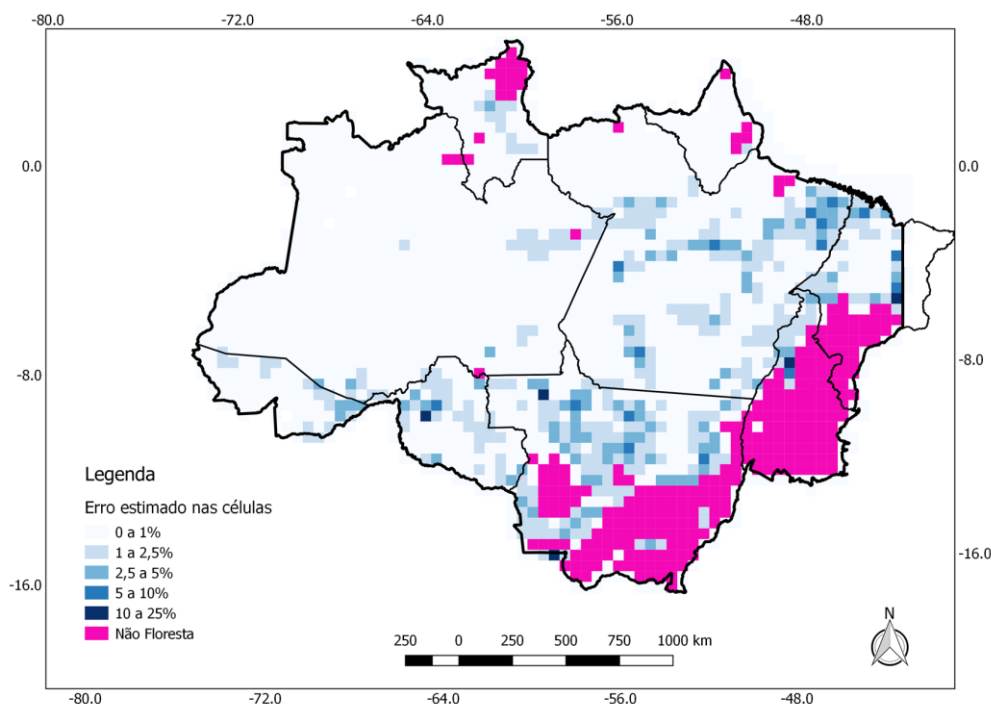
onde P_ProdesFlor corresponde a porcentagem de desmatamento 2014 em relação à área de floresta na célula, PLAND corresponde a porcentagem da área desmatada na célula e TABO indica a área (ha) do maior polígono desmatamento presente na célula.

4.5.2 Espacialização do modelo estatístico

Aplicando a Equação 4.1 para todo conjunto de células de 50 x 50 km foi possível modelar e estimar a probabilidade se obter erro no mapeamento do PRODES, sendo 0 (0%) interpretado como baixa probabilidade de ocorrência

de erro e 1 (100%) como alta probabilidade. O mapa da Figura 4.9 apresenta a espacialização das células estratificadas por faixa de erro.

Figura 4.9 - Mapa da probabilidade de erros estimados por célula.

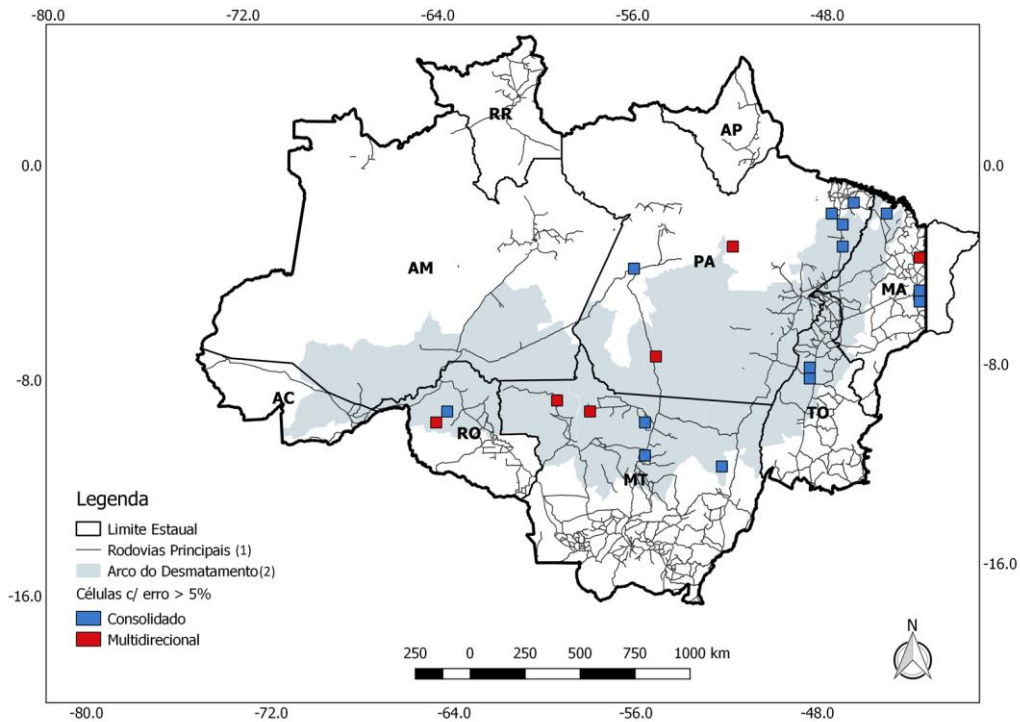


Fonte: Produção do autor.

Analisando o mapa verifica-se que a grande maioria das células válidas que compõem a área de estudo tem erro estimado abaixo de 1%, ou seja, pelo modelo de regressão aplicado, nestas células há pouca probabilidade de se encontrar um erro no mapeamento do PRODES de 2014. É possível também identificar que as 20 células onde se concentram as maiores faixas de erro - superior a 5%, são próximas às regiões onde historicamente se concentram áreas com desmatamento já estabelecido na Amazônia Legal, conhecido como arco do desmatamento e a região de influência das rodovias BR-163, BR-2030 (Transamazônica) e BR-010 (Belém-Brasília). A Figura 4.10 apresenta a localização dessas células divididas por padrão de desmatamento, sendo 14 células correspondentes ao padrão Consolidado e 6 ao padrão Multidirecional. Estes padrões apresentam uma variável significativa do modelo que indica a

alta porcentagem de área total desmatada na célula (PLAND), que para este conjunto de 21 células variou de 37% a 94% (média de 68%).

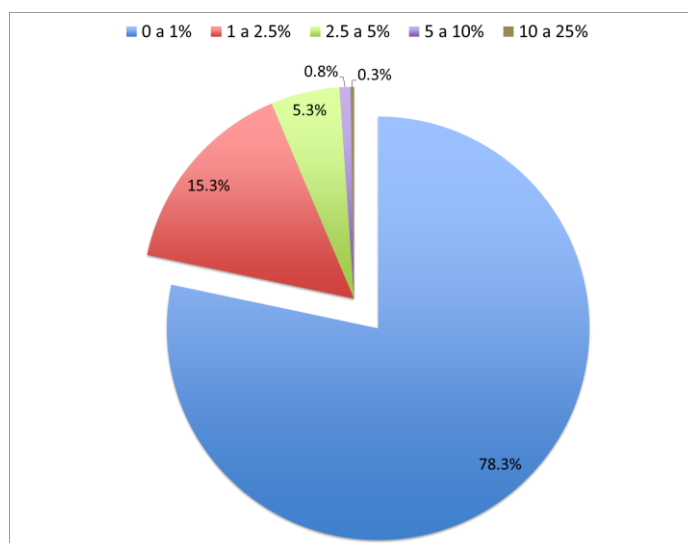
Figura 4.10 - Localização das 20 células com erro estimado maior que 5%.



Fonte: Produção do autor.

A Figura 4.11 exibe a distribuição do número de células válidas por faixa de erro estimado pelo modelo de regressão, excluindo-se as células com predomínio da classe Não Floresta. Verifica-se que 78% destas células apresentam erros menores que 1% (1496 de 1913 células), enquanto somente 0,3% (5 de 1913 células) apresentam erros estimados maiores que 10%.

Figura 4.11 - Proporção de células por faixa de erro estimado.



Fonte: Produção do autor.

A Tabela 4.11 mostra a distribuição dos erros estimados nas células por faixa de probabilidade para cada padrão de desmatamento. É possível observar que a faixa de células com menor erro estimado (0 a 1%) concentra grande parte das células (78%), independente do padrão do desmatamento nela encontrado, sendo o padrão Difuso o que corresponde ao maior percentual nesta faixa (28%). Para as faixas cujos erros estimados são mais altos - maior que 10%, por exemplo, o número de células é baixo (0,3%), independente do padrão. Estes resultados confirmam que no PRODES 2014 são baixas as probabilidades de verificação de erros no mapeamento. Na seção seguinte são apresentados os erros estimados pelo modelo nas células que representam diferentes padrões e proporções de erro. O comportamento das variáveis que mais influenciaram os resultados é analisado e discutido para cada padrão.

Tabela 4.11 - Distribuição dos erros estimados por faixa de probabilidade.

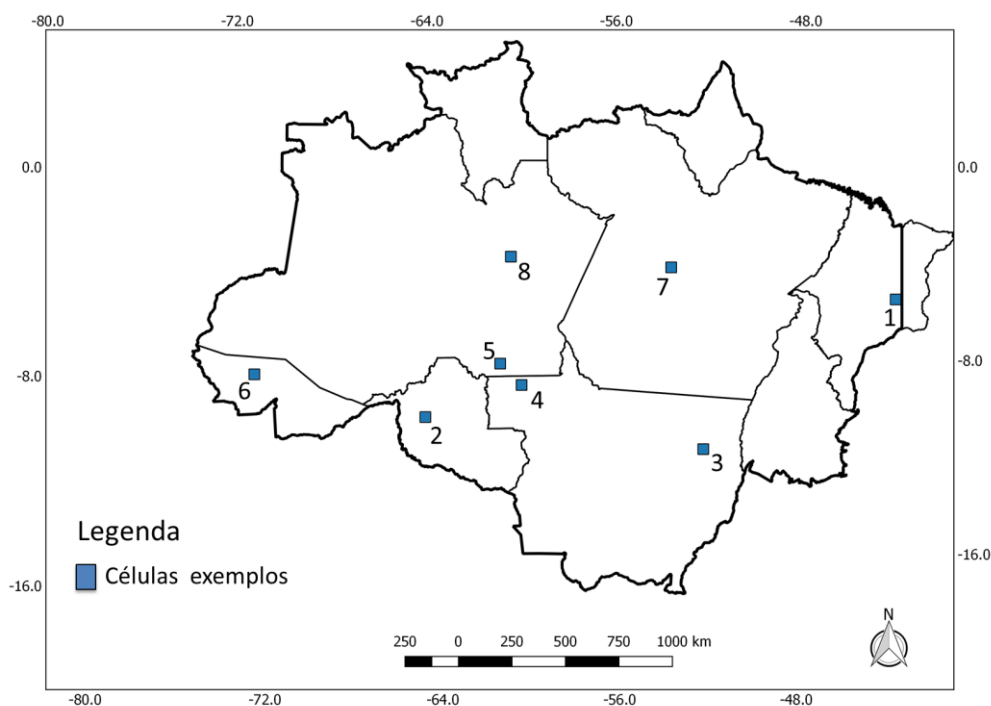
Tipologia	0 a 1%	1 a 2,5%	2,5 a 5%	5 a 10%	10 a 25%
Linear	13,3	0,1	0,1	0,0	0,1
Consolidado	2,6	4,0	2,4	0,6	0,1
Multidirecional	4,7	6,3	1,8	0,2	0,1
GeometricoRegular	10,2	3,9	0,6	0,0	0,0
EspinhaPeixe	0,9	0,9	0,4	0,0	0,0
Difuso	28,4	0,1	0,0	0,0	0,0
Floresta	18,1	0,0	0,0	0,0	0,0
Total	78,3	15,3	5,3	0,8	0,3

Fonte: Produção do autor.

4.5.3 Análise das células por tipologia

Esta seção apresenta uma análise individual de algumas células para avaliar e discutir a aplicação e os resultados gerados pelo modelo. Foram selecionadas três células com erros estimados maiores que 5% que apresentaram padrão de desmatamento Consolidado, Geométrico Regular e Multidirecional. Também foram selecionadas cinco células com erro estimado menor que 1%. A localização das células utilizadas nesta análise é mostrada na Figura 4.12.

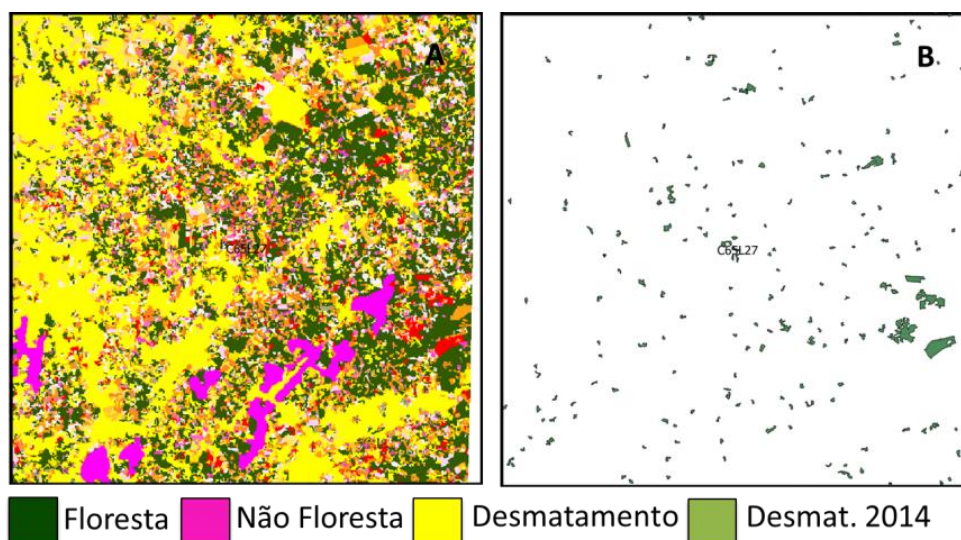
Figura 4.12 - Localização das células analisadas individualmente.



Fonte: Produção do autor.

Célula nº 1 (localização na Figura 4.12): ID C65L27 – padrão Consolidado: Esta célula apresenta um erro estimado considerado alto de 0,16. Na Figura 4.13 (A) é possível verificar que a região apresenta desmatamento acumulado de grande extensão, indicado pela variável PLAND (68%). Nesta célula 199 polígonos de desmatamento foram detectados com área média de 16 ha – Figura 4.13 (B). Uma das variáveis de maior peso no modelo foi a P_Prodes2014Flor (0,049) que representa a relação da área desmatada em 2014 sobre a área de floresta remanescente e uma relação negativa com o erro. Verifica-se que a área apresenta alta complexidade dos polígonos, cujas formas são bastante irregulares, este fato foi capturado pela métrica AWMSI (40,2) que, todavia não foi selecionada pelo modelo. Células nestas condições apresentam uma alta probabilidade de verificação de erro no mapeamento dos polígonos.

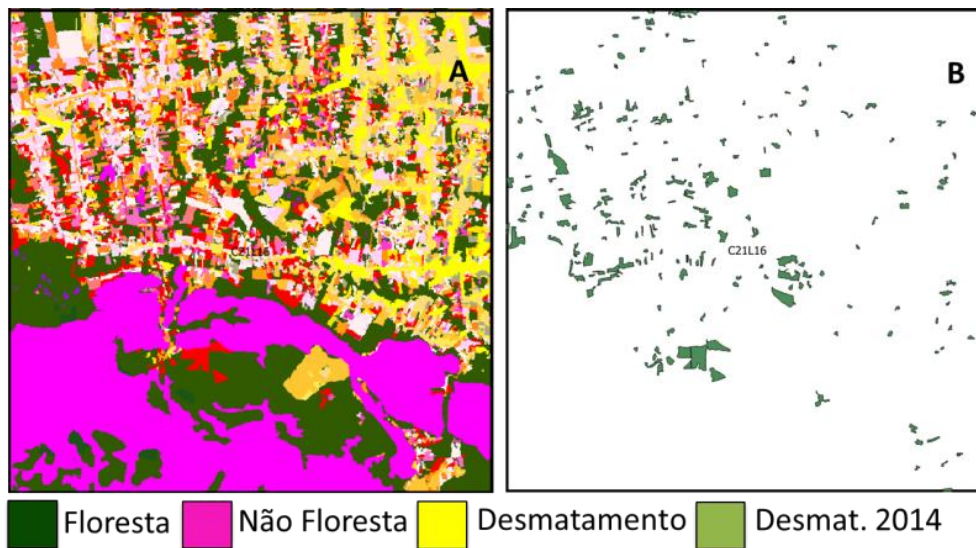
Figura 4.13 - Célula padrão Consolidado.



Fonte: Produção do autor.

Célula n° 2 (localização na Figura 4.12): ID C21L16 – padrão Multidirecional: Esta célula apresenta um erro estimado alto de 0,11. Na Figura 4.14 (A) é possível verificar que a região apresenta desmatamento intenso e desordenado em grande parte da célula - indicado pela variável PLAND (48%) e poucos remanescentes florestais. Foram detectados 220 polígonos de desmatamento de 2014 com área média de 28 ha - Figura 4.14 (B). A variável P_Prodes2014Flor (0,09) indica um razoável desmatamento de 2014 em relação ao remanescente florestal. Esta configuração espacial indica, baseado no modelo de regressão, probabilidade alta de encontrar erro no mapeamento do desmatamento.

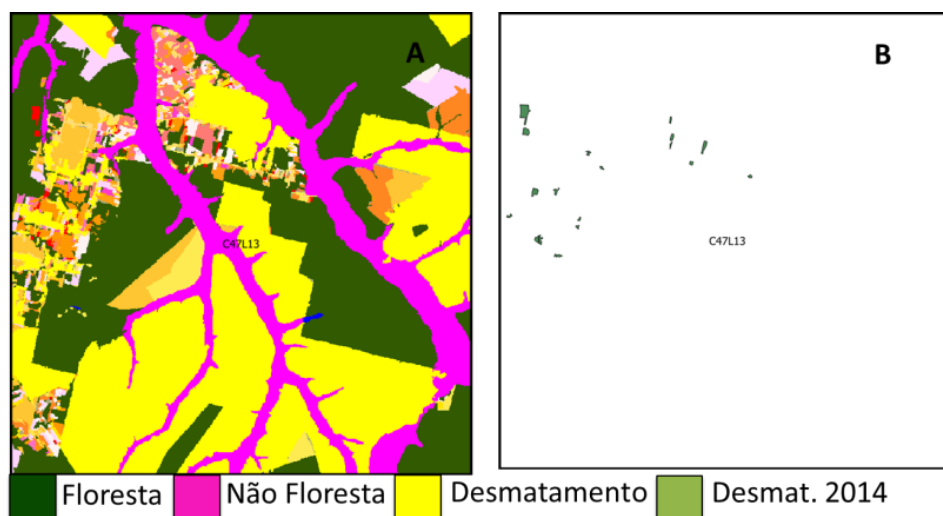
Figura 4.14 - Célula padrão Multidirecional.



Fonte: Produção do autor.

Célula n° 3 (localização na Figura 4.12): ID C47L13 – padrão Geométrico Regular: Esta célula apresenta um erro considerado mediano que foi estimado em 0,05. Na Figura 4.15 (A) é possível verificar que a região apresenta desmatamento bem estabelecido e em formato regular em grande parte da célula - indicado pela variável PLAND (49%). Foram detectados 21 polígonos de desmatamento de 2014 nesta célula com média é de 20 ha – Figura 4.15 (B). A variável P_Prodes2014Flor indica baixa porcentagem de desmatamento de 2014 sobre a área de floresta (0,0049). Células com poucas e pequenas áreas desmatadas contém um valor baixo dessa variável, que associado à alta porcentagem de desmatamento acumulado apresenta, pelo modelo, uma média probabilidade de verificação de erro.

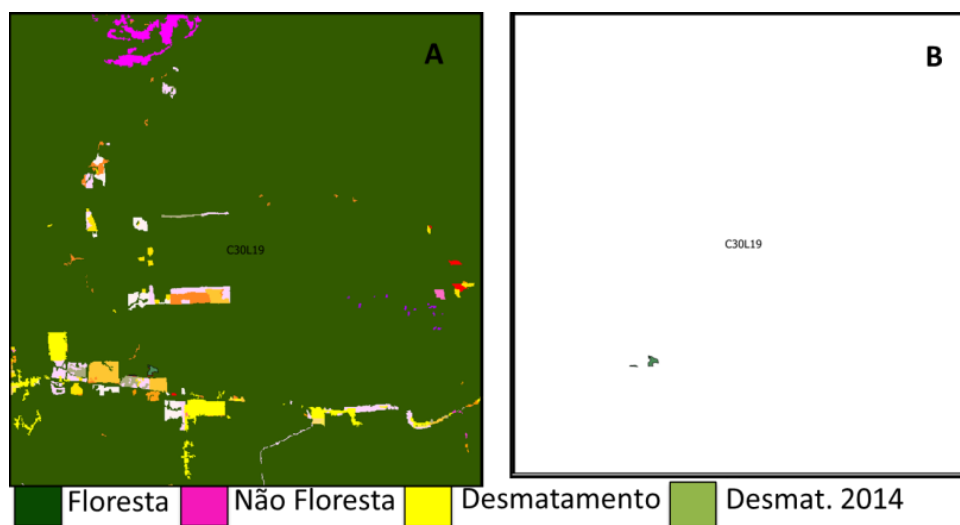
Figura 4.15 - Célula padrão Geométrico Regular (exemplo 1).



Fonte: Produção do autor.

Célula n° 4 (localização na Figura 4.12): ID C30L19 – padrão Geométrico Regular: Esta célula apresenta um baixo erro estimado de 0,006. Na Figura 4.16 (A) é possível verificar que a região apresenta pouco desmatamento e em formato regular em grande parte da célula - indicado pela variável PLAND (3,8%) e alta porcentagem de floresta. A variável P_Prodes2014Flor indica baixa porcentagem de desmatamento de 2014 sobre a área de floresta (0,0003). Foram detectados 2 polígonos de desmatamento de 2014 nesta célula – Figura 4.16 (B), cuja área média é de 36 ha. O baixo percentual de desmatamento e consequentemente a alta concentração de floresta na célula apontado nas variáveis do modelo indica a baixa probabilidade de encontrar erro no mapeamento do desmatamento nesta célula.

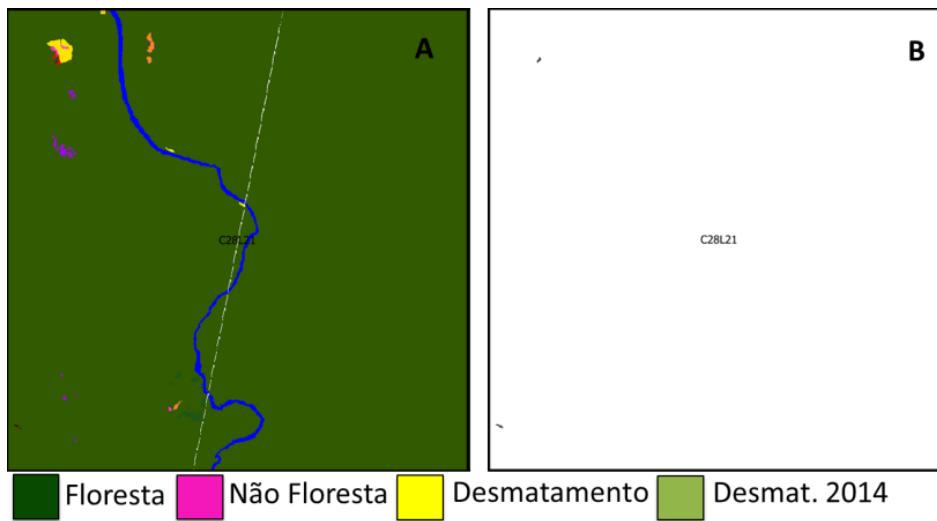
Figura 4.16 - Célula padrão Geométrico Regular (exemplo 2).



Fonte: Produção do autor.

Célula nº 5 (localização na Figura 4.12): ID C28L21 – padrão Difuso: Esta célula apresenta um baixo erro estimado de 0,005. Na Figura 4.17 (A) é possível verificar que a região apresenta pouco desmatamento e distribuído difusamente no interior da célula, indicado pela variável PLAND (0,29%). A variável P_Prodes2014Flor indica baixa porcentagem de desmatamento 2014 sobre a área de floresta (0,0001) e alta porcentagem de floresta. Foram detectados dois polígonos de desmatamento de 2014 nesta célula – Figura 4.17 (B), cuja área média é de 7 ha. O baixo percentual de desmatamento e consequentemente a alta concentração de floresta na célula apontado nas variáveis do modelo indica a baixa probabilidade de encontrar erro no mapeamento do desmatamento nesta célula.

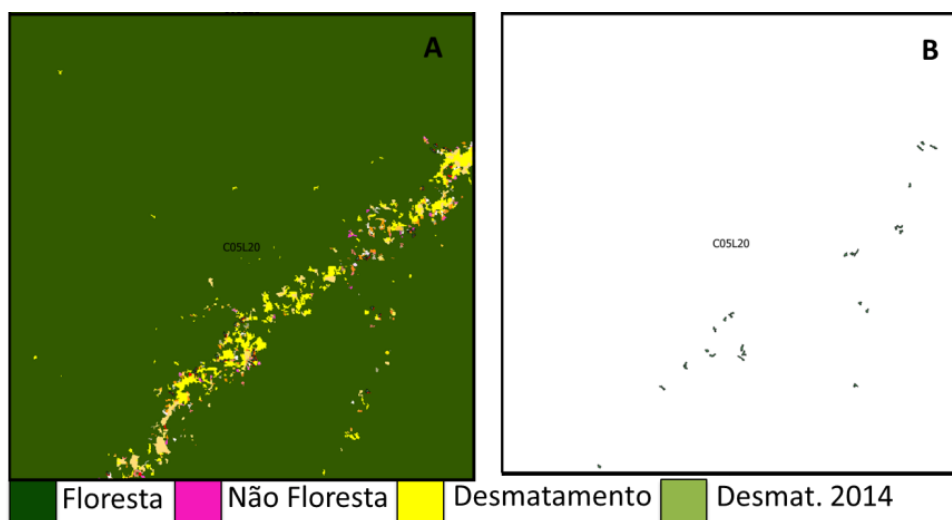
Figura 4.17 - Célula padrão Difuso.



Fonte: Produção do autor.

Célula nº 6 (localização na Figura 4.12): ID C05L20 – padrão Linear: Esta célula apresenta um baixo erro estimado de 0,006. Na Figura 4.18 (A) é possível verificar que a região apresenta pouco desmatamento e distribuído de forma linear no interior da célula, indicado pela variável PLAND (3%), já a variável P_Prodes2014Flor indica baixa porcentagem de desmatamento de 2014 sobre a área de floresta (0,0009). Foram detectados 23 polígonos de desmatamento de 2014 – Figura 4.18 (B), cuja área média é de 9 ha. Esta configuração espacial indica baseado no modelo de regressão, baixa probabilidade de encontrar erro no mapeamento do desmatamento nesta célula.

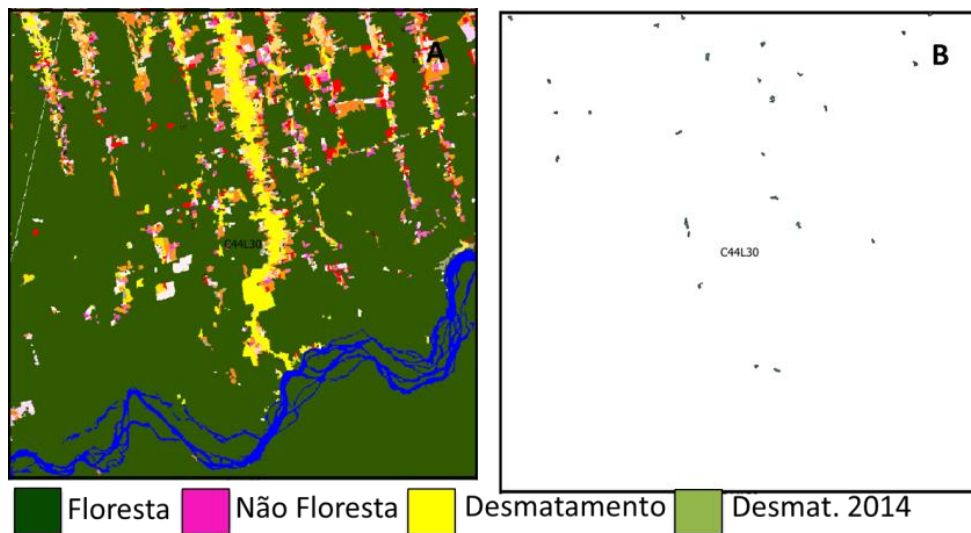
Figura 4.18 - Célula padrão Linear.



Fonte: Produção do autor.

Célula nº 7 (localização na Figura 4.12): ID C44L30 – padrão Espinha de Peixe:
Esta célula apresenta um baixo erro estimado de 0,005. Na Figura 4.19 (A) é possível verificar que a região apresenta desmatamento distribuído de forma linear e alongado no interior da célula, calculado em 15% pela variável PLAND. Já a variável P_Prodes2014Flor indica baixa porcentagem de desmatamento de 2014 sobre a área de floresta (0,0014). Foram detectados 37 polígonos de desmatamento de 2014 nesta célula – Figura 4.19 (B), cuja área média é de 7 ha. Esta configuração espacial indica baseado no modelo de regressão, baixa probabilidade de encontrar erro no mapeamento do desmatamento nesta célula.

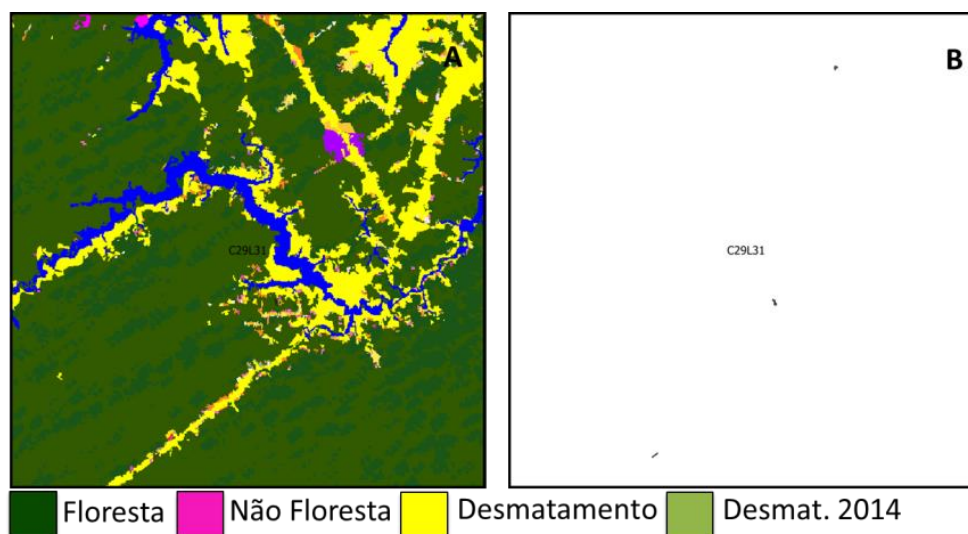
Figura 4.19 - Célula padrão Espinha de Peixe.



Fonte: Produção do autor.

Célula nº 8 (localização na Figura 4.12): ID C29L31 – padrão Multidirecional: Esta célula apresenta um baixo erro estimado de 0,008. Na Figura 4.20 (A) é possível verificar que a região apresenta desmatamento de formas variadas e irregulares no interior da célula calculado em 15% pela variável PLAND. A variável P_Prodes2014Flor indica baixa porcentagem de desmatamento sobre a área de floresta (0,0002). Foram detectados 3 polígonos de desmatamento de 2014 nesta célula – Figura 4.20 (B), cuja área média é de 8 ha. Esta configuração espacial indica baseado no modelo de regressão, baixa probabilidade de encontrar erro no mapeamento do desmatamento nesta célula.

Figura 4.1 - Célula padrão Multidirecional.



Fonte: Produção do autor.

4.5.4 Resumo das análises do modelo

Com base nos dados extraídos do modelo de regressão foi possível verificar que os padrões de desmatamento Consolidado e Multidirecional apresentaram as maiores estimativas de erros dentre os analisados. Considerando somente as 100 células com os maiores erros estimados, cujo valores variam de 3 a 16%, 54 pertencem ao padrão Consolidado e 32 ao Multidirecional. Dentre estas 54 células do padrão Consolidado a média da métrica PLAND ficou acima de 75% - indicando uma alta porcentagem de desmatamento acumulado na célula. Analisando as 417 células com erros estimados entre 1 a 16%, 28% pertencem ao padrão Consolidado e 32% ao Multidirecional. As métricas P_Prodes2014Flor e PLAND tem relação direta com o erro estimado, já a métricas TABO tem relação inversa.

5 CONSIDERAÇÕES FINAIS

O desenvolvimento deste estudo pretendeu estabelecer uma metodologia consistente e reprodutível que permitisse calcular estimativas confiáveis dos níveis de qualidade do mapeamento produzido pelo PRODES na Amazônia Legal para o ano de 2014, e duas perguntas principais nortearam as análises realizadas. A primeira relativa à identificação de locais onde há maior ou menor confiabilidade nos resultados do mapeamento PRODES e ainda quais fatores contribuem para o aumento do erro deste mapeamento. Análises iniciais demonstraram que o desmatamento na Amazônia Legal se tornou um evento pouco frequente na série histórica apresentando uma redução de 40% no número de polígonos mapeados entre 2002 (104.681 polígonos) e 2014 (41.729 polígonos).

Para responder às perguntas formuladas, partiu-se de uma regionalização da área de estudo representada por células de 50 x 50 km, em que padrões de desmatamento foram classificados a partir de uma tipologia estabelecida, com uso de métricas de ecologia da paisagem, técnicas de mineração de dados e classificação por árvore de decisão. Foi desenvolvido um esquema amostral para validação do mapeamento a partir de sorteio aleatório-estratificado de aproximadamente 1800 pontos que foram validados por três auditores independentes com uso de uma ferramenta construída especificamente para esta finalidade. Com base na validação dos pontos amostrais, das variáveis extraídas das métricas, além de outras variáveis obtidas no processamento dos dados, foi desenvolvido um modelo de regressão logística multivariado, ajustado para gerar uma superfície contínua que apontasse para regiões com maiores e menores probabilidade de ocorrência de erro no mapeamento do PRODES para a região da Amazônia Legal.

Três variáveis foram selecionadas pelo modelo, apresentando relação estatisticamente significativa, sendo elas: P_Prodes2014Flor (associada à porcentagem de desmatamento do PRODES 2014 em relação a área de floresta) PLAND (associada à porcentagem da área desmatada na célula) e TABO (associada à área do maior polígono de desmatamento no interior da

célula). Aplicado o modelo a todo conjunto do espaço celular utilizado foi possível espacializar a área de estudo em regiões com maior e menor probabilidade de se encontrar erro no mapeamento do PRODES. Esta espacialização mostrou que 78% das células apresentam erro estimado menor que 1% enquanto somente 0.3% apresentam erro estimado maior que 10%. Quando analisada a distribuição do espacial das 20 células concentradas na faixa superior a 5% de erro estimado, verifica-se que são próximas às regiões onde historicamente se concentram áreas com desmatamento já bem estabelecido na Amazônia Legal, principalmente no estado do Mato Grosso e associadas a dois padrões – Consolidado e Multidirecional. Estes padrões apresentam uma variável significativa do modelo que indica a área total desmatada na célula (PLAND), que para este conjunto de 20 células variou de 37% a 94% (média de 68%) indicando que em células com alta concentração de desmatamento acumulado, aumenta-se a probabilidade de ocorrência de erro no mapeamento. Com base nos dados extraídos do modelo de regressão foi possível verificar que os padrões de desmatamento Consolidado e Multidirecional apresentaram as maiores estimativas de erros dentre os analisados, chegando a representar 60% do total das células com erro estimado maior que 1%.

A metodologia utilizada neste estudo, apoiada no uso do mapa de padrões de desmatamento para relacioná-los como fonte de erros de mapeamento, associada à validação de pontos amostrais e a modelagem estatística da probabilidade de verificação de erro se mostrou adequada. Sua aplicação deveria ser acoplada ao esquema de produção sistemática nos mapeamentos futuros do PRODES, possibilitando assim que os dados produzidos fossem divulgados conjuntamente com seus índices de erros regionalizados, tanto para a Amazônia Legal como para os demais biomas brasileiros.

Desta forma seria possível associar a estes produtos uma métrica de qualidade, que enriqueceria sobremaneira todo conjunto de dados. Destaca-se ainda que esta metodologia não é restrita a forma como os dados do PRODES são produzidos atualmente, por técnica de interpretação visual de imagens de satélite. Ela também poderia ser aplicada para avaliação de mapeamentos de

desmatamentos obtidos por processos automatizados e/ou semi-automatizados de classificação de imagens.

Como sugestão de trabalhos futuros, indica-se adotar a metodologia desenvolvida neste trabalho para outros anos do mapeamento do PRODES a fim de conferir se o resultado aqui encontrado se mantém ou mesmo aplica-la aos dados do PRODES Cerrado, cujos dados foram recém-divulgados pelo INPE.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMI, M.; MELLO, M. P.; AGUIAR, D. A. ; RUDORFF, B. F. T. ; SOUZA, A. F. A web platform development to perform thematic accuracy assessment of sugarcane mapping in south-central Brazil. **Remote Sensing**, v. 4, p. 3201-3214, 2012.

ADAMI, M; GOMES, A. R. ; BELUZZO, A. ; COELHO, A. S. ; VALERIANO, D. ; RAMOS, F. S. ; NARVAES, I. S. ; BROWN, I. F. ; OLIVEIRA, I. D. ; SANTOS, L. B. ; MAURANO, LUIS EDUARDO PINHEIRO ; WATRIN, O. S. ; GRACA, P. M. L. A. A confiabilidade do PRODES: estimativa da acurácia do mapeamento do desmatamento no estado Mato Grosso. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18, 2017, Santos. **Anais...** São José dos Campos: INPE, 2017. p. 4189-4196.

ALMEIDA, C. A.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; ADAMI, M.; VENTURIERI, A.; DINIZ, C. G.; DESSAY, N.; DURIEUX, L.; GOMES, A. R. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. **Acta Amazônica**, Manaus, v. 46, p. 291-302, 2016.

BECKER, B. K. **Amazônia**. São Paulo: Ática, 1990. v. 1. 112p.

BECKER, B. K. **Amazônia**: geopolítica na virada do III milênio. São Paulo: Garamond, 2004. v. 1. 172p.

BECKER, B. K. Geopolítica da Amazônia. **Estudos Avançados**, São Paulo, v.19, n.53, p. 71-86, 2005.

BRASIL. **Lei nº 1.806, de 6 de Janeiro de 1953**. Brasília, 1963. Disponível em: <<http://www2.camara.leg.br/legin/fed/lei/1950-1959/lei-1806-6-janeiro-1953-367342-publicacaooriginal-1-pl.html>>. Acesso em: 20 mar. 2017.

BRASIL. **Decreto-lei nº 1.106, de 16 de junho de 1970**. Brasília, 1970. Disponível em: <http://www.planalto.gov.br/ccivil_03/Decreto-Lei/1965-1988/Del1106.htm>. Acesso em: 20 maio 2018.

BRASIL. MINISTÉRIO DO MEIO AMBIENTE- MMA. **Biodiversidade**. 2017. Disponível em: <http://www.mma.gov.br/estruturas/secex_consumo/_arquivos/5%20-%20mcs_biodiversidade.pdf>, 2017. Acesso em: 05 mar. 2018.

BRASIL. MINISTÉRIO DO MEIO AMBIENTE - MMA. **O nível de referência de emissões florestais do Brasil para pagamentos por resultados de redução de emissões provenientes do desmatamento no bioma Amazônia.** 2017. Disponível em: <<http://redd.mma.gov.br/images/central-de-midia/pdf/artigos/notainformativa-FREL-MMA-03-fev-15.pdf>>. Acesso 05 mar. 2018.

BRASIL. MINISTÉRIO DO MEIO AMBIENTE - MMA. **Plano de ação para prevenção e controle do desmatamento na Amazônia Legal.** 2017. Disponível em: <<http://www.mma.gov.br/florestas/controle-e-prevenção-do-desmatamento/plano-de-ação-para-amazônia-ppcdam>>. Acesso em 30 jan. 2017.

BUJLOW, T.; RIAZ, M. T.; PEDERSEN, J. M. A method for classification of network traffic based on C5.0 Machine Learning Algorithm. In: ICNC'12: INTERNATIONAL CONFERENCE ON COMPUTING, NETWORKING AND COMMUNICATIONS, 2012, Maui, Hawaii, USA. **Proceedings...** IEEE, 2012. p.237-241.

CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U.M.; GARRIDO, J. Spring: integrating remote sensing and gis by object-oriented data modelling. **Computers & Graphics**, v. 20, p. 395-403, 1996.

CÂMARA, G.; CARVALHO, M. S. Análise espacial de eventos. In DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (Eds). **Análise espacial de dados geográficos.** Brasília: Embrapa, 2004. p. 1-15.

CÂMARA, G.; VINHAS, L.; REIS FERREIRA, K.; RIBEIRO DE QUEIROZ, G.; CARTAXO MODESTO DE SOUZA, R.; VIEIRA MONTEIRO, M. TerraLib: an open source GIS library for large-scale environmental and socio-economic application. In: HALL, G.B.; Leahy, M. G. (Eds.). **Open source approaches in spatial data handling.** Berlin: Springer, 2008. p. 247–270.

CAMARGO, E. C. G.; DRUCK, S.; CÂMARA, G. Análise de superfícies por geoestatística linear. In DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (Eds.). **Análise espacial de dados geográficos.** Brasília: EMBRAPA, 2004. p.1-37.

CARD, D. H. Using known map category marginal frequencies to improve estimates of thematic map accuracy. **Photogrammetric Engineering and Remote Sensing**, v. 48, n.3, p. 431–439, 1982.

- CELESTIAL GREEN. **Independent validation of PRODES report 2014.** Disponível em: <http://www.celestialgreenventures.com/wp-content/uploads/2013/06/Report-on-the-Independent-Verification-of-PRODES-Forest-Change-Data_CGV_December2014.pdf>. Acesso em: 28 jan. 2018.
- CHAMBERS, J.Q.; HIGUCHI, N.; TRIBUZY, E.S.; TRUMBORE, S.E. Sink for a century: carbon sequestration in the Amazon. **Nature**, London, v. 410, n.6827, p. 429–429, 2001.
- CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 49, n. 12, p. 1671-1678, 1991.
- CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices.** New York: Lewis, 1999, 137p.
- CONNOR E. F. **Advanced biometry: logistic regression.** San Francisco: StateUniversity. Disponível em: <<http://userwww.sfsu.edu/efc/classes/biol710/logistic/logisticreg.pdf>>. Acesso em: 24 abr. 2018.
- CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise multivariada para os cursos de administração, ciências contábeis e economia.** São Paulo: Atlas, 2007. 568p.
- DINIZ, C. G.; SOUZA, A. A. A.; SANTOS, D. C.; DIAS, M. C.; LUZ, N. C.; MORAES, D. R. V.; MAIA, J. S.; GOMES, A. R.; NARVAES, I. S.; VALERIANO, D. M.; MAURANO, L. E. P.; ADAMI, M. DETER-B: the new Amazon near real-time deforestation detection system. **IEEE J-STARS**, v. 8, p. 1-10, 2015. DOI: 10.1109/JSTARS.2015.2437075
- DORAIS, A.; CARDILLE, J. Strategies for incorporating high-resolution Google Earth databases to guide and validate classifications: understanding deforestation in Borneo. **Remote Sensing**, v. 3, p. 1157–1176, 2011.
- DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (Eds.). **Análise espacial de dados geográficos.** Brasília: EMBRAPA, 2004.
- ESCADA, M. I. S. **Evolução de padrões de uso e cobertura da terra na região Centro-Norte de Rondônia. 2003.** 264 p. (INPE-10209-TDI/899). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2003.
- EWERS, R. M.; LAURANCE, W. F. Scale-dependent patterns of deforestation in the Brazilian Amazon. **Environmental Conservation**, v. 33, n.3, p. 203-211, 2006.

FELGUEIRAS, C. A.; DRUCK, S.; MONTEIRO, A. M. V. Análise espacial de superfícies por geoestatística: o enfoque por indicação. In: DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (Eds.). **Análise espacial de dados geográficos**. Brasília: EMBRAPA, 2004. p.1-25.

FERREIRA, E.; DANTAS, A. A. A.; MORAIS, A.R. Exatidão na classificação de fragmentos de matas em imagem do satélite Cbers-CCD, no município de Lavras. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007, Florianópolis. **Anais...** São José dos Campos: INPE, 2007. p. 887-894.

FINER, M.; NOVOA, S.; WEISSE, M. J.; PETERSEN, R.; MASCARO, J.; SOUTO, T.; STEARNS, F.; MARTINEZ, R. G. Combating deforestation: from satellite to intervention. **Science**, v. 360, p. 1303-1305, 2018. DOI:10.1126/science.aat1203.

FONTE, C. C.; GONÇALVES, L. M. S. **Spatialization of classification accuracy using spatially constrained confusion matrixes and classification uncertainty**. Coimbra: Instituto de Engenharia e Computadores de Coimbra, 2015.

FOOD AND AGRICULTURE ORGANIZATION - FAO. **Map accuracy assessment and area estimation: a practical guide**. 2016. Disponível em: <http://www.fao.org/3/a-i5601e.pdf>. Acesso em: 21 mar. 2018.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, p. 185–20, 2001.

FOODY, G. M. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. **International Journal of Remote Sensing**, v. 26, n.6, p. 1217-1228, 2005.

FUNDO AMAZÔNIA. **Projetos apoiados**. 2017. Disponível em: <http://www.fundoamazonia.gov.br/FundoAmazonia/fam/site_pt/Esquerdo/Projetos_Apoiados/Lista_Projetos/INPE>. Acesso em: 15 jan. 2017.

GAVLAK, A. A.; ESCADA, M. I. S.; MONTEIRO, A. M. V. Dinâmica de padrões de mudança de uso e cobertura da terra na região do Distrito Florestal Sustentável da BR-163. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. **Anais...** São José dos Campos: INPE, 2011. p. 6152–6160.

GEIST, H. J.; ERIC F. L. Proximate causes and underlying driving forces of tropical deforestation. **BioScience**, v. 52, n.2, p.143-150, 2002.

HOSMER JUNIOR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. New York: John Wiley & Sons, 2013.

HUDSON, W. D. Correct formulation of the kappa coefficient of agreement. **Photogrammetric Engineering and Remote Sensing**, v. 53, n. 4, p. 421-422, 1987.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Mapa integrado dos zoneamentos ecológico-econômicos dos estados da Amazônia Legal**. 2017. Disponível em: <http://www.ibge.gov.br/home/geociencias/geografia/mapas_doc5.shtm> Acesso em: 24 fev. 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Produção agrícola municipal: culturas temporárias e permanentes**. 2016. Disponível em :< <http://www.sidra.ibge.gov.br>> Acesso em: 22 fev. 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Fronteira agrícola: Amazônia Legal**. 2015. Disponível em: < <https://www.ibge.gov.br/geociencias-novoportal/cartas-e-mapas/mapas-regionais/15976-fronteira-agricola-amazonia-legal.html>>. Acesso em: 16 fev. 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Censo agropecuário 2006**. 2006. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 16 fev. 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Mapa de biomas e de vegetação**. 2004. Disponível em: <<https://ww2.ibge.gov.br/home/presidencia/noticias/21052004biomashtml.shtm>> Acesso em: 20 mar. 2018.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Monitoramento da cobertura florestal da Amazônia por satélites**. 2008. Disponível em: <<http://www.obt.inpe.br/deter/avaliacao/RelatorioMonitoramento.pdf>>. Acesso em: 27 abr. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **TerraView**: 2010. Disponível em: < <http://www.dpi.inpe.br/terraview>>. Acesso em: 09 jan. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Monitoramento da cana-de-açúcar via imagens de satélite, Projeto Cansat**. 2016: Disponível em: <http://www.dsr.inpe.br/laf/canasat/>. Acesso em: 09 jan. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Projeto PRODES**: Projeto de Estimativa de Desflorestamento da Amazônia: taxas anuais do desmatamento – 1988 até 2016. 2017. Disponível em http://www.obt.inpe.br/prodes/prodes_1988_2016n.htm. Acesso em: 21 jan. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **GeoDMAFeatures**. 2017. Disponível em: <http://wiki.dpi.inpe.br/doku.php?id=geodma:features>. Acesso em: 15 fev. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Sistema de Detecção de Desmatamento em Tempo Real (DETER)**. 2017. Disponível em: < <http://www.obt.inpe.br/deter>>. Acesso em: 12 mar. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE. **Spring**: manual do usuário. Disponível em: <<http://www.dpi.inpe.br/spring/portugues/tutorial/modelagem.html>>. Acesso em: 19 jan. 2018.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **TerraLib and TerraView**. 2017. Disponível em: <http://www.dpi.inpe.br/terraview>. Acesso em: 19 abr. 2018.

JENSEN, J. R. **Sensoriamento remoto do ambiente**: uma perspectiva em recursos terrestres. São José dos Campos, SP: Parênteses, 2009. 598 p.

KINTISCH E. Improved monitoring of rainforests helps pierce haze of deforestation. **Science**, v.316, p.536-537, 2007.

KORTING, T.S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA: Geographic Data Mining Analyst a framework for GIScience. **Computers & Geosciences**, v. 57, p. 133-145, 2013.

KORTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S.; SILVA, F. C.; SILVA, M. P. S. GeoDMA: a novel system for spatial data mining. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, 2008, Pisa, Italy. **Proceedings...** Pisa: IEEE, 2008. v. 10. p. 975-978.

LANDIS, J.R.; KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p. 159-174, 1977.

LE QUERE C.; RAUPACH M.R.; CANADELL J.G. Trends in the sources and sinks of carbon dioxide. **Nature Geosciences**, v. 2, p. 831–836, 2009.

- LEE, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. **International Journal of Remote Sensing**, v. 26:7, p. 1477-1491, 2005.
- LOARIE S. R.; ASNER G.P.; FIELD, C. B. Boosted carbon emissions from Amazon deforestation. **Geophysical Research Letters**, v.36, n.14, p.1–5, 2009.
- MAGALHÃES, J. P. **A ocupação desordenada da Amazônia**. Brasília: Completa, 1990. 111 p.
- MAURANO, L. E.; ADAMI, M. Ferramentas Web-Gis para avaliar exatidão de mapeamento de uso e cobertura da terra no Cerrado brasileiro. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017, Santos. **Anais...** São José dos Campos: INPE, 2017. p. 462-469.
- MCGARIGAL, K.; MARKS, B. J. **FRAGSTATS**: spatial pattern analysis program for quantifying landscape structure. Washington: United State Department of Agriculture, 1995. 132 p.
- MCROBERTS, R.E. Satellite image-based maps: scientific inference or pretty pictures? **Remote Sensing of Environment**, v. 115, p. 715-724, 2011.
- NEPSTAD, D.; MCGRATH, D.; STICKLER, C.; ALENCAR, A.; AZEVEDO, A.; SWETTE, B.; BEZERRA, T.; DIGIANO, M.; SHIMADA, J.; SEROA DA MOTTA, R.; ARMIJO, E.; CASTELLO, L.; BRANDO, P.; HANSEN, M.C.; MCGRATH-HORN M.; CARVALHO, O.; HESS, L. Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. **Science**, v. 344, p. 1118-1123, 2014.
- NETER J.; KUTNER, M.H.; NACHTSHEIM, C.J.; WASSERMAN, W. **Applied linear statistical models**. 4. ed. New York: McGraw-Hill, 1996. 318p.
- OBORN, I.; MAGNUSSON, U.; BENGTSSON, J.; VREDE, K.; FAHLBECK, E.; JENSEN, E.S.; WESTIN, C.; JANSSON, T.; HEDENUS, F.; LINDHOLM, S. H.; STENSTRÖM, M.; JANSSON, B.; RYDHMER, L. **Five scenarios for 2050**: conditions for agriculture and land use. [S.l.]: Swedish University of Agricultural Sciences, 2011.
- PANDYA, R.; PANDYA J. C5.0 Algorithm to improved decision tree with feature selection and reduced error pruning. **International Journal of Computer Applications**, v. 117, n. 16, p. 18-21, 2015.
- PONTIUS, R. G. J. R; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n.15, p. 4407-4429, 2011.

PONTIUS, R. G.; SANTACRUZ, A. Quantity, exchange, and shift components of difference in a square contingency table. **International Journal of Remote Sensing**, v. 35, p. 7543-7554, 2014.

QUEIROZ, G. R.; FERREIRA, K. R.; VINHAS, L.; CÂMARA, G.; COSTA, R. W.; SOUZA, R. C. C. M.; MAUS, V. W.; SANCHEZ, A. WTSS: um serviço web para extração de séries temporais de imagens de sensoriamento remoto. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17., 2015, João Pessoa, PB. **Anais...** São José dos Campos: INPE, 2015. p. 7553-7560.

QUINLAN, J. R. **C4.5**: programs for machine learning. San Mateo: Morgan Kaufmann, 1993. 302p.

QUINLAN, J.R. **C5.0**: an informal tutorial. 2017. Disponível em: <<https://www.rulequest.com/see5-unix.html>>. Acesso em: 26 fev. 2018.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2017. Disponível em: <https://www.R-project.org/>.

RIBEIRO, F. G. **Uso da regressão logística na estimação da probabilidade de reincidência de jovens infratoras**. 2015. Monografia (Bacharelado em Estatística) - Universidade de Brasília, Brasília, 2015. Disponível em: <http://bdm.unb.br/handle/10483/13133>.

RUDOLPHO, L. S.; BRAGHIROLI, G.; REFOSCO, J. C.; SANTIAGO, A. G.; SABOYA, R. T. Aplicação de técnicas de geoprocessamento e métricas da paisagem na análise temporal da cobertura florestal da Bacia do Ribeirão Fortaleza em Blumenau/SC. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16., 2013, Foz do Iguaçu, PR. **Anais....** São José dos Campos: INPE, 2013. p. 1742-1749.

RUDORFF, B.F.T.; ADAMI, M.; AGUIAR, D.A.; MOREIRA M.A.; MELLO M.P.; FABIANI, L.; AMARAL, D.F.; PIRES, B.M. The soy moratorium in the Amazon biome monitored by remote sensing images. **Remote Sensing**, v.3, p.185-202, 2011.

SAITO, É. A. **Caracterização de trajetórias de padrões de ocupação humana na Amazônia Legal por meio de mineração de dados**. 2010. 158 p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010.

SAITO, E. A.; FONSECA, L. M. G.; ESCADA, M. I. S.; KORTING, T. S. Efeitos da mudança de escala em padrões de desmatamento na Amazônia. **Revista Brasileira de Cartografia**, v.63, n.3, p.401-414, 2011.

SATURNINO, H. M. S. **Modelação e mapeamento da probabilidade de incêndio florestal**. 2012. 107p. Dissertação (Mestrado em Sistemas de Informação Geográfica) – Instituto Politécnico de Castelo Branco, Castelo Branco, Portugal, 2012.

SILVA, R. G. C. Amazônia globalizada: da fronteira agrícola ao território do agronegócio: o exemplo de Rondônia. **Revue Franco-Brésilienne de Géographie**, n. 23, 2015.

SILVA, M. P. S.; CÂMARA, G.; SOUZA, R. C. M.; VALERIANO, D. M.; ESCADA, M. I. S. Mining patterns of change in remote sensing image databases. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, 5., 2005. **Proceedings...** 2005.

SILVA, M. P. S.; CÂMARA, G.; ESCADA, M. I. S.; SOUZA, R. C. M. Remote-sensing image mining: detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, v.29, n.16, p. 4803-4822, 2008.

SHIMABUKURO, Y. E.; BATISTA, G. T.; MELLO, E. M. K.; MOREIRA, J.C.; DUARTE, V. Using shade fraction image segmentation to evaluate deforestation in Landsat Thematic Mapper images of the Amazon Region. **International Journal of Remote Sensing**, v. 19, p. 535 – 541, 1998.

SMEDES, H. W. The truth about ground truth. In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 10., 1975. **Proceedings...** Ann Arbor, Michigan: Environmental Research Institute of Michigan, 1975, p. 821-823.

SPRACKLEN, D.V.; ARNOLD, S. R.; TAYLOR, C. M. Observations of increased tropical rainfall preceded by air passage over forests. **Nature**, v. 489, p. 282–285, 2012.

SUAREZ, A. F.; CANDEIAS, A. L. B. Avaliação de exatidão da classificação de dados de sensoriamento remoto para o município de Maragogipe. In: SIMPÓSIO BRASILEIRO DE CIÊNCIAS GEODÉSICAS E TECNOLOGIAS DA GEOINFORMAÇÃO, 4., 2012, Recife. **Anais....** Recife: Universidade Federal de Pernambuco, 2012. v. 1. p. 1-10.

TARDIN, A. T.; SANTOS, A. P.; NOVO, E. M. L. M. **Projeto SUDAM**: uso da dados do Landsat no estudo do impacto da implantação de projetos agropecuários da Amazônia. São Jose dos Campos: INPE, 1978. 18 p. Disponível em: <<http://urlib.net/6qtX3pFwXQZ3r59YD6/GLCUV>>.

TARDIN, A. T.; RODRIGUES, J. E.; ABDON, M. M.; NOVAES, R. A.; CHEN, S. C.; DUARTE, V.; SHIMABUKURO, Y. E.; SANTOS, A. P.; LEE, D. C. L.; MAIA, F. C. S.; MENDONCA, F. J.; ASSUNCAO, G. V. **Levantamento de áreas de desmatamento na Amazônia Legal através de imagens do Satélite Landsat**. São Jose dos Campos: INPE, 1979. 62 p. Disponível em: <<http://urlib.net/6qtX3pFwXQZ3r59YD6/GLCRo>>.

TERRAAMAZON. **TerraAmazon**. 2017: Disponível em: <http://terraamazon.org/index.php/pt/sobre>. Acesso em: 09 mar. 2017

TSUCHIYA, Í. **Regressão logística aplicada na análise espacial de dados arqueológicos**. 2002. 86 p. Dissertação (Mestrado em Ciências Cartográficas) - Universidade Estadual Paulista, Presidente Prudente, 2002.

VALENTIM, J. F.; ANDRADE, C. M. S. Tendências e perspectivas da pecuária bovina na Amazônia brasileira. **Amazônia**, v. 4, p. 9-32, 2009.

VALERIANO, D. M.; ESCADA, M. I. S.; CÂMARA, G.; AMARAL, S.; MAURANO, L. E.; RENNÓ, C. D.; ALMEIDA, C. A.; MONTEIRO, A. M. V. O monitoramento do desmatamento. In: MARTINE, G. (Org.). **População e sustentabilidade na era das mudanças ambientais globais**. Belo Horizonte: ABEP, 2012. p. 223-238.

VIEIRA, I. C. G.; ARAÚJO, R.; TOLEDO, P. M. Dinâmicas produtivas, transformações no uso da terra e sustentabilidade na Amazônia. In: SILFET, N.; CARDOSO, M.; MAGALHÃES, W.; LASTRES, H. (Orgs.). **Um olhar territorial para o desenvolvimento da Amazônia**. Rio de Janeiro: BNDES, 2014. p. 370-395. DOI: 10.13140/2.1.5180.7682.

WERTH D, A. R. The local and global effects of Amazon deforestation. **Journal of Geophysical Research**, v. 107, n.D20, p. 551–558, 2002.

ANEXO A – CÓDIGO FONTE DO PROGRAMA R C5.0

O programa abaixo, escrito em linguagem R, foi utilizado para realizar a classificação para obtenção do mapa de padrões de desmatamento a partir das métricas de paisagem extraídas com uso do GeoDMA. Este programa utiliza o algoritmo de classificação supervisionado baseado no conceito de árvore de decisão versão C 5.0, conforme descrito na Seção 3.3.5.

```
#-----
library(C50)
library(DiagrammeR)
source("src/C502GraphViz.R")

# ----- Inputs -----
pathOutput <- "./output/"
inputCSV <- "./data/MetricasGEODMA_Prodes2014.csv"
outputCSVName <- "cl"
outputTreeFileName <- "tree_Full"
outputSummaryFileName <- "summaryFull"
#trainingField <- "trainClass"

#-----
# Read the CSV
myData <- read.csv(inputCSV, sep = ";", na = "")
# Remove FID from training data
trainingData <- subset(myData, select=-c(id, col, row, SHDI, SHEI, PR, PRD, SIDI, SIEI, X))
# Read the interesting classes
classes <- levels(trainingData$trainClass)

# ----- First classification -----
input <- subset(trainingData,select=-c(trainClass))
output <- trainingData[, "trainClass"]

for(i in 1:100)
{
  result <- C5.0(input, output, control = C5.0Control(noGlobalPruning = TRUE, minCases=4,
  sample=0.75))
  outputTreeFile <- paste(pathOutput, outputTreeFileName, "-", i, ".dot", sep="")

  # Plot the tree using GraphViz
  C5.0.graphviz(result, outputTreeFile)
  grViz(outputTreeFile)
  # Output Summary TXT file
  summary(result)
  outputSummaryFile <- paste(pathOutput, outputSummaryFileName, "-", i, ".txt", sep="")
  write(toString(summary(result)), file = outputSummaryFile)

  # Output CSV classification
  myData["class"] <- predict(result, trainingData)
  outputData <- subset(myData, select=c(id, class))
  outputCSV <- paste(pathOutput, outputCSVName, "_", i, ".csv", sep="")
  write.table(outputData, file = outputCSV, append = FALSE, sep = ";", row.names = FALSE)
}
}
```