

UNIVERSIDADE FEDERAL DE GOIÁS  
INSTITUTO DE INFORMÁTICA

EVANDO NATAL FERNANDES DE OLIVEIRA

**Uma abordagem semissupervisionada  
para classificação de pastagens usando  
séries temporais de NDVI**

Goiânia - Goiás  
2020

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR  
VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES  
NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das dissertações e teses disponibilizados são de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o autor e o orientador firmam o compromisso de que ele não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

**1. Identificação do material bibliográfico:**    ☒ **Dissertação**        ☐ **Tese**

**2. Identificação da Tese ou Dissertação:**

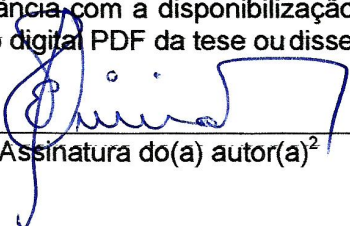
Nome completo do autor: Evando Natal Fernandes de Oliveira

Título do trabalho: Uma abordagem semissupervisionada para classificação de pastagens usando séries temporais de NDVI.

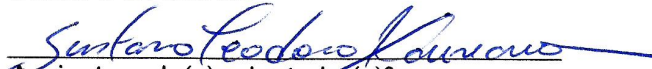
**3. Informações de acesso ao documento:**

Concorda com a liberação total do documento ☒ **SIM**                      ☐ **NÃO**<sup>1</sup>

Independente da concordância com a disponibilização eletrônica, é imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

  
Assinatura do(a) autor(a)<sup>2</sup>

Ciente e de acordo:

  
Assinatura do(a) orientador(a)<sup>2</sup>

Data: 28/01/2020

<sup>1</sup> Neste caso o documento será embargado por até um ano a partir da data de defesa. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

<sup>2</sup> As assinaturas devem ser originais sendo assinadas no próprio documento, imagens coladas não serão aceitas.

EVANDO NATAL FERNANDES DE OLIVEIRA

# **Uma abordagem semissupervisionada para classificação de pastagens usando séries temporais de NDVI**

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

**Área de concentração:** Ciência da Computação.

**Orientador:** Prof. Dr. Gustavo Teodoro Laureano

Goiânia - Goiás  
2020

Ficha de identificação da obra elaborada pelo autor, através do  
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

de Oliveira, Evando Natal Fernandes

Uma abordagem semissupervisionada para classificação de  
pastagens usando séries temporais de NDVI [manuscrito] / Evando  
Natal Fernandes de Oliveira. - 2020.

LXXXVII, 87 f.: il.

Orientador: Prof. Dr. Gustavo Teodoro Laureano.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto  
de Informática (INF), Programa de Pós-Graduação em Ciência da  
Computação, Goiânia, 2020.

Bibliografia.

Inclui siglas, mapas, fotografias, abreviaturas, gráfico, tabelas,  
algoritmos, lista de figuras, lista de tabelas.

1. Séries Temporais. 2. NDVI. 3. MODIS. 4. Modelo Linear de  
Mistura Temporal. 5. Classificação Semissupervisionada. I. Laureano,  
Gustavo Teodoro, orient. II. Título.

CDU 004





UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

**ATA DE DEFESA DE DISSERTAÇÃO**

Ata nº **07/2020** da sessão de Defesa de Dissertação de **Evando Natal Fernandes de Oliveira**, que confere o título de Mestre em Ciência da Computação, na área de concentração em Ciência da Computação.

Aos vinte e oito dias do mês de janeiro de dois mil e vinte, a partir das dezesseis horas, na sala 150 do Instituto de Informática, realizou-se a sessão pública de Defesa de Dissertação intitulada **“Uma abordagem semissupervisionada para classificação de pastagens usando séries temporais de NDVI”**. Os trabalhos foram instalados pelo Orientador, Professor Doutor Gustavo Teodoro Laureano (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professor Doutor Leandro Leal Parente (LAPIG/UFG), membro titular externo; e Professor Doutor Anderson da Silva Soares (INF/UFG), membro titular interno. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor Gustavo Teodoro Laureano, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos vinte e oito dias do mês de janeiro de dois mil e vinte.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **LEANDRO LEAL PARENTE, Usuário Externo**, em 28/01/2020, às 18:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Gustavo Teodoro Laureano, Professor do Magistério Superior**, em 28/01/2020, às 18:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Anderson Da Silva Soares**,  
**Professor do Magistério Superior**, em 28/01/2020, às 18:17,  
conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º,  
do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



A autenticidade deste documento pode ser conferida no site  
[https://sei.ufg.br](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)  
[/sei/controlador\\_externo.php?acao=documento\\_conferir&](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)  
[id\\_orgao\\_acesso\\_externo=0](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1107936**  
e o código CRC **984A3F51**.

---

**Referência:** Processo nº 23070.048317/2019-17

SEI nº 1107936

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Evando Natal Fernandes de Oliveira**

Tem graduação em Licenciatura em Matemática pela Pontifícia Universidade Católica de Goiás – (PUC, 1981), pós-graduação *lato sensu* em Geoprocessamento pelo Centro Federal de Educação Tecnológica de Goiás (CEFET-GO, 2005), além de uma especialização em Gerenciamento de Projetos de TI pela George Washington University (2003). Com mais de quarenta anos de experiência na área de tecnologia da informação, coordenou, por mais de quinze anos, grandes projetos de geoprocessamento aplicado às telecomunicações nas empresas Telegoiás, Brasil Telecom e Oi, tendo adquirido grande experiência na elaboração de Termos de Referência para aquisição e manutenção de Sistemas de Informações Geoespaciais (GIS) e também para aquisição e auditoria de dados geoespaciais. Durante quatro anos prestou consultoria especializada em geoprocessamento para a Secretaria de Planejamento da Prefeitura de Goiânia, gerenciando processos de aquisição e auditoria de dados geoespaciais. Tem um bom conhecimento sobre mapeamento aerofotogramétrico, geodésia por satélite, sensoriamento remoto, modelo de elevação digital e processamento digital de imagens. Possui vasta experiência na utilização dos softwares de código aberto PostGIS, GeoServer, OpenLayers, Saga e QGIS, todos patrocinados pela Open Source Geospatial Foundation (OSGeo). Atualmente é mestrando em Ciência da Computação na Universidade Federal de Goiás – UFG, sendo bolsista da CAPES.

Esta Dissertação é dedicada à minha filha Valentina, minha maior incentivadora e  
companheira diária, independente das circunstâncias!

---

## Agradecimentos

---

Meus agradecimentos:

À UFG, por ter me dado a oportunidade de ter tido um ensino de alta qualidade e gratuito.

À CAPES, por ter me concedido um auxílio financeiro por meio de uma bolsa de estudos.

Ao Prof. Dr. Gustavo Teodoro Laureano, por sua disposição em ser meu orientador, mesmo que eu não me enquadrasse no perfil médio dos candidatos ao mestrado em Ciência da Computação.

Aos meus familiares, por terem sido compreensivos quanto às minhas ausências nos eventos que ocorreram neste período.

À minha companheira Ana Paula, pelo incentivo, exemplo e solidariedade.

Não há passageiros na nave espacial Terra. Somos todos tripulação.

**Herbert Marshall McLuhan,**

.

---

## Resumo

---

de Oliveira, Evando N. F.. **Uma abordagem semissupervisionada para classificação de pastagens usando séries temporais de NDVI**. Goiânia - Goiás, 2020. 86p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

O conhecimento e o gerenciamento da cobertura e do uso do solo são fundamentais num cenário de aumento na produção de alimentos, devido ao crescimento da população mundial e à mudança de seus hábitos alimentares, com destaque para o incremento no consumo de proteína animal, o que demanda um aumento no rebanho bovino e, conseqüentemente, nas áreas de pastagens. Um aliado dos gestores e da comunidade tem sido, já há um bom tempo, o sensoriamento remoto, com sua produção de dados abundantes sobre a superfície terrestre em variadas resoluções espaciais, espectrais e temporais, com particular destaque para os índices de vegetação. Um destes índices, o NDVI, é calculado e disponibilizado com dados gerados pelo sensor MODIS em forma de uma série temporal, que é uma das fontes de informação mais usadas para a classificação das mais variadas feições vegetais. Neste trabalho, apresentamos uma metodologia de classificação de pastagens que compreende, numa primeira etapa, o uso do Modelo Linear de Mistura Temporal – MLMT, sendo que os membros finais foram obtidos a partir de uma classificação não supervisionada. Num segundo momento, é feita a rotulação dos dados a partir de um mapa de pastagens que tem melhor resolução espacial que os dados gerados pelo MODIS e que foi produzido pelo Laboratório de Processamento de Imagens e Geoprocessamento da Universidade Federal de Goiás (LAPIG – UFG). Em seguida, é construído o modelo de classificação que é aplicado aos dados a serem classificados e sua qualidade é aferida por meio da comparação com outro mapa de pastagens, também produzido pelo LAPIG, com resolução espacial semelhante aos dados classificados. A metodologia aqui utilizada apresentou resultados com qualidade compatível com outros trabalhos que tiveram abordagens de treinamento puramente supervisionado para a classificação de pastagens, usando a mesma base de dados.

### Palavras-chave

séries temporais; NDVI; pastagens; clusterização; classificação supervisionada; MLMT; MLME; MODIS

---

## Abstract

---

de Oliveira, Evando N. F.. **A semi-supervised approach for pastures classification using NDVI time series**. Goiânia - Goiás, 2020. 86p. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

The knowledge and the management of land cover and land use are fundamental in a scenario of food production increasing, due to growth of the world population and change in its eating habits, especially the increase in animal protein consumption, what demands an increase in the herd of cattle and, consequently, in the pasture areas. Remote sensing has been an ally of public managers and the community for a long time, with its abundant data production about the terrestrial surface in different spatial, spectral and temporal resolutions, with particular emphasis on vegetation indices. One of these indices, the NDVI, is calculated and made available from the data generated by the MODIS satellite sensor as a time series, which is one of the most used sources of information for the classification of the most varied types of vegetation. In this work, we present a methodology for classifying pastures that comprises, in a first step, the use of the Linear Temporal Mixture Model – LTMM, with the final members being obtained from an unsupervised classification method. Secondly, the data are labeled from a pasture map produced by the Processing of Images and Geoprocessing Laboratory of the Federal University of Goiás (LAPIG - UFG), which has better spatial resolution than the data generated by MODIS. Then, a classification model is constructed to be applied to the classifying data and its quality is measured by comparison with another pasture map, also produced by LAPIG, with the same spatial resolution than the classified data. The methodology used here presented results with quality compatible with other studies that had purely supervised training approaches for the classification of pastures, using the same data base.

### Keywords

time series; NDVI; pasture; clustering; semisupervised classification; LTMM; LSMM; MODIS



---

# Sumário

---

|                                                                                          |    |
|------------------------------------------------------------------------------------------|----|
| Lista de Figuras                                                                         | 14 |
| Lista de Tabelas                                                                         | 16 |
| 1 Introdução                                                                             | 18 |
| 1.1 Objetivos e contribuições                                                            | 20 |
| 1.1.1 Contribuições                                                                      | 21 |
| 1.2 Organização da dissertação                                                           | 21 |
| 2 Materiais                                                                              | 23 |
| 2.1 Regiões de interesse                                                                 | 23 |
| 2.2 Séries temporais para classificação de pastagens                                     | 25 |
| 2.2.1 O índice de vegetação NDVI                                                         | 25 |
| 2.2.2 Séries temporais de NDVI                                                           | 25 |
| 2.3 Mapas de pastagens                                                                   | 27 |
| 2.3.1 Mapa de pastagens – origem Landsat 8                                               | 28 |
| 2.3.2 Mapa de pastagens – origem MOD13Q1                                                 | 29 |
| 2.4 Softwares e linguagens                                                               | 32 |
| 3 Metodologia                                                                            | 33 |
| 3.1 Classificação não supervisionada                                                     | 34 |
| 3.1.1 Modelo Linear de Mistura Espectral - MLME                                          | 34 |
| 3.1.2 Modelo Linear de Mistura Temporal – MLMT                                           | 35 |
| 3.1.3 Obtenção dos Membros Finais                                                        | 35 |
| 3.1.3.1 Tratamento de ruídos, <i>outliers</i> e dados faltantes                          | 36 |
| 3.1.3.2 Membros finais obtidos com o algoritmo PPI                                       | 43 |
| 3.1.3.3 Membros finais obtidos por clusterização                                         | 45 |
| 3.1.4 Cálculo das Frações de Abundância                                                  | 52 |
| 3.2 Modelo de classificação supervisionada                                               | 54 |
| 3.2.1 Definição de parâmetros para rotulação das pastagens                               | 54 |
| 3.2.1.1 Percentuais de pastagens dos <i>pixels</i> NDVI                                  | 55 |
| 3.2.1.2 Correlação entre percentuais de pastagens e frações de abundância                | 55 |
| 3.2.1.3 Definição do coeficiente de multiplicação para seleção das frações de abundância | 56 |
| 3.2.2 Preparação do mapa verdade de campo                                                | 59 |
| 3.2.3 Matriz de Confusão                                                                 | 60 |
| 3.3 Execução da classificação supervisionada                                             | 62 |

|          |                                                                              |           |
|----------|------------------------------------------------------------------------------|-----------|
| <b>4</b> | <b>Resultados e Discussões</b>                                               | <b>63</b> |
| 4.1      | Primeira etapa                                                               | 63        |
| 4.1.1    | Obtenção dos membros finais                                                  | 63        |
| 4.1.2    | Frações de abundância                                                        | 65        |
| 4.2      | Segunda etapa                                                                | 66        |
| 4.2.1    | Cálculo do percentual de pastagens                                           | 67        |
| 4.2.2    | Correlação entre o percentual de pastagens e as frações de abundância        | 67        |
| 4.2.3    | Coeficiente de multiplicação para seleção das frações de abundância          | 68        |
| 4.2.4    | Validação em relação ao mapa LAPIG250                                        | 70        |
| 4.2.4.1  | Geração de grade compatível com o LAPIG250                                   | 70        |
| 4.2.4.2  | Comparação das classificações dos recortes de referência com o mapa LAPIG250 | 71        |
| 4.3      | Terceira etapa                                                               | 72        |
| 4.3.1    | Classificação dos recortes de classificação                                  | 72        |
| 4.3.1.1  | Recorte CDG                                                                  | 72        |
| 4.3.1.2  | Recorte CLV                                                                  | 73        |
| 4.3.1.3  | Recorte LUZ                                                                  | 74        |
| 4.3.2    | Definição de um novo recorte - REC39K                                        | 75        |
| 4.3.3    | Comparação com os resultados do mapa LAPIG250                                | 76        |
| <b>5</b> | <b>Conclusões</b>                                                            | <b>78</b> |
| 5.1      | Sobre a metodologia aplicada                                                 | 78        |
| 5.2      | Comparação com outros trabalhos                                              | 80        |
| 5.3      | Trabalhos futuros                                                            | 80        |
|          | <b>Referências Bibliográficas</b>                                            | <b>82</b> |

---

## Lista de Figuras

---

|      |                                                                                                       |    |
|------|-------------------------------------------------------------------------------------------------------|----|
| 1.1  | Resumo da metodologia adotada neste trabalho                                                          | 20 |
| 2.1  | Regiões de Interesse – Recortes de referência (em vermelho) e de classificação (em verde)             | 24 |
| 2.2  | Recorte de referência – SLMB                                                                          | 24 |
| 2.3  | Recorte de referência – NCX                                                                           | 25 |
| 2.4  | Cálculo do NDVI a partir da radiação solar                                                            | 26 |
| 2.5  | Perfil NDVI de um pixel durante o ano de 2015                                                         | 26 |
| 2.6  | Mapa Temático de NDVI do Estado de Goiás - Referência 01/01/2015                                      | 28 |
| 2.7  | Mapa de pastagens de Goiás – LAPIG030<br>Ano 2015                                                     | 30 |
| 2.8  | Mapa de pastagens de Goiás – LAPIG250<br>Ano 2015                                                     | 32 |
| 3.1  | Primeira etapa da metodologia                                                                         | 34 |
| 3.2  | Recorte de Goiânia e Gráfico de dispersão                                                             | 35 |
|      | (a) Recorte Goiânia (Bandas Vermelho + Infravermelho)                                                 | 35 |
|      | (b) Gráfico de dispersão dos <i>pixels</i> da imagem                                                  | 35 |
| 3.3  | Perfis NDVI originais dos <i>pixels</i> selecionados<br>Recorte SLMB – 2015                           | 39 |
| 3.4  | Perfis NDVI com ruídos dos <i>pixels</i> selecionados<br>Recorte SLMB – 2015                          | 40 |
| 3.5  | Distâncias Euclidianas entre as séries originais e as séries ruidosas<br>filtradas com Savitzky-Golay | 41 |
| 3.6  | Séries ruidosas após serem filtradas com Savitzky-Golay                                               | 41 |
| 3.7  | Distâncias Euclidianas entre as séries originais e as séries ruidosas<br>filtradas com MaxMin         | 42 |
| 3.8  | Séries ruidosas após serem filtradas com MaxMin                                                       | 42 |
| 3.9  | [Recorte SLMB - Localização dos membros finais gerados pelo PPI                                       | 44 |
| 3.10 | [Recorte SLMB - perfil NDVI dos membros finais gerados pelo PPI                                       | 45 |
| 3.11 | Comparação entre Distâncias Euclidiana e DTW                                                          | 47 |
|      | (a) Recorte SLMB - Distância Euclideana                                                               | 47 |
|      | (b) Recorte NCX - Distância DTW                                                                       | 47 |
| 3.12 | Recorte SLMB - Localização dos membros finais gerados pelo PAM                                        | 51 |
| 3.13 | Recorte SLMB - perfil NDVI dos membros finais gerados pelo PAM                                        | 51 |
| 3.14 | Exemplo das frações de abundância de um conjunto de <i>pixels</i> - Recorte<br>SLMB                   | 52 |
| 3.15 | Segunda etapa da metodologia                                                                          | 54 |
| 3.16 | Área abrangida pelos <i>pixels</i> do mapa de pastagens e da série NDVI                               | 55 |

|      |                                                                                                        |    |
|------|--------------------------------------------------------------------------------------------------------|----|
| 3.17 | Quartis das frações de abundância                                                                      | 57 |
| (a)  | Quartis das frações de abundância - recorte NCX                                                        | 57 |
| (b)  | Quartis das frações de abundância - recorte SLMB                                                       | 57 |
| 3.18 | Gráfico da seleção do coeficiente <i>cm</i> - recorte NCX                                              | 59 |
| 3.19 | Percentual de pastagens nos <i>pixels</i> NDVI                                                         | 60 |
| (a)  | NDVI > 0% pastagem                                                                                     | 60 |
| (b)  | NDVI $\geq$ 20% de pastagens                                                                           | 60 |
| (c)  | NDVI $\geq$ 40% de pastagens                                                                           | 60 |
| (d)  | NDVI $\geq$ 50% de pastagens                                                                           | 60 |
| 3.20 | Terceira etapa da metodologia                                                                          | 62 |
| 4.1  | Membros finais dos recortes de referência                                                              | 64 |
| (a)  | Membros finais - recorte SLMB                                                                          | 64 |
| (b)  | Membros finais - recorte NCX                                                                           | 64 |
| 4.2  | Frações de abundância - recorte NCX<br>x11(azul) - x22(laranja) - x3(verde) - x4(cinza) - x5(vermelho) | 65 |
| 4.3  | Frações de abundância - recorte SLMB<br>x1(azul) - x2(laranja) - x3(verde) - x4(cinza) - x5(vermelho)  | 66 |
| 4.4  | Percentual de pastagens nos recortes de referência NCX e SLMB                                          | 67 |
| 4.5  | Correlação entre as frações de abundância e os percentuais de pastagens                                | 68 |
| (a)  | Recorte NCX                                                                                            | 68 |
| (b)  | Recorte SLMB                                                                                           | 68 |
| 4.6  | Gráfico da seleção do coeficiente <i>cm</i> - recorte NCX                                              | 69 |
| 4.7  | Gráfico da seleção do coeficiente <i>cm</i> - recorte SLMB                                             | 70 |
| 4.8  | Desencontro entre as grades NDVI x LAPIG250                                                            | 71 |
| (a)  | Grade NDVI 231,656 m                                                                                   | 71 |
| (b)  | Mapa LAPIG250 sobre grade NDVI                                                                         | 71 |
| (c)  | Grade do mapa LAPIG250                                                                                 | 71 |
| (d)  | Grades NDVI e LAPIG250                                                                                 | 71 |
| 4.9  | Recorte CDG                                                                                            | 73 |
| 4.10 | Recorte CLV                                                                                            | 74 |
| 4.11 | Recorte LUZ                                                                                            | 75 |
| 4.12 | Classificador NCX aplicado ao recorte REC39K                                                           | 76 |
| (a)  | Recorte REC39K                                                                                         | 76 |
| (b)  | Mapa LAPIG030 sobre recorte REC39K                                                                     | 76 |
| (c)  | Pastagens classificadas pelo modelo NCX                                                                | 76 |
| (d)  | Erros de Omissão - Não Pastagens                                                                       | 76 |

---

## Lista de Tabelas

---

|     |                                                                                             |    |
|-----|---------------------------------------------------------------------------------------------|----|
| 2.1 | Recortes de referência e de classificação                                                   | 23 |
| 2.2 | Tamanho real dos <i>pixels</i> MODIS em latitude zero                                       | 27 |
| 3.1 | Tabela de pixels nos quais foram inseridos ruídos                                           | 39 |
| 3.2 | Matrizes de confusão das clusterizações com distâncias Euclidiana e DTW                     | 48 |
| 3.3 | Matrizes de confusão das clusterizações com os algoritmos <i>k-means</i> e <i>k-medoids</i> | 50 |
| 3.4 | Tabela de seleção do coeficiente <i>cm</i> – recorte NCX                                    | 58 |
| 4.1 | Tabela de seleção do coeficiente <i>cm</i> - recorte NCX                                    | 69 |
| 4.2 | Tabela de seleção do coeficiente <i>cm</i> - recorte SLMB                                   | 70 |
| 4.3 | Matriz de confusão - recorte SLMB                                                           | 71 |
| 4.4 | Matriz de confusão - recorte NCX                                                            | 72 |
| 4.5 | Matriz de confusão - recorte CDG                                                            | 73 |
| 4.6 | Matriz de confusão - recorte CLV                                                            | 74 |
| 4.7 | Matriz de confusão - recorte LUZ                                                            | 75 |
| 4.8 | Matriz de confusão - recorte REC39K                                                         | 76 |
| 4.9 | Desempenho dos modelos de classificação                                                     | 77 |

---

## Lista de Abreviaturas

---

|           |                                                                                         |
|-----------|-----------------------------------------------------------------------------------------|
| ANC       | Abundance Non-negativity Constraint                                                     |
| ASC       | Abundance Sum-to-one Constraint                                                         |
| CSV       | Comma Separated Values                                                                  |
| DTW       | Dynamic Time Warping                                                                    |
| EOS       | Earth Observing System                                                                  |
| GEOTIFF   | Geographic Tagged Image File Format                                                     |
| GIS       | Geographic Information System                                                           |
| LAPIG     | Laboratório de Processamento de Imagens e Geoprocessamento da UFG                       |
| LAPIG030  | Mapa gerado pelo LAPIG com origem no satélite Landsat 8, com 30 m de resolução espacial |
| LAPIG250  | Mapa gerado pelo LAPIG com origem no sensor MODIS, com 250 m de resolução espacial      |
| MLME      | Modelo Linear de Mistura Espectral                                                      |
| MLMT      | Modelo Linear de Mistura Temporal                                                       |
| MOD13Q1   | Mapa de índices de vegetação gerado a partir do sensor MODIS                            |
| MODIS     | Moderate Resolution Imaging Spectroradiometer                                           |
| NASA      | National Aeronautics and Space Administration                                           |
| NDVI      | Normalized Difference Vegetation Index                                                  |
| NNLS      | Non-Negative Least Squares                                                              |
| OLI       | Operational Land Imager                                                                 |
| PAM       | Partitioning Around Medoids                                                             |
| PASTL8MOD | Mapa resultado da associação do LAPIG030 com o produto MOD13Q1                          |
| PPI       | Pixel Purity Index                                                                      |
| SITS      | Satellite Image Time Series - SITS                                                      |
| UCS       | Unconstrained Least Squares                                                             |
| USGS      | U.S. Geological Survey                                                                  |

## Introdução

---

Nas últimas décadas, houve um aumento considerável no consumo de alimentos *per capita* e o padrão de alimentação mundial vem migrando de uma dieta básica, composta de raízes e tubérculos, para uma dieta mais rica em óleos vegetais e produtos animais [Alexandratos e Bruinsma 2012]. Em seu *Statistical Pocketbook* de 2018 [FAO 2018], a Organização das Nações Unidas para a Alimentação e a Agricultura informa que, em 2016 a agricultura era responsável pela ocupação de um terço do total de terras, sendo que aproximadamente 75% deste quantitativo correspondia a áreas de pastagens. No Brasil, segundo estudo efetuado pelo Projeto TerraClass [ALMEIDA et al. 2016], as pastagens cobriam, no ano de 2008, 62% da área desflorestada da Amazônia Legal Brasileira.

A expansão de terras para o desenvolvimento da pecuária, contudo, traz consigo implicações ambientais tais como o desmatamento das florestas e precisa de uma atenção especial. Mais do que quantificar uma área ocupada por pastagens, é importante conhecer sua localização e sua utilização ao longo do tempo. A constatação de que não houve crescimento na área total de pastagens em um determinado período pode significar, por exemplo, que tenha havido a conversão de algumas delas em áreas de cultivo, aliada a um desmatamento equivalente às áreas convertidas [Parente e Ferreira 2018].

Para entender as mudanças climáticas e as variações na cobertura e uso do solo, a NASA, por meio de um programa chamado *Earth Observing System (EOS)*, mantém um conjunto de satélites que geram, em intervalos e órbitas regulares, dados de observação terrestre, denominados de séries temporais de imagens de satélite (*Satellite Image Time Series - SITS*). Dentre vários produtos que são obtidos a partir destas séries temporais, estão os índices de vegetação, que são utilizados para medir a densidade de verde em uma determinada região, sendo que o mais comum destes índices é o *Normalized Difference Vegetation Index – NDVI* [NASA/NDVI].

Os satélites AQUA e TERRA, ambos da constelação do programa EOS, possuem um sensor chamado MODIS (*Moderate Resolution Imaging Spectroradiometer*), a partir do qual é gerado um valor de NDVI a cada dezesseis dias, para toda a superfície terrestre. Graças a esta periodicidade, é possível acompanhar o comportamento fenológico dos diversos tipos de cobertura do solo, inclusive do objeto deste trabalho, que são as áreas de

pastagens.

Vários métodos e algoritmos têm sido empregados para a classificação de feições terrestres a partir das informações geradas pelas SITS. Alguns trabalhos usam métodos de classificação não supervisionada como, por exemplo, em [Gonçalves et al. 2017], onde é apresentada uma solução para monitoramento de plantações de cana de açúcar no Estado de São Paulo, por meio de clusterização de SITS de baixa resolução. Outros trabalhos, contudo, usam o processo de classificação supervisionada pixel a pixel como, por exemplo, [Parente et al. 2017] e [Parente e Ferreira 2018], que aplicaram o algoritmo *Random Forest* para produzirem mapas de pastagens abrangendo todo o território brasileiro.

Segundo [Theodoridis e Koutroumbas 2003], quando há o conhecimento prévio das características que determinam as classes de agrupamento de um conjunto de objetos, este conjunto é dito rotulado; por outro lado, quando não há conhecimento *a priori* de como as características dos dados definem suas classes, este conjunto é denominado como não rotulado. Havendo um conjunto de dados rotulados e se este conjunto pode ser usado para classificar outros objetos de natureza semelhante, aplicamos o método de classificação supervisionada. Se, por outro lado, o conjunto de dados é não rotulado, a alternativa é o método de classificação não supervisionada<sup>1</sup>, no qual procura-se descobrir os padrões dos objetos por meio de medidas de semelhança entre eles. Um terceiro método é a classificação semissupervisionada, quando há apenas uma pequena parcela dos dados rotulados e a busca de informações sobre os dados não rotulados pode ajudar na tarefa de classificação. Também no caso da clusterização o método semissupervisionado pode ser aplicado fazendo com que a pequena parcela de dados rotulados sejam utilizados como moderador para incluir ou não um ponto em um determinado cluster.

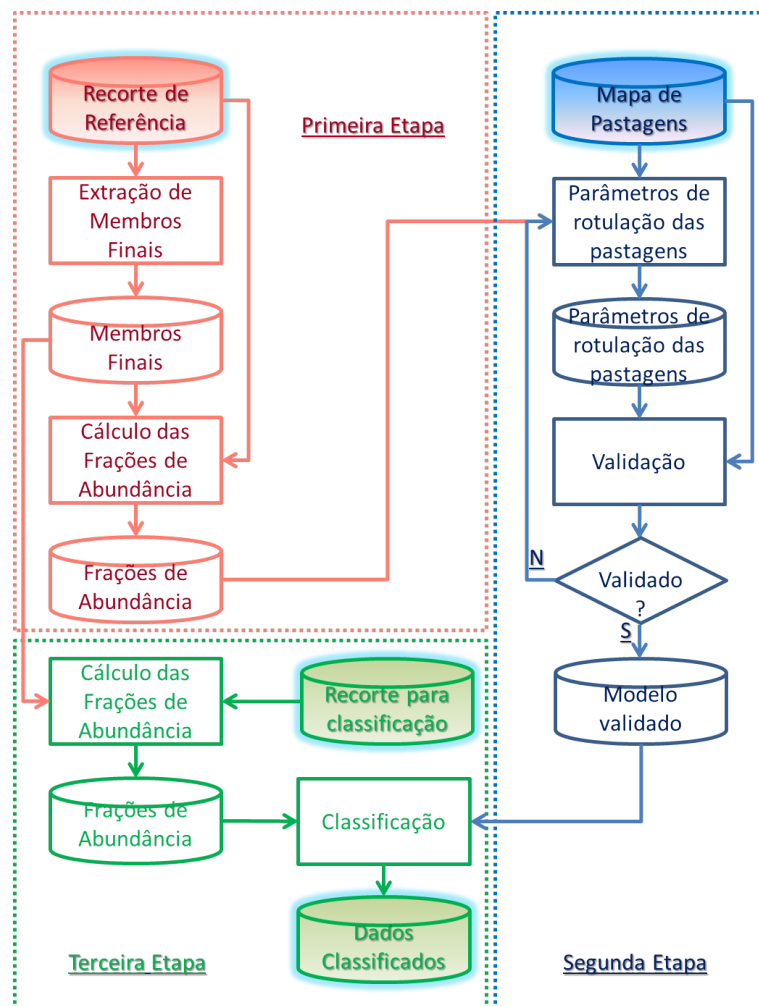
Esta dissertação apresenta uma abordagem semissupervisionada para a classificação de pastagens a partir de séries temporais de NDVI e é composta de três etapas, conforme resumido na figura 1.1. A primeira etapa inicia com uma classificação não supervisionada, baseada no Modelo Linear de Mistura Temporal – MLMT. Este método é uma adaptação do Modelo Linear de Mistura Espectral – MLME que usa dados temporais no lugar de dados espectrais. O produto do MLMT é uma matriz de abundâncias na qual os valores temporais de NDVI de cada pixel são substituídos por uma combinação linear de valores que representam as proporções com que alguns *pixels* do conjunto – denominados de membros finais – conseguem representar as características do pixel em questão.

Na segunda etapa é construído um modelo de classificação no qual os *pixels* da matriz de abundâncias são rotulados a partir de um mapa de pastagens, utilizado como

---

<sup>1</sup>O meio acadêmico tem adotado o termo "clusterização"(um anglicismo adaptado ao Português, do original em Inglês *clustering*), para designar método de classificação não supervisionada





**Figura 1.1:** Resumo da metodologia adotada neste trabalho

verdade de campo. São definidos, também, os valores de abundância que representam as áreas de pastagens a serem aplicados na classificação de novas áreas.

A terceira etapa consiste na aplicação do modelo de classificação a um novo conjunto de dados com base nas frações de abundância já calculadas e nos critérios de seleção dos *pixels* de pastagens definidos na segunda etapa.

## 1.1 Objetivos e contribuições

Os principais objetivos deste trabalho são a apresentação de uma metodologia de classificação de pastagens com baixo custo de recursos computacionais, materiais e humanos, produzindo resultados com qualidade compatível com outras iniciativas que usaram os mesmos materiais, mas que aplicaram métodos distintos.

### 1.1.1 Contribuições

Nesta metodologia apresentamos o filtro MaxMin, para o tratamento de ruídos, *outliers* e dados faltantes em séries temporais de NDVI. Este filtro, que foi validado nesta metodologia por este autor e desenvolvido por seu orientador, mostrou-se adequado a este trabalho por ser mais sensível na diferenciação entre os ruídos efetivos e as variações das séries temporais de NDVI causadas pelas características fenológicas da vegetação e pela intervenção humana.

Apresentamos também uma comparação entre o processo de seleção de membros finais por meio de uma clusterização e um dos processos tradicionais, via algoritmo PPI.

Ainda como contribuição, apresentamos uma metodologia para rotulagem de dados a partir da associação de um mapa com frações de abundância com um mapa contendo as áreas de pastagens.

## 1.2 Organização da dissertação

Este Capítulo abordou a necessidade do monitoramento das áreas de pastagens, tendo em vista sua grande importância na produção de alimentos e também por serem uma das principais causas de desmatamentos das florestas. Uma das maneiras de se fazer esse monitoramento é por meio da tecnologia de sensoriamento remoto, que produz dados periódicos de observação terrestre – as séries temporais. Dentre as várias informações disponibilizadas em forma de série temporal, estão os índices de vegetação, sendo que o NDVI é o mais popular dentre eles. Foram abordados os métodos de classificação de dados, que se dividem em métodos não supervisionados, supervisionados e semissupervisionados. Por fim, foi feito um resumo sobre a metodologia utilizada para a classificação semissupervisionada de pastagens a partir de séries temporais de NDVI.

O Capítulo 2 discorre sobre as características e a localização geográfica das regiões de interesse deste trabalho, bem como os materiais que foram utilizados na aplicação da metodologia. É apresentado um detalhamento da série temporal de NDVI produzida pelo sensor MODIS e também dos mapas de pastagens produzidos pelo LAPIG. Ao final do Capítulo, são listados os softwares e linguagens que foram utilizados neste trabalho.

O Capítulo 3 apresenta a metodologia deste trabalho, com o detalhamento de quais foram os algoritmos utilizados e qual foi a sequência de processamento obedecida. Quando necessário, foram apresentados os trabalhos e a teoria relativos ao algoritmo sendo discutido. Nos casos em que houve necessidade de escolher entre dois algoritmos, consta o estudo que justificou a opção por um deles.

O Capítulo 4 apresenta os resultados que foram obtidos em cada uma das etapas, bem como as avaliações e discussões que tais resultados geraram. A Conclusão está no Capítulo 5, seguida das Referências Bibliográficas, que encerram o presente documento.

## Materiais

### 2.1 Regiões de interesse

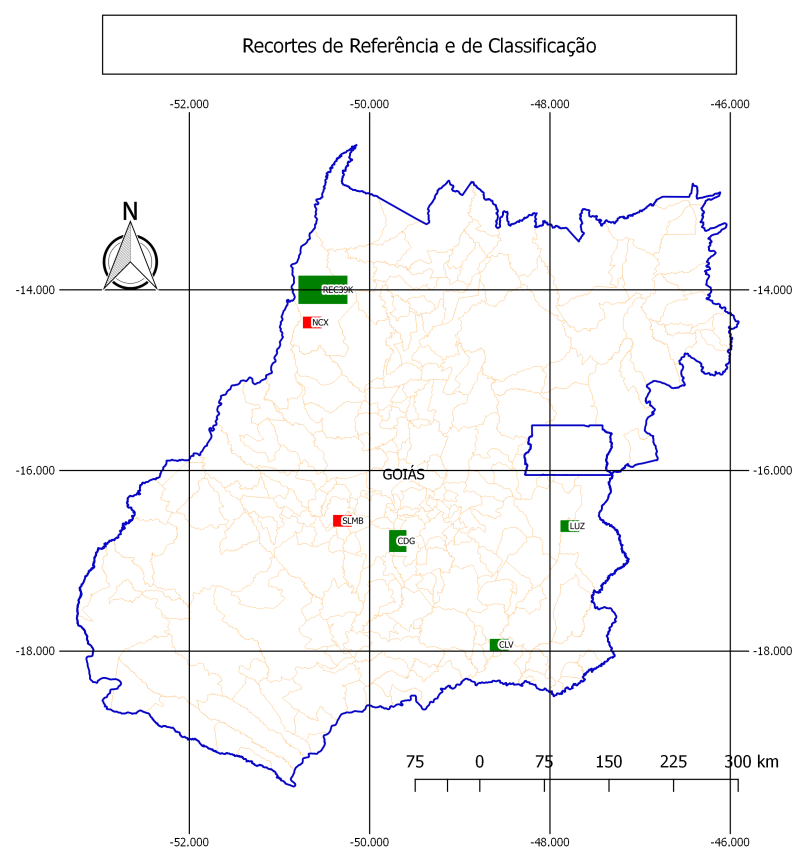
Conforme previsto em nossa metodologia, um recorte espacial de uma série temporal de NDVI é usado como referência para a classificação de séries temporais de outros recortes. Estes recortes foram selecionados de modo a garantir que as imagens representassem a diversidade de coberturas do solo, tais como pastagens, florestas, áreas de agricultura, solo exposto, áreas urbanas etc. Foram selecionadas duas áreas para serem usadas como recortes de referência (na cor vermelha) e outras quatro áreas na condição de recortes de classificação (na cor verde), conforme figura 2.1. Inicialmente, havia a previsão de uso de apenas três recortes de classificação mas, durante a aplicação da metodologia e em decorrência dos resultados obtidos, optou-se por criar um quarto recorte, o REC39K, com maior proximidade geográfica e com feições semelhantes às de um dos recorte de referência que, no caso, foi o recorte NCX.

A tabela 2.1 mostra a localização geográfica de cada um dos recortes, sua área em quilômetros quadrados, as quantidades de linhas, colunas e o respectivo número de pixels relativo ao grid NDVI, descrito a seguir na Seção 2.2.

**Tabela 2.1:** *Recortes de referência e de classificação*

| RECORTE | MUNICÍPIO                 | NORTE    | SUL      | LESTE    | OESTE    | ÁREA(km <sup>2</sup> ) | LINHAS | COLUNAS | PIXELS |
|---------|---------------------------|----------|----------|----------|----------|------------------------|--------|---------|--------|
| NCX     | NOVA CRIXÁS               | -14,2967 | -14,4263 | -50,5318 | -50,7410 | 327,44                 | 62     | 100     | 6200   |
| SLMB    | SÃO LUIZ DOS MONTES BELOS | -16,4925 | -16,6222 | -50,1972 | -50,4064 | 323,86                 | 62     | 100     | 6200   |
| CLV     | CALDAS NOVAS              | -17,8644 | -17,9941 | -48,4594 | -48,6685 | 321,44                 | 62     | 100     | 6200   |
| CDG     | CAMPESTRE DE GOIÁS        | -16,6598 | -16,9024 | -49,5928 | -49,7852 | 556,86                 | 116    | 92      | 10672  |
| LUZ     | LUZIANIA                  | -16,5511 | -16,6807 | -47,6751 | -47,8843 | 323,80                 | 62     | 100     | 6200   |
| REC39K  | NOVA CRIXÁS               | -13,8429 | -14,1565 | -50,2474 | -50,7912 | 2.063,07               | 150    | 260     | 39000  |

O recorte SLMB é um dos recortes de referência usados para a construção dos modelos de classificação. Ele está localizado na cidade de São Luiz de Montes Belos e tem como características a presença de áreas urbanas, mata densa, agricultura e pastagens (figura 2.2).



**Figura 2.1:** Regiões de Interesse – Recortes de referência (em vermelho) e de classificação (em verde)



**Figura 2.2:** Recorte de referência – SLMB

O recorte NCX, localizado na cidade de Nova Crixás, abrange basicamente áreas de pastagens e mata densa (figura 2.3).



**Figura 2.3:** Recorte de referência – NCX

## 2.2 Séries temporais para classificação de pastagens

### 2.2.1 O índice de vegetação NDVI

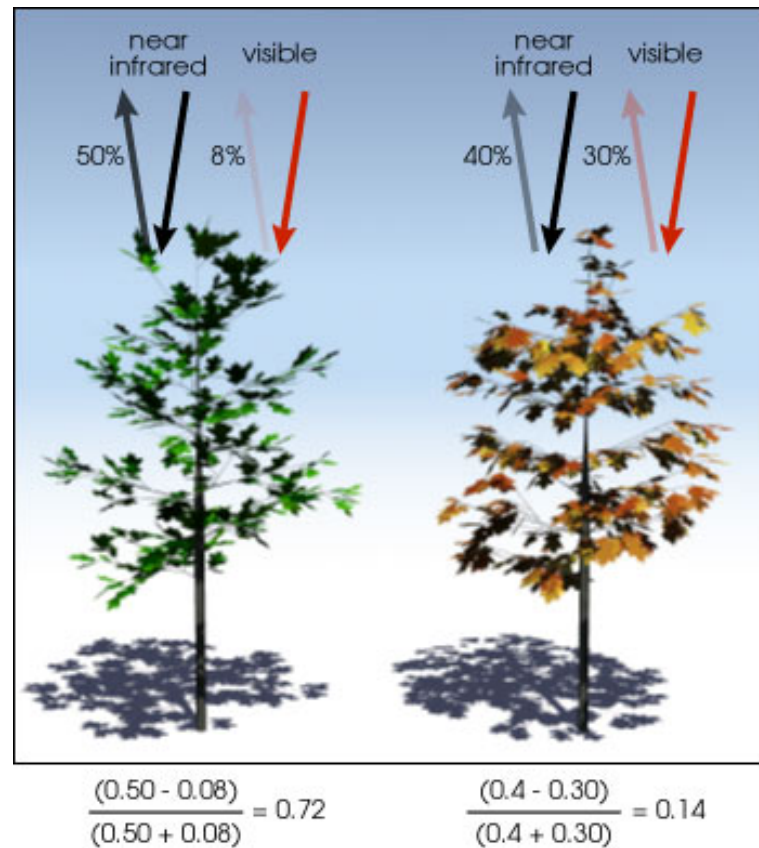
Os índices de vegetação são indicadores utilizados para medir o vigor e a saúde da vegetação e usualmente são obtidos a partir de informações coletadas por sensores remotos embarcados em satélites, que captam a reflexão da luz solar em várias faixas de comprimentos de onda – as chamadas bandas espectrais. Os índices de vegetação são calculados a partir de combinações entre estas bandas, principalmente entre aquelas situadas nas regiões do vermelho (*Red*) e do infravermelho próximo (*Near Infrared – NIR*) [Abdou et al. 1996].

Segundo [NASA/NDVI], a clorofila, que é o pigmento das folhas das plantas, absorve um grande percentual da luz visível (intervalo de 400 a 700 *nm*) e reflete uma grande parte da radiação na faixa do infravermelho próximo (intervalo de 700 a 1100 *nm*). O NDVI, um dos índices de vegetação mais utilizados, tem valores no intervalo de [-1,+1] e é calculado a partir dos valores de reflexão da luz nas bandas *Red* e *NIR*, de acordo com a equação 2-1 e ilustrado na figura 2.4.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (2-1)$$

### 2.2.2 Séries temporais de NDVI

O MODIS é um instrumento embarcado nos satélites AQUA e TERRA, ambos da constelação do programa EOS e que recobrem toda a superfície terrestre a cada dois dias [NASA/EARTHDATA]. Seus sensores captam a reflexão da radiação solar em 36

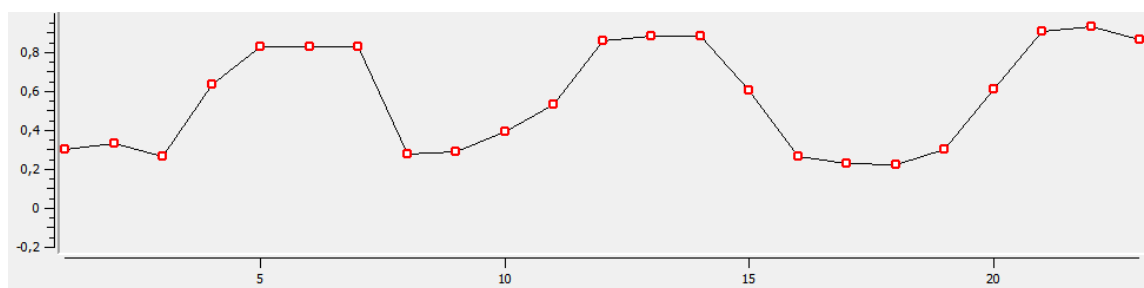


**Figura 2.4:** Cálculo do NDVI a partir da radiação solar

Fonte: [NASA/NDVI]

bandas espectrais entre  $0,405 \mu m$  e  $14,385 \mu m$  e produzem dados com resoluções espaciais aproximadas de 250 m, 500 m e 1.000 m [NASA/MODIS].

Um dos produtos gerados pelo MODIS é o MOD13Q1, que fornece, a cada dezesseis dias, um valor de NDVI para cada pixel da superfície terrestre [Didan 2015]. O algoritmo de geração do MOD13Q1 escolhe, dentre os valores coletados para cada pixel, o que tiver menor interferência de nuvens, o melhor ângulo de visão e o maior valor do período. A figura 2.5 mostra um exemplo da série temporal de um pixel com suas 23 leituras ao longo de um ano.



**Figura 2.5:** Perfil NDVI de um pixel durante o ano de 2015

Os dados do MODIS são gerados na projeção sinusoidal contínua, que mantém a área projetada igual à área original da superfície terrestre. Estes dados são identificados

pelo tamanho aproximado de seus *pixels*, cujo tamanho real difere de seus identificadores, como descrito na tabela 2.2 [Giglio L.].

**Tabela 2.2:** *Tamanho real dos pixels MODIS em latitude zero*

| Identificador | Tamanho   |
|---------------|-----------|
| 250 m         | 231,656 m |
| 500 m         | 463,313 m |
| 1000 m        | 926,625 m |

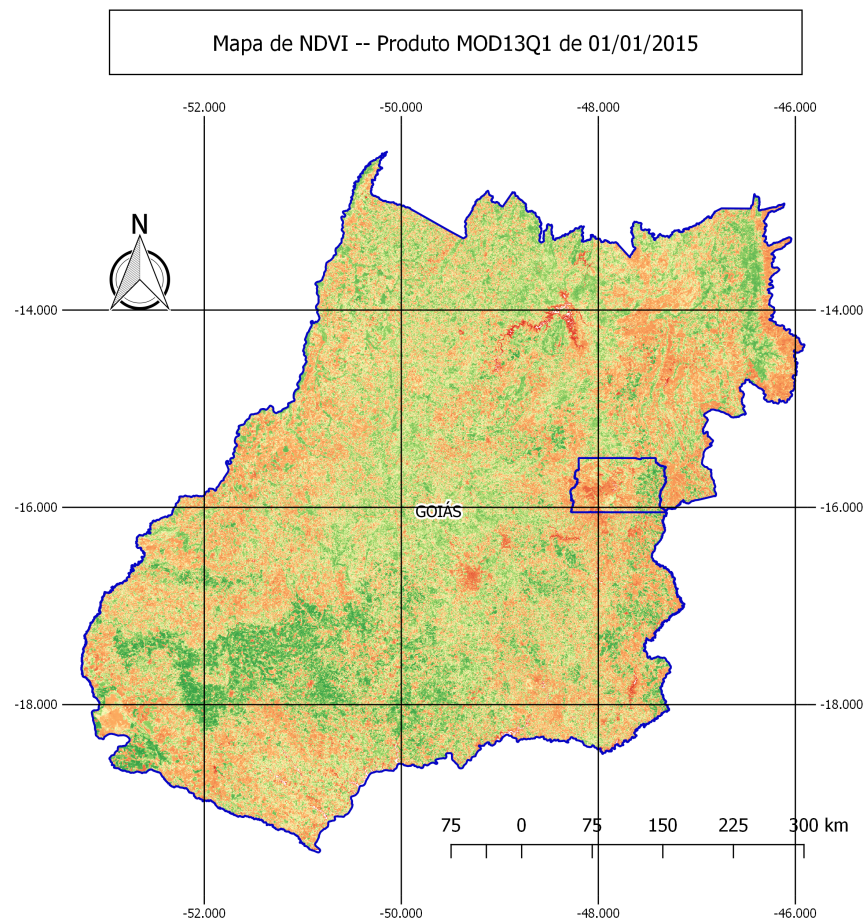
A NASA disponibiliza, para usuários cadastrados, a possibilidade de efetuar *download* das séries temporais de NDVI de forma gratuita [NASA/EARTHDATA]. O LAPIG mantém uma base de dados do produto MODIS/MOD13Q1 para o território brasileiro, a partir da qual foi feita a extração dos dados de NDVI referentes ao Estado de Goiás do período 2000 – 2018, nos formatos GEOTIF e CSV. Cada pixel das imagens GEOTIF possui um identificador que também está presente nos dados em formato CSV, possibilitando a conexão entre estas duas formas de representação dos dados. Os dados em formato CSV foram carregados localmente em uma base de dados Postgre/SQL no formato PostGIS e foram utilizados neste trabalho em conjunto com os dados GEOTIF. Para manter a compatibilidade temporal com os mapas de pastagens descritos nos itens 2.3.1 e 2.3.2, os dados referentes aos recortes de referência e de classificação corresponderam ao ano de 2015.

A figura 2.6 mostra um mapa temático dos valores de NDVI referentes aos dados gerados pelo MODIS em 01/01/2015. Os dados foram normalizados para o intervalo [0,1] sendo que, quanto mais próximo de zero (cor vermelha), menos vegetação e, quanto mais próximo de um, mais vegetação (cor verde).

## 2.3 Mapas de pastagens

Segundo [Parente et al. 2017], até recentemente a melhor informação disponível sobre as pastagens no território brasileiro eram os mapas de cobertura vegetal produzidos em 2002 pelo Programa Brasileiro para Conservação e Uso Sustentável da Biodiversidade Brasileira (PROBIO). Neste trabalho foram utilizados dois mapas de pastagens produzidos pelo LAPIG a partir de 2015. O primeiro deles, que tem sua origem em imagens do satélite Landsat 8, foi usado como conhecimento especializado na etapa de rotulação dos dados de pastagens e na validação dos resultados. O segundo mapa utilizado foi construído a partir de séries temporais NDVI do produto MOD13Q1 e serviu de comparação com o processo de classificação executado, tendo em vista possuírem características semelhantes.





**Figura 2.6:** Mapa Temático de NDVI do Estado de Goiás -  
Referência 01/01/2015

### 2.3.1 Mapa de pastagens – origem Landsat 8

O Landsat 8 é um dos satélites que fazem parte do programa EOS e foi lançado em 2013 numa parceria com a *U.S. Geological Survey (USGS)*. Ele libera novas cenas a cada 16 dias e um de seus sensores embarcados é o *Operational Land Imager (OLI)*, que capta nove bandas espectrais com resolução de 30 m. Dentre elas, estão as bandas *Red* e *NIR*, utilizadas no cálculo do NDVI [NASA/Landsat8].

O mapa de pastagens gerado a partir das imagens do satélite Landsat 8 referentes ao ano de 2015, faz parte da iniciativa Mapbiomas e está disponível para *download* para usuários cadastrados [LAPIG-UFG 2017]. Foram necessárias 380 cenas para cobrir toda a área mapeada, totalizando 6.700 imagens. As bandas espectrais que foram utilizadas para o cálculo de diversos índices, dentre os quais o índice de vegetação NDVI, foram as bandas verde (*Green*), vermelha (*Red*), infravermelha próxima (*NIR*), infravermelha de ondas curtas (*SWIR*) e infravermelha termal (*TIR*).

O conjunto de dados de treinamento foi gerado de maneira automática e aleatória, em quantidades iguais de *pixels* de pastagem e não pastagem, segundo um mapa

síntese de pastagens previamente elaborado pela própria equipe do LAPIG. Os dados de treinamento foram regionalizados de maneira que, para cada cena analisada, fossem selecionados 300 pontos extraídos da própria cena e mais 300 pontos de cada uma das oito cenas adjacentes, totalizando 2.700 pontos.

O algoritmo de classificação supervisionada utilizado foi o *Random Forest*<sup>1</sup> composto de 500 árvores de decisão, sendo que cada uma delas classificou os pixels como sendo ou não da classe pastagem. Computou-se o percentual de quantas árvores, do total de 500, classificaram um pixel como pastagem e estabeleceu-se um limiar de 60% para que ele fosse considerado como tal. O mapa obtido foi ainda submetido a melhorias com base em máscaras de áreas urbanas, unidades de conservação etc, até a sua versão final.

A validação do processo foi feita pela análise visual de 5.000 pontos (2.500 de cada classe), executada por uma equipe de cinco profissionais treinados para a tarefa. Foram utilizadas imagens Landsat 8 das estações chuvosa e seca, imagens de alta resolução do Google Earth, séries temporais MODIS NDVI e informações auxiliares. Apenas os pontos para os quais houve concordância de, pelo menos, três dos analistas, foram considerados para a avaliação da qualidade. A acurácia geral foi de 87%, enquanto que as acurácias do produtor e do usuário alcançaram 90% e 83%, respectivamente (veja definições destas medidas de acurácia no item 3.2.3).

Constatou-se que, para 80% do território brasileiro, houve disponibilidade média de apenas 12 cenas livres de nuvens e de sombras de nuvens, das 23 cenas geradas anualmente. O bioma Cerrado, para o qual houve a melhor taxa de disponibilidade, teve 56% de sua área com 12 ou mais cenas de boa qualidade; por outro lado, 35% do bioma Amazonas teve uma ou nenhuma cena de boa qualidade no ano inteiro, muito em função da alta concentração de nuvens e da baixa disponibilidade de pontos de controle no solo para um efetivo processo de correção dos dados.

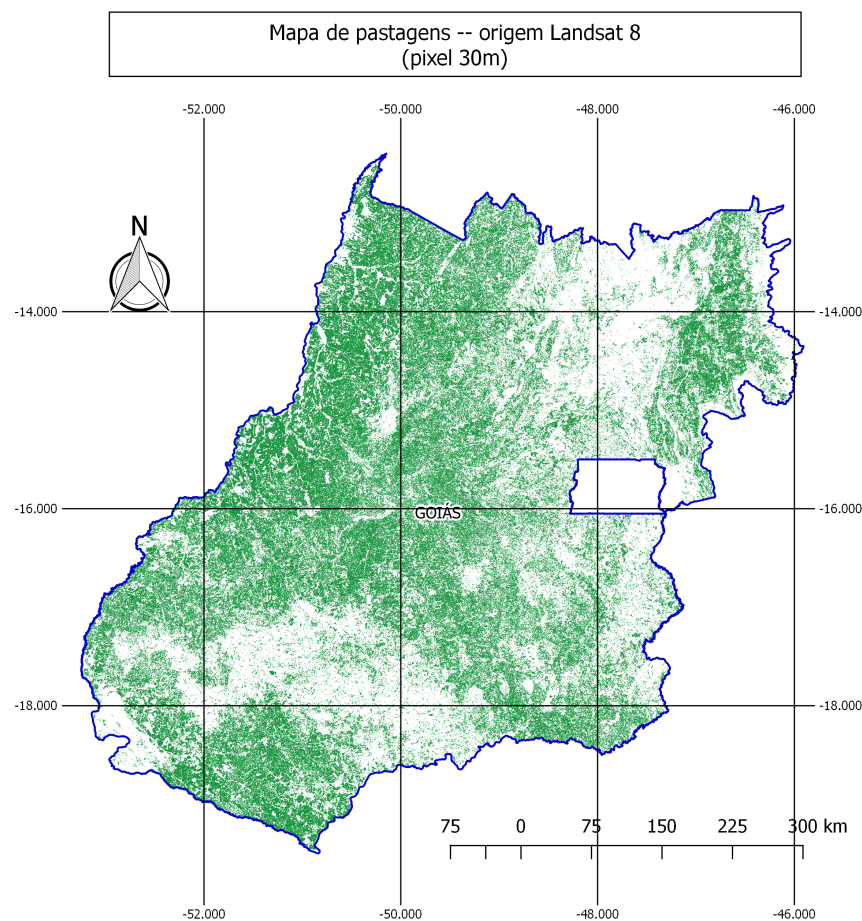
A figura 2.7 mostra as pastagens do Estado de Goiás segundo o mapa produzido a partir das imagens do satélite Landsat 8. Para facilitar sua identificação, ele será denominado neste documento de LAPIG030, em função da resolução espacial do satélite Landsat 8.

### 2.3.2 Mapa de pastagens – origem MOD13Q1

O LAPIG produziu, também, outro mapa contendo as áreas de pastagens do território brasileiro no período de 2000 a 2016 [Parente e Ferreira 2018]. Como fon-

---

<sup>1</sup>*Random Forest* (floresta aleatória) é um classificador que consiste em uma coleção de classificadores estruturados em árvores  $\{h(x, \theta_k), k = 1, \dots\}$  em que os  $\{\theta_k\}$  são vetores aleatórios distribuídos de forma idêntica e independentes e cada árvore emite um voto unitário para a classe mais popular na entrada  $x$ . [Breiman e Cutler 2001]



**Figura 2.7:** Mapa de pastagens de Goiás – LAPIG030  
Ano 2015

tes de dados, foram utilizadas séries temporais de NDVI geradas pelo sensor MODIS [Didan 2015], que são disponibilizados a cada 16 dias, totalizando 23 emissões anuais. Foram selecionados os *pixels* que não continham nuvens ou sombras de nuvens, de acordo com informações de qualidade dos *pixels* contidas no produto MOD13Q1.

O mapa LAPIG030 foi utilizado como fonte de dados para o treinamento do algoritmo que gerou este mapa. Os seus *pixels* foram reamostrados, da resolução espacial de 30 m do Landsat 8 para a resolução espacial de 250 m do MODIS. Para cada ano do período, foi construído um conjunto de treinamento a partir da grade de 380 cenas do Landsat 8, de maneira que, para cada cena analisada, fossem selecionados 300 pontos extraídos da própria cena e mais 300 pontos de cada uma das oito cenas adjacentes, totalizando 2.700 pontos para cada cena.

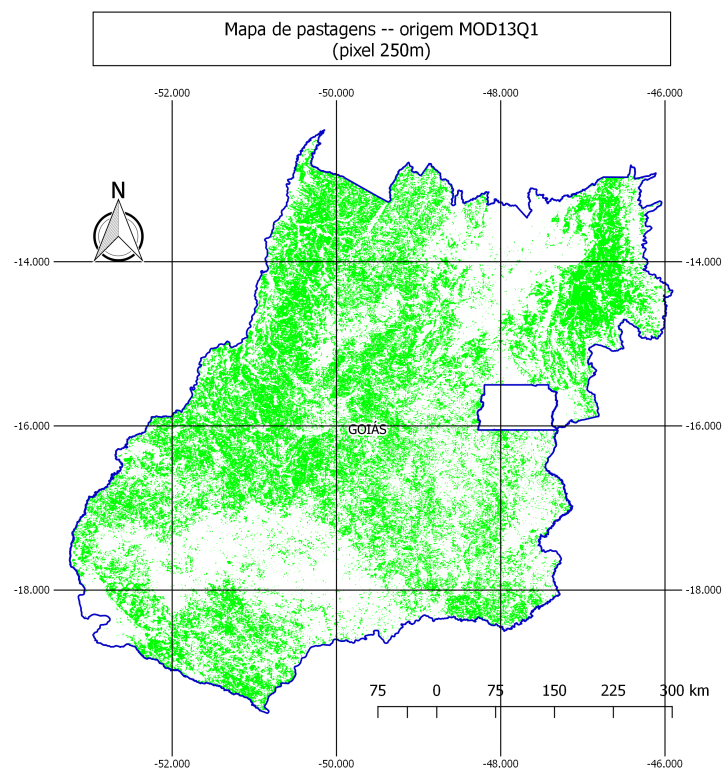
Foi feita uma primeira classificação para cada ano do período, considerando os dados de treinamento do mapa LAPIG030, na qual foi aplicado o algoritmo de classificação supervisionada *Random Forest* com 500 árvores de decisão. Cada pixel foi classificado, em cada uma das 500 árvores, como sendo ou não da classe pastagem e,

aqueles que foram classificados como pastagens por mais de 60% das 500 árvores, assim o foram considerados. Uma segunda classificação foi executada, sendo que, desta vez os dados de treinamento foram obtidos a partir do resultado da primeira classificação, mas o processo de seleção dos *pixels* foi o mesmo da primeira classificação, ou seja, 2.700 pixels por cena. Ao final, foram treinados 12.920 modelos, resultado de 380 cenas Landsat 8 por cada um dos 17 anos do período, em duas classificações.

À semelhança do que havia sido feito na validação do mapa LAPIG030, neste também foram analisados 5.000 pontos (2.500 de cada classe), sendo que foram selecionados para o processo de avaliação de qualidade, apenas aqueles nos quais houve concordância de pelo menos três dos cinco analistas treinados para a tarefa. A análise foi feita com base em 34 imagens Landsat 8 geradas ao longo do período de 2000 a 2016 (2 imagens por ano), tendo a acurácia geral para o período ficado em torno de 80%.

Na discussão dos resultados, os autores concluíram que os dados do sensor MODIS, apesar de terem uma baixa resolução espacial, possuem uma alta taxa de disponibilidade de observações livres de nuvens ou de sombras de nuvens, numa média de 18 observações anuais por pixel, suficientes para a detecção de variações sazonais nas áreas de pastagens.

A figura 2.8 mostra as pastagens do Estado de Goiás referentes ao ano de 2015, segundo o mapa produzido a partir dos dados gerados pelo sensor MODIS. Para facilitar sua identificação, ele será denominado neste documento de LAPIG250, em função da resolução espacial do sensor MODIS.



**Figura 2.8:** Mapa de pastagens de Goiás – LAPIG250  
Ano 2015

## 2.4 Softwares e linguagens

Os softwares e linguagens utilizados neste estudo são do tipo de código aberto, sob a Licença Pública Geral GNU. (*General Public Licence GNU*<sup>2</sup>).

- **RStudio (Versão 1.0.153)** – é um ambiente de desenvolvimento integrado para a linguagem de programação **R (Versão 3.4.1)**, que é voltada, principalmente, para a geração de gráficos e execução de cálculos estatísticos [RStudio].
- **QGIS (Versão 2.18.27)** – é um Sistema de Informações Geográficas (*Geographic Information System – GIS*) para processamento e análise de dados espaciais e que trabalha com vários formatos de dados vetoriais e raster. Pode-se conectar com diversas fontes de dados espaciais, inclusive bancos de dados e servidores de mapas, locais ou remotos [QGIS].
- **PostGIS (Versão 2.4.4 - r16526)** – é um módulo do banco de dados **PostgreSQL (Versão 10.4)** para o armazenamento, processamento e análise de objetos geoespaciais cujo manuseio é feito por meio da linguagem **SQL** [PostGIS].

---

<sup>2</sup>GNU é um sistema operacional e um conjunto de softwares livres e a sua sigla é um acrônimo recursivo para "GNU's Not Unix!"

## Metodologia

---

Este capítulo apresenta uma metodologia em três etapas que foi elaborada para a classificação semissupervisionada de pastagens a partir de séries temporais de NDVI. Ao descrever seus métodos e algoritmos, também foram abordados trabalhos correlatos e seus referenciais teóricos. Quando houve necessidade de escolha entre dois algoritmos, foi apresentado o estudo que justificou a opção por um deles.

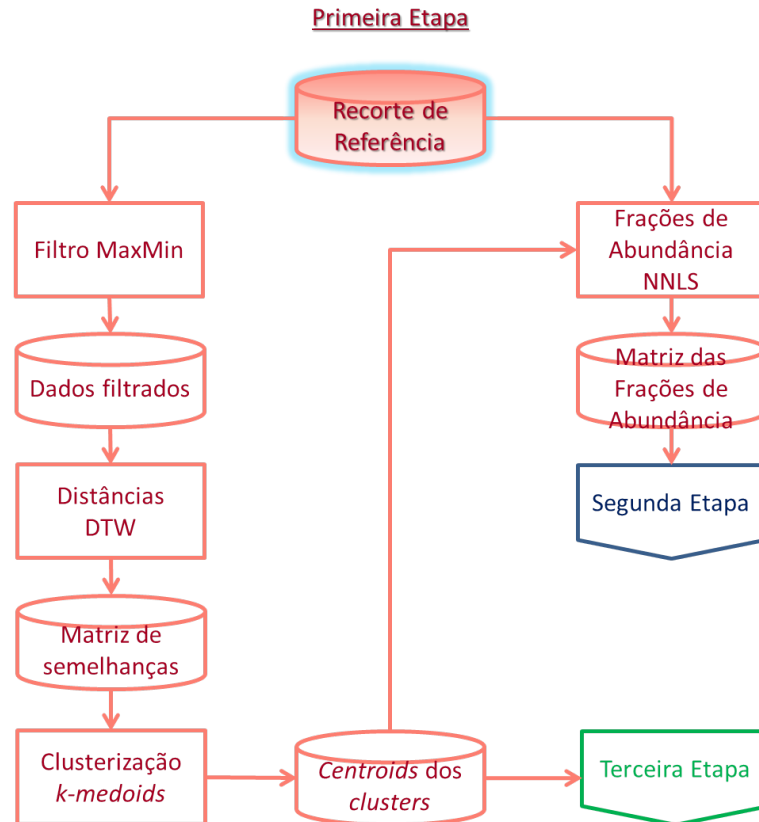
A primeira etapa é uma classificação do tipo não supervisionada, baseada no Modelo Linear de Mistura Temporal – MLMT. Neste método os *pixels* de um recorte espacial de referência de uma série temporal de NDVI são agrupados a partir de suas características e, depois, um representante de cada um destes grupos é utilizado para compor um conjunto denominado de membros finais (ou *endmembers*). O passo seguinte é o cálculo das frações de abundância, que são as proporções com que cada membro final participa, por meio de uma combinação linear, da composição de cada um dos *pixels* misturados.

A segunda etapa é a construção do modelo de classificação supervisionada, que tem como insumo as frações de abundância e um conjunto de polígonos referentes às áreas de pastagens da mesma região do recorte de referência, que é utilizado como fonte para a rotulação. No caso do presente trabalho, os polígonos foram obtidos a partir do mapa de pastagens LAPIG030 descrito no item 2.3. Os critérios para identificação e rotulação dos *pixels* como pastagens são definidos a partir da análise da correlação existente entre as frações de abundância e o percentual de pastagens existente em cada pixel. Este percentual é calculado a partir da sobreposição de uma grade do recorte de referência com o mapa de pastagens. Depois que os pixels tiverem sido rotulados, obedecendo aos critérios definidos anteriormente, é feita uma validação da rotulação com o mapa de pastagens, pixel a pixel.

A terceira etapa consiste na classificação de um novo conjunto de dados aplicando-se as frações de abundância já calculadas e os critérios de seleção dos *pixels* de pastagens definidos na segunda etapa.

### 3.1 Classificação não supervisionada

A figura 3.1 mostra o diagrama detalhado desta primeira etapa, na qual são obtidos os membros finais e calculadas as frações de abundância a serem utilizados nas etapas seguintes.



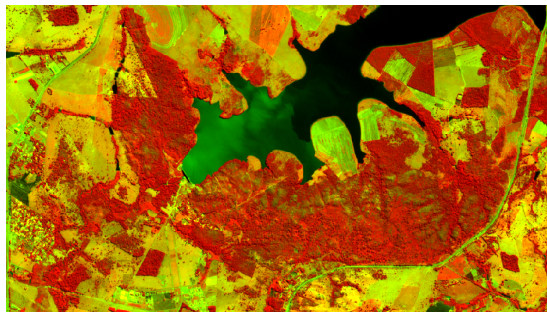
**Figura 3.1:** Primeira etapa da metodologia

#### 3.1.1 Modelo Linear de Mistura Espectral - MLME

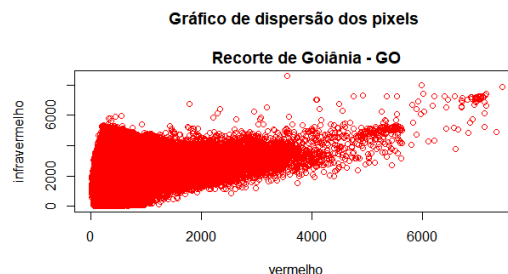
Segundo [Shimabukuro e Ponzoni 2019], dois problemas estão associados à relação entre a resolução espacial dos sensores remotos e o tamanho dos objetos na superfície sendo imageada. O primeiro é quando os objetos são menores do que a resolução espacial do sensor e um pixel da imagem contém uma mistura dos fluxos de radiação de diferentes objetos. O segundo problema é quando o objeto é maior do que a resolução espacial do sensor e ocupa um ou mais *pixels* de forma parcial. Ainda há uma terceira situação na qual um pixel contém apenas um tipo de objeto e é denominado de pixel puro. Os *pixels* puros, quando representados em um gráfico de dispersão, ocupam as extremidades da figura geométrica formada pelo gráfico e são denominados de **membros finais** (ou *end-members*). Decorre daí que os *pixels* localizados no interior da figura geométrica são uma mistura dos membros finais em variadas proporções denominadas de **frações de abun-**



**dância.** A figura 3.2 mostra um gráfico de dispersão de uma imagem com duas bandas espectrais (vermelho e infravermelho) formando uma figura aproximadamente triangular, cujos vértices representam os elementos naturais vegetação, solo e água (ou sombra).



(a) Recorte Goiânia (Bandas Vermelho + Infravermelho)



(b) Gráfico de dispersão dos pixels da imagem

**Figura 3.2:** Recorte de Goiânia e Gráfico de dispersão

O MLME, que foi construído para ser utilizado na análise de imagens multi ou hiperespectrais, tem sido utilizado também para análise de dados temporais [Ferreira 2003], onde as bandas multiespectrais são substituídas por bandas temporais.

### 3.1.2 Modelo Linear de Mistura Temporal – MLMT

O Modelo Linear de Mistura Temporal (MLMT) é uma variação do MLME, no qual as  $m$  bandas multiespectrais são substituídas por  $t$  bandas temporais, que contêm as informações fenológicas dos diferentes tipos de cobertura de solo. Segundo [Zhuo et al. 2018], o perfil temporal NDVI de um pixel misto pode ser considerado como uma combinação linear de perfis temporais NDVI de membros finais. [Chi, Kim e Kang 2016] afirmam que o MLMT é algebricamente semelhante ao MLME, mas que ele utiliza o domínio do tempo ao invés do domínio espectral. Ainda de acordo com [Chi, Kim e Kang 2016], o MLMT assume que as trajetórias temporais dos dados são linearmente independentes e que os *pixels* nas imagens estão em espaços lineares

### 3.1.3 Obtenção dos Membros Finais

A obtenção dos membros finais está intimamente ligada à natureza dos dados analisados, de maneira que os *pixels* que forem selecionados possam representar, por meio de uma combinação linear entre eles, todo o conjunto de *pixels* da imagem. Os membros finais podem ser obtidos de diversas maneiras, como assinaturas espectrais disponíveis em bibliotecas, por meio de análise e seleção visual diretamente na imagem ou, então, por meio de algoritmos. Tendo em vista que uma série anual de NDVI é composta por 23 valores, medidos a cada 16 dias, não há representação visual que suporte esta quantidade



de informações simultaneamente. Desta maneira, os algoritmos são a melhor opção para a obtenção dos membros finais de uma série NDVI. Antes de se aplicar o algoritmo que gera os membros finais, é necessário fazer o tratamento de ruídos, *outliers* e dados faltantes.

### 3.1.3.1 Tratamento de ruídos, *outliers* e dados faltantes

Os dados gerados por sensores remotos geralmente contêm ruídos causados por condições atmosféricas adversas, seja pela presença de nuvens, sombras de nuvens ou poeira, bem como pela visão dos sensores fora do nadir [Hird e McDermid 2009]. Se uma observação for afetada por algum tipo destes ruídos e seu valor encontrar-se além de um limite estabelecido em relação ao valor médio do grupo, esta observação é considerada um *outlier*. Tais ruídos geralmente diminuem a reflectância na banda do infravermelho próximo e podem causar distorções, por exemplo, no cálculo do NDVI, que é feito a partir dessa faixa do espectro eletromagnético.

Os ruídos de menor amplitude costumam passar por um processo de suavização enquanto que os *outliers*, que costumam aparecer em poucas quantidades, às vezes podem ser eliminados do conjunto de dados. Contudo, geralmente, à semelhança do que também acontece com os dados faltantes, eles são substituídos por uma estimativa baseada nos pontos da vizinhança, considerando que estes provavelmente possuem valores semelhantes ao ponto a ser substituído.

Apesar de o filtro Savitzky-Golay ser amplamente utilizado em artigos que foram analisados neste trabalho [Foerster et al. 2014, Leinenkugel et al. 2013, Pereira et al. 2018, Spruce et al. 2018], ele produz um sinal próximo ao original, mas com todos os valores suavizados, uma vez que ele atua em toda a extensão da série. Com o objetivo de obter um sinal próximo ao original e, ao mesmo tempo, livre de valores destoantes do comportamento esperado de uma série temporal de NDVI, apresentamos uma proposta de um outro algoritmo para tratamento de ruídos, *outliers* e dados faltantes, denominado de Filtro MaxMin. Foi feita uma comparação entre as duas opções para determinar qual delas seria a melhor em se tratando de séries temporais de NDVI.

**Filtro Savitzky-Golay** Abraham Savitzky e Marcel Golay [Savitzky e Golay 1964] propuseram, em 1964, um algoritmo para filtragem e suavização de dados com base no ajuste linear de um polinômio pelo método dos mínimos quadrados. O algoritmo parte do princípio de que os dados da série devem estar em intervalos fixos e igualmente espaçados. Esta restrição torna possível uma simplificação de um problema de mínimos quadrados, pois cada valor pode ser multiplicado por um coeficiente de convolução pré-estabelecido, que depende apenas do número de pontos do filtro.

Considerando um conjunto de pontos  $N\{y_j\}$  com  $(j = 1, \dots, n)$ , o processo de filtragem ou suavização deste conjunto consiste na substituição do ponto central  $f_i$  de

uma janela deslizante  $m$ , pelo ajuste linear de um polinômio composto pelos valores do próprio ponto  $f_i$  e dos pontos adjacentes, conforme o tamanho da janela e de acordo com a equação 3-1

$$Y_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i} \quad (3-1)$$

Onde:

$n$  é a quantidade de pontos do conjunto  $N$ ,

$j$  é limitado pelo intervalo  $\frac{m+1}{2} \leq j \leq n - \frac{m-1}{2}$

para evitar problemas nas bordas do conjunto

Por exemplo, na suavização de uma dada série de pontos  $N\{y_j\}$  com  $(j = 1, \dots, n)$ , feita pelo ajuste de um polinômio de grau 3, com  $m = 5$ ,  $i = (-2, -1, 0, 1, 2)$  e  $C = (-3, 12, 17, 12, -3)$ , o  $j$ -ésimo ponto a ser suavizado,  $Y_j$  é dado pela equação 3-2

$$Y_j = \frac{1}{35}(-3y_{j-2} + 12y_{j-1} + 17y_j + 12y_{j+1} - 3y_{j+2}) \quad (3-2)$$

Onde: 35 é o fator de normalização, igual à soma de  $C$ , que são os coeficientes constantes para polinômios cúbicos

Como o filtro Savitzky-Golay atua no ponto central de uma janela deslizante, o começo e o fim da série devem ser calculados de uma forma diferente. Uma destas formas é adicionar, em ordem inversa, os  $\frac{m-1}{2}$  primeiros pontos antes do início da série e os  $\frac{m-1}{2}$  últimos pontos depois do fim da série. Por exemplo, uma série  $\{y_1, y_2, \dots, y_n\}$  com uma janela com  $m = 5$ , seria expandida para a série  $\{y_3, y_2, y_1, y_2, \dots, y_n, y_{n-1}, y_{n-2}\}$ .

**Filtro MaxMin** O sinal de NDVI tem a particularidade de poder ser afetado pelas condições climáticas, tais como a presença de nuvens e sombras. Estas situações ocorrem de forma abrupta e isolada, levando o sinal a assumir valores próximos aos limites do seu domínio. A mudança repentina de um valor de NDVI, seguido de uma recuperação também repentina, é naturalmente improvável em um curto intervalo de tempo. Por outro lado, pequenas variações desse sinal são possíveis devido às particularidades e características da vegetação observada. A filtragem linear de sinais tem por consequência a substituição de cada ponto por uma medida estimada localmente. Esse procedimento atenua a presença de ruído, mas também pode comprometer algumas variações que são características do tipo de cobertura.

O filtro MaxMin, proposto neste estudo, estabelece um modo de selecionar quais pontos da série temporal devem ser re-estimados. Assume-se, inicialmente, que um ponto é candidato a *outlier* se ele for o valor máximo ou mínimo em uma janela de vizinhança de tamanho  $m$ . Em seguida, os pontos selecionados são submetidos a uma avaliação

estatística que classifica cada ponto como um *outlier*, substituindo-o por uma estimativa, ou não-*outlier*, preservando seu valor original. Os valores de máximo e mínimo são dados pelas equações 3-3 e 3-4:

$$M_j = \max_{\frac{1-m}{2} \leq i < \frac{m+1}{2}} \{y_{j+i}\} \quad (3-3)$$

$$m_j = \min_{\frac{1-m}{2} \leq i < \frac{m+1}{2}} \{y_{j+i}\} \quad (3-4)$$

O filtro MaxMin é definido então pela equação 3-5:

$$y'_j = \begin{cases} \hat{y}_j & , (y_j = m_j \mid y_j = M_j) \ \& \ e_j > k\sigma_j \\ y_j & , \text{caso contrário} \end{cases} \quad (3-5)$$

Onde:  $\hat{y}_j$  é a uma estimativa para  $y_j$ ,  $e_j = |y_j - \hat{y}_j|$   
e  $k\sigma_j$  é o limiar de classificação

A estimativa  $\hat{y}_j$  é calculada a partir do ajuste polinomial na mesma janela de vizinhança dada pela equação 3-6:

$$\hat{y}_j = \sum_{i=\frac{1-m}{2}}^{\frac{m+1}{2}} C_i \cdot y_{j+i} \quad (3-6)$$

Considerando que o erro produzido pelo ajuste polinomial de uma janela de vizinhança em um sinal livre de ruído possui uma distribuição normal, o ponto deve ser considerado *outlier* se seu erro de ajuste for superior a um limiar definido pela distribuição dos erros de seus vizinhos. Desse modo, o valor de limiar  $\sigma_j$  é dado pela equação 3-7:

$$\sigma_j = \frac{1}{m-1} \sum_{i=\frac{1-m}{2}}^{\frac{m+1}{2}} (e_{j+i} - \bar{e}_j)^2 \quad (3-7)$$

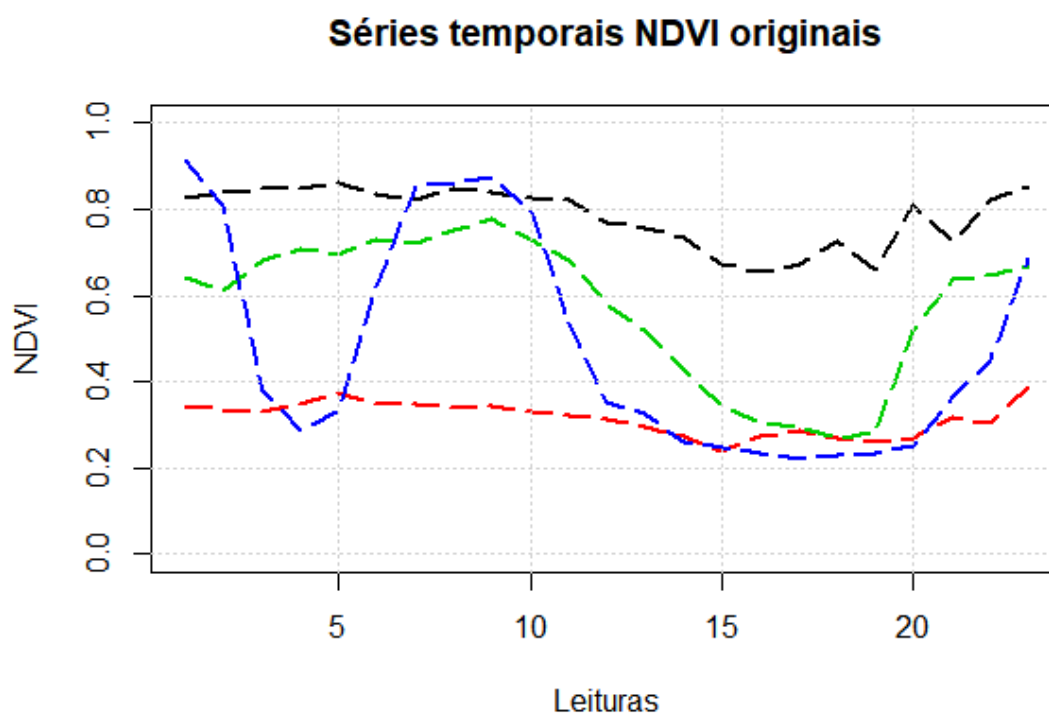
Onde:  $k$  é um fator de ponderação do desvio padrão do erro  
e  $\bar{e}_j$  a média dos erros de estimação

**Comparação entre os filtros Savitzky-Golay e MaxMin** Para verificar o desempenho dos filtros Savitzky-Golay e MaxMin, foi feito um teste comparativo entre eles a partir da seleção de quatro séries temporais de *pixels* pertencentes ao recorte SLMB, representativos das feições de área urbana, mata, pastagem e agricultura. A seleção destas séries

foi feita por análise visual de imagens de satélite<sup>1</sup> e de seus perfis NDVI ao longo do ano de 2015. O teste consistiu na inserção proposital de ruídos nos valores destas séries, de maneira que fossem alterados os seus comportamentos usuais. Em seguida, as séries alteradas pelos ruídos foram submetidas aos dois filtros para verificar qual deles produziria resultados mais próximos dos valores originais. A tabela 3.1 identifica os *pixels* selecionados e mostra quais colunas foram alteradas (das 23 existentes na série anual NDVI), bem como seus valores antes e depois de serem alterados. A figura 3.3 mostra os perfis NDVI originais dos *pixels* selecionados.

**Tabela 3.1:** Tabela de pixels nos quais foram inseridos ruídos

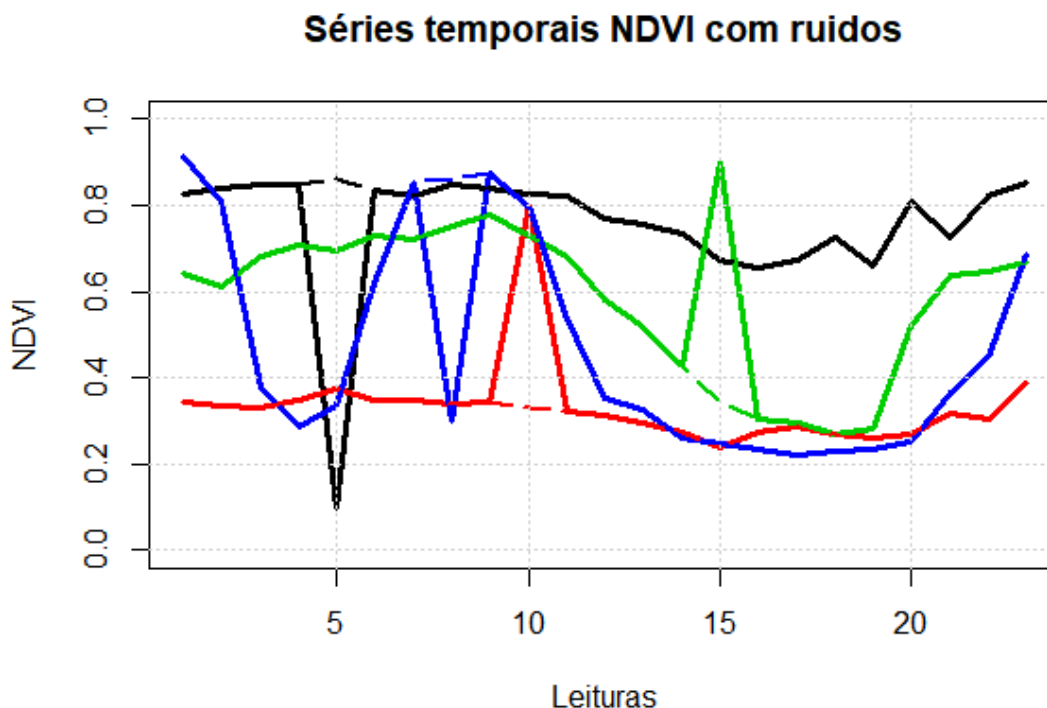
| Id. Pixel | Feição      | Latitude  | Longitude | Coluna Alterada | Vr. Original | Vr. Com Ruído |
|-----------|-------------|-----------|-----------|-----------------|--------------|---------------|
| 6879437   | Mata        | -50,34884 | -16,49357 | 5               | 0,8623       | 0,1           |
| 6918045   | Área urbana | -50,37603 | -16,51657 | 10              | 0,3325       | 0,8           |
| 6984793   | Pastagem    | -50,29447 | -16,55631 | 15              | 0,3441       | 0,9           |
| 6991827   | Agricultura | -50,26937 | -16,56049 | 8               | 0,8614       | 0,3           |



**Figura 3.3:** Perfis NDVI originais dos pixels selecionados  
Recorte SLMB – 2015

<sup>1</sup>Imagens disponibilizadas pelo Google por meio do *plugin* QuickMapServices do QGIS

A figura 3.4 mostra os perfis NDVI após terem sido alterados com a inserção de ruídos. As linhas tracejadas mostram os perfis originais das séries.

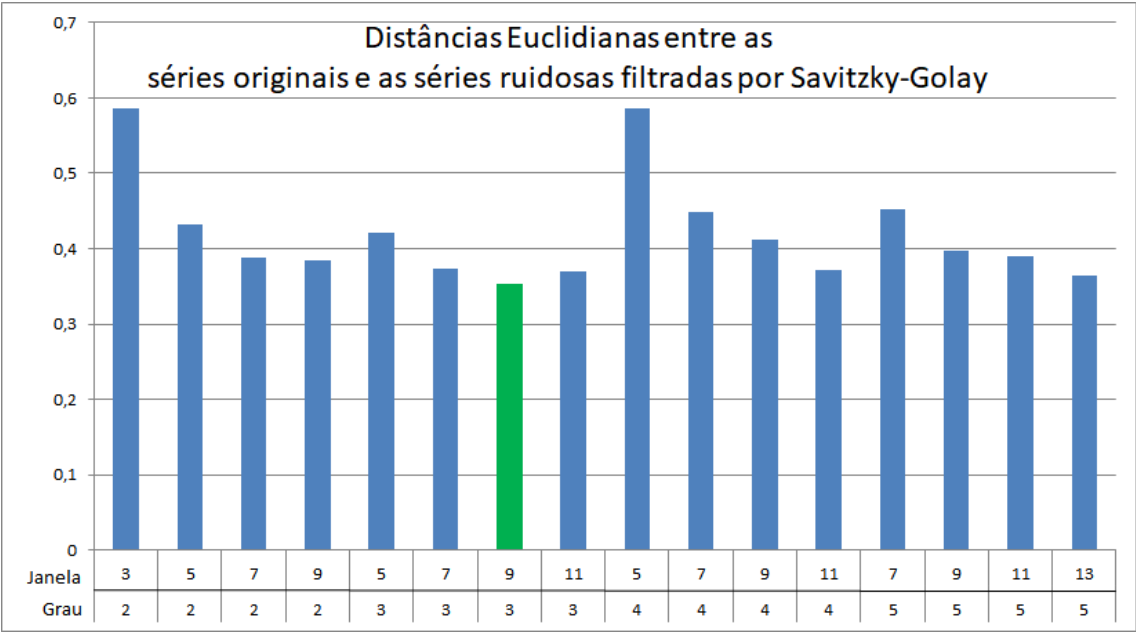


**Figura 3.4:** *Perfis NDVI com ruídos dos pixels selecionados  
Recorte SLMB – 2015*

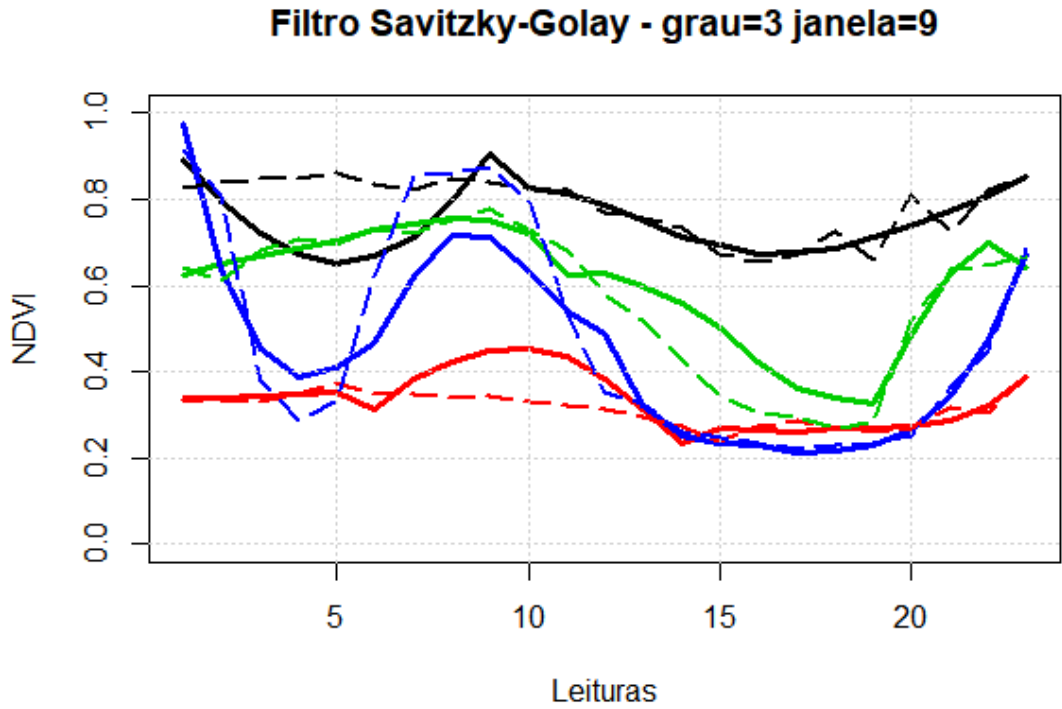
Após a inserção dos ruídos nos valores originais, as séries foram submetidas a um processo de filtragem usando os filtros Savitzky-Golay e MaxMin. Para cada um deles foram exercitadas várias combinações de seus parâmetros à procura daquela combinação que melhor aproximasse o valor filtrado do valor original, antes da inserção do ruído. Essa proximidade foi medida por meio do cálculo da distância euclidiana entre as séries originais e as com ruídos.

O gráfico de barras da figura 3.5 mostra os valores das distâncias Euclidianas entre as séries originais e as séries ruidosas após terem sido filtradas com o filtro Savitzky-Golay. No eixo x, há duas linhas representando as várias combinações dos parâmetros "grau do polinômio"(linha inferior) e "tamanho da janela"(linha superior). O menor valor de distância, igual a **0,35333**, foi obtido com o "grau de polinômio"igual a 3 e "tamanho da janela"igual a 9.

A figura 3.6 mostra o gráfico das séries ruidosas após terem sido filtradas pelo filtro Savitzky-Golay, com os parâmetros que geraram o máximo de semelhança em relação aos valores originais. O gráfico mostra, também, os valores originais em linhas tracejadas, possibilitando a comparação visual entre as séries.

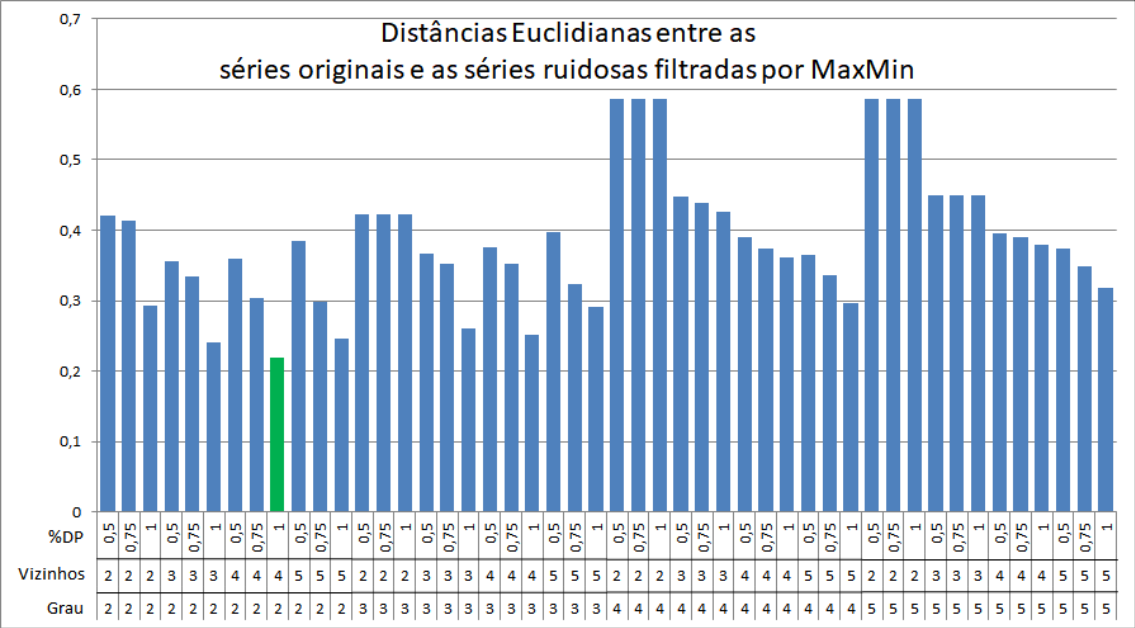


**Figura 3.5:** Distâncias Euclidianas entre as séries originais e as séries ruidosas filtradas com Savitzky-Golay



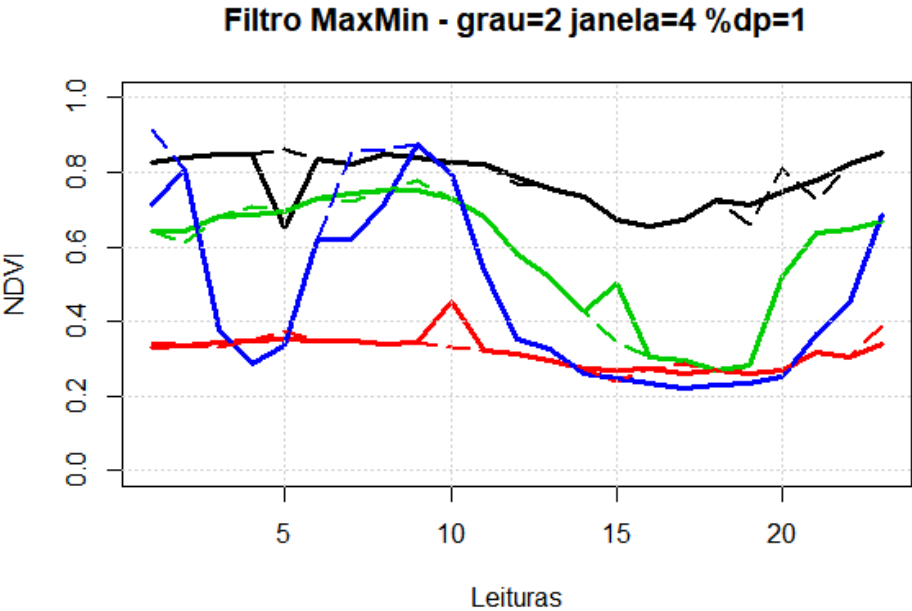
**Figura 3.6:** Séries ruidosas após serem filtradas com Savitzky-Golay

A figura 3.7 mostra o gráfico de barras com os valores das distâncias Euclidianas entre as séries originais e as séries ruidosas após terem sido filtradas com o filtro MaxMin. No eixo x, há três linhas representando as várias combinações dos parâmetros



**Figura 3.7:** Distâncias Euclidianas entre as séries originais e as séries ruidosas filtradas com MaxMin

"grau do polinômio"(linha inferior), "número de vizinhos"(linha intermediária) e "percentual do desvio padrão"(linha superior). O menor valor de distância, igual a **0,2195**, foi obtido com o "grau de polinômio"igual a 2, o "número de vizinhos"igual a 4 e o "percentual do desvio padrão"igual a 1.



**Figura 3.8:** Séries ruidosas após serem filtradas com MaxMin

A figura 3.8 mostra o gráfico das séries ruidosas após terem sido filtradas pelo

filtro MaxMin, com os parâmetros que geraram o máximo de semelhança em relação aos valores originais. O gráfico mostra, também, os valores originais em linhas tracejadas, possibilitando a comparação visual entre as séries.

Os resultados mostraram que o filtro MaxMin foi o que conseguiu uma melhor aproximação entre os valores ruidosos filtrados e os valores originais das séries de teste que foram selecionadas para a comparação.

### 3.1.3.2 Membros finais obtidos com o algoritmo PPI

Dentre os algoritmos desenvolvidos para obtenção de membros finais, [Chang et al. 2008] afirmam que o *PPI - Pixel Purity Index* é um dos mais populares e descrevem o seu funcionamento conforme algoritmo 3.1. Este algoritmo busca identificar os *pixels* presentes nos vértices de uma geometria convexa, no espaço de características, de um conjunto de dados que, segundo [Boardman J. W. e Green 1995], representam os *pixels* puros daqueles dados.

---

#### Algoritmo 3.1: Algoritmo PPI

---

**Entrada:**  $V$  – conjunto de vetores  $d$ -dimensionais relativos aos *pixels* a serem analisados;

$k$  – número de *skewers* a serem gerados;

$t$  – limiar de proximidade;

$m$  – número de membros finais a serem selecionados;

**Saída:** Um conjunto  $P$  de  $m$  *pixels* puros denominados de membros finais;

**início**

Gerar um conjunto  $S$  de  $k$  vetores aleatórios unitários com  $d$  dimensões, denominados *skewers*;

**para** cada vetor  $v$  do conjunto  $V$  **faça**

**para** cada vetor  $s$  do conjunto  $S$  **faça**

        Fazer a projeção ortogonal do vetor  $v$  no vetor  $s$  *skewers*;

**se** a distância entre projeção de  $v$  e a extremidade do vetor  $s$  estiver dentro do limiar  $t$  **então**

            Computar 1 voto para o vetor  $v$

Gerar o conjunto  $P$  com os  $m$  vetores que receberam as maiores votações

---

Após aplicarmos o algoritmo PPI, verificamos que as áreas urbanas, de mata e de agricultura foram as mais votadas na identificação dos membros finais. As áreas urbanas, por terem um perfil de NDVI estável o ano inteiro com valores baixos; as áreas de mata, por também apresentarem um perfil de NDVI estável com valores altos e as áreas de agricultura, por terem um perfil de NDVI bastante característico de valores baixos e altos relacionados ao ciclo de colheita e replantio. As áreas de pastagens não estiveram entre as



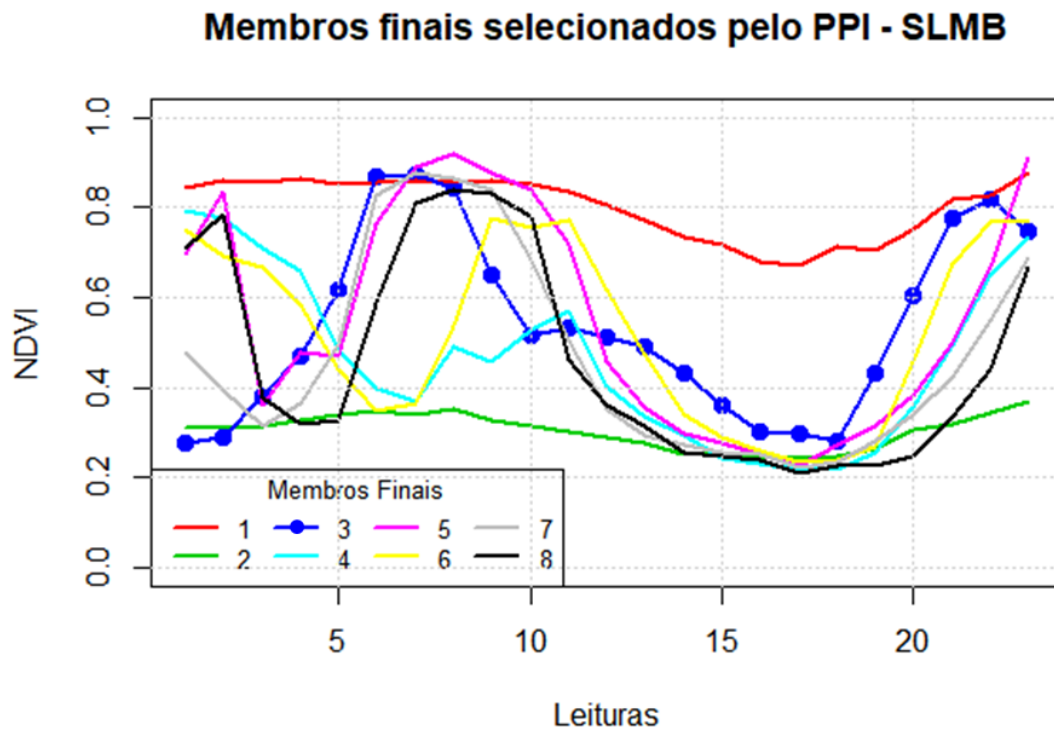
mais votadas por terem um comportamento fenológico que não atinge valores extremos de NDVI e por serem, tipicamente, uma mistura de feições. Além da pastagem propriamente dita, que é uma vegetação rasteira, há a presença de solo exposto nas trilhas feitas pelo gado, pequenas edificações, árvores isoladas etc.

Utilizando o mapa de pastagens LAPIG030, foi feita uma verificação visual para identificar os potenciais membros finais que estavam localizados em áreas de pastagens. A figura 3.9 mostra que, dos oito *pixels* que tiveram mais de 5% na etapa de votação do PPI, apenas um deles correspondeu a uma área de pastagens (pixel 3). Além disso, os *pixels* 1 (mata) e 2 (área urbana) obtiveram 47% dos votos, por estarem, respectivamente nos limites superior e inferior do perfil de NDVI.



**Figura 3.9:** *[Recorte SLMB - Localização dos membros finais gerados pelo PPI]*

O gráfico da figura 3.10 revela que, apesar de o pixel 3 estar localizado em uma área coberta pelo mapa LAPIG030, visualmente o seu perfil NDVI (linha azul destacada com pontos) não é compatível com o comportamento fenológico usual de uma área de pastagens.



**Figura 3.10:** *[Recorte SLMB - perfil NDVI dos membros finais gerados pelo PPI]*

Em função dos resultados insatisfatórios obtidos com o algoritmo PPI, foi adotada uma nova estratégia em nossa metodologia, que é a obtenção dos membros finais a partir dos centros dos *clusters* gerados por um algoritmo de clusterização.

### 3.1.3.3 Membros finais obtidos por clusterização

A outra opção de algoritmo para a obtenção dos membros finais foi a classificação não supervisionada, ou clusterização. Neste tipo de classificação, todos os *pixels*, inclusive aqueles que não possuem características extremas como é o caso das áreas de pastagens, são agrupados com outros elementos semelhantes. Cada um dos grupos tem um representante que reflete a média das características dos demais integrantes de seu grupo.

O processo adotado nesta metodologia consiste na geração de uma matriz de semelhanças/dessemelhanças dos dados que já foram previamente filtrados e esta matriz serve como argumento para o algoritmo de clusterização. Daí, os centros dos *clusters* gerados são escolhidos como membros finais da área de referência.

**3.1.3.3.1 Medidas de Proximidade** As medidas de semelhança e dessemelhança, conforme afirmam [Aghabozorgi, Shirkhorshidi e Teh 2015], são fundamentais no pro-

cesso de clusterização de séries temporais, pois determinam o quão semelhantes elas são. As medidas de distância mais populares são a Distância Euclidiana e a *Dynamic Time Warping* (DTW), sendo que a primeira é mais adequada para séries que possuem igual número de ocorrências e cujas semelhanças ocorrem em cada etapa do tempo; a DTW é mais adequada para as séries com semelhança na forma, sendo que estas não precisam, necessariamente, ter o mesmo tamanho. As duas medidas são utilizadas para encontrar semelhanças em séries temporais baseadas na forma, sendo que a semelhança no tempo é considerada um caso especial de semelhança na forma.

**Distância Euclidiana** A Distância Euclidiana é uma medida simples, mas sua acurácia a torna competitiva em um grande número de situações [Aghabozorgi, Shirkhorshidi e Teh 2015]. Contudo, havendo distorções entre os dados de duas séries, ou a distorção deve ser removida antes de se aplicar a Distância Euclidiana ou deve-se utilizar uma medida mais adequada [Batista et al. 2013].

A Distância Euclidiana entre duas séries temporais  $A$  e  $B$ , de tamanho  $n$ , tal que  $A = \{a_1, a_2, \dots, a_n\}$  e  $B = \{b_1, b_2, \dots, b_n\}$ , é calculada de acordo com a equação 3-8:

$$ED(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (3-8)$$

**Dynamic Time Warping - DTW** O algoritmo DTW se destaca por sua eficiência em minimizar o deslocamento e a distorção no tempo, esticando ou compactando localmente duas séries temporais, de maneira que seja possível detectar formas semelhantes em diferentes fases [Senin 2009]. As séries não precisam ter o mesmo tamanho mas é requerido que os valores obedeçam ao mesmo intervalo de tempo. O DTW também é conhecido por seu alto consumo computacional, fazendo com que, ao longo do tempo, ele fosse sendo modificado com o objetivo de melhorar sua performance. A versão aqui utilizada emprega a programação dinâmica, com tempo de execução  $O(N \times M)$ .

Dadas duas séries temporais  $A$  e  $B$ , tal que  $A = \{a_1, a_2, \dots, a_N\}, N \in \mathbb{N}$  e  $B = \{b_1, b_2, \dots, b_M\}, M \in \mathbb{N}$ , deve-se calcular uma matriz de custos acumulados  $N \times M$ , segundo as equações 3-9 (primeira linha), 3-10 (primeira coluna) e 3-11 (para os demais elementos da matriz):

$$D_{(1,j)} = \sum_{k=1}^j C(a_1, b_k), j \in [1, M] \quad (3-9)$$

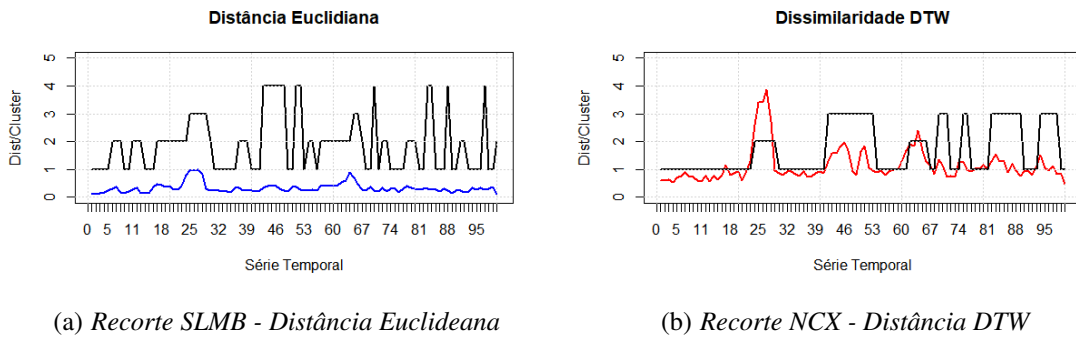
$$D_{(i,1)} = \sum_{k=1}^i C(a_k, b_1), i \in [1, N] \quad (3-10)$$

$$D_{(i,j)} = \min\{D_{(i-1,j-1)}, D_{(i-1,j)}, D_{(i,j-1)}\} + C(a_i, b_j), i \in [1, N], j \in [1, M] \quad (3-11)$$

Uma vez que a matriz de custos acumulados esteja calculada, o passo seguinte é determinar o "caminho distorcido" (*warping path*) que vai do ponto  $P_{(N,M)}$  até o ponto  $P_{(1,1)}$ , sendo que o próximo ponto  $P$  é determinado conforme equação 3-12.

$$P = \min\{P_{(i-1,j-1)}, P_{(i-1,j)}, P_{(i,j-1)}\}, i \in [1, N], j \in [1, M] \quad (3-12)$$

**3.1.3.3.2 Comparação entre Distâncias Euclidiana e DTW** A aplicação das distâncias Euclidiana ou DTW geram uma matriz de semelhanças contendo a distância de cada um dos *pixels* da imagem para os demais *pixels*. A linha azul na figura 3.11(a) e a linha vermelha na figura 3.11(b) mostram, graficamente, a distância entre o primeiro pixel da imagem e os demais 99 *pixels*, correspondendo à primeira linha da imagem e aos 100 primeiros elementos da matriz de semelhanças. A intensidade com que as linhas azul e vermelha variam, indica o quão distante o primeiro pixel está dos *pixels* de 2 a 100. É esperado que os *pixels* que estejam à mesma distância do pixel 1 sejam classificados no mesmo cluster.



**Figura 3.11:** Comparação entre Distâncias Euclidiana e DTW

As linhas pretas, em ambas as sub-figuras da figura 3.11, representam os *clusters* nos quais cada um dos 100 primeiros pixels da imagem foram classificados. É possível observar que, no caso da clusterização feita a partir da matriz de distâncias DTW, as classes são mais conectadas espacialmente, mesmo que o seu gráfico de distâncias apresente mais variações do que no caso da clusterização feita a partir da matriz de distâncias Euclidean. Corroborando com esta constatação, a tabela 3.2 mostra duas

matrizes de confusão<sup>2</sup> nas quais é possível constatar que os resultados de classificação obtidos a partir da distância DTW apresentam menor erro total do que aqueles obtidos a partir da distância Euclidiana.

**Tabela 3.2:** Matrizes de confusão das clusterizações com distâncias Euclidiana e DTW

| Recorte:        |   | SLMB       |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | PAM-EUC    |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 4.564      | 1.049  | 5.613        | 18,69%           |
|                 | N | 98         | 489    | 587          | 16,70%           |
| Total Colunas   |   | 4.662      | 1.538  | 6.200        | 18,50%           |
| Erro de Omissão |   | 2,10%      | 68,21% | Erro Total   |                  |

| Recorte:        |   | SLMB       |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | PAM-DTW    |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 4.380      | 613    | 4.993        | 12,28%           |
|                 | N | 282        | 925    | 1.207        | 23,36%           |
| Total Colunas   |   | 4.662      | 1.538  | 6.200        | 14,44%           |
| Erro de Omissão |   | 6,05%      | 39,86% | Erro Total   |                  |

**3.1.3.3.3 Clusterização** De acordo com [Aghabozorgi, Shirkhorshidi e Teh 2015], o processo de clusterização pode ser classificado em seis grupos: particionamento, hierárquico, baseado em grade, baseado em modelo, baseado em densidade de clustering e algoritmos de clusterização multi-etapas. Os algoritmos que pertencem ao grupo particionamento dividem os dados em várias partições, nas quais os componentes são agrupadas pela semelhança de suas características. Dentre os algoritmos por particionamento estão o *k-means* e o *k-medoids*, que foram analisados e comparados neste trabalho. O prefixo *k* presente no nome destes dois algoritmos significa o número de *clusters* que eles devem produzir, sendo que esta quantidade deve ser informada quando da execução de um destes algoritmos.

**Algoritmo de clusterização *k-means*** Este algoritmo de clusterização por particionamento gera *k clusters* de *n* objetos não rotulados, sendo que cada cluster tem um protótipo, que é o valor médio de seus objetos. O *k-means* é sensível à presença de *outliers* que porventura estejam presentes nos dados a serem clusterizados. Contudo, ele é um dos algoritmos mais populares e foi a opção adotada em um bom número de trabalhos consultados, quais sejam: [Gonçalves et al. 2014] na clusterização tridimensional de dados de NDVI, coeficiente de reflexão (albedo) e temperatura da superfície terrestre para mapear a variabilidade do uso do solo no estado de São Paulo; [Mercado et al. 2016] em seu estudo, a partir de séries temporais de imagens de satélite, sobre a eutrofização da bacia do norte do Mar de Alborão, na parte ocidental do Mar Mediterrâneo; [Gonçalves et al. 2017], para fazer a clusterização de séries temporais de NDVI para monitoramento da cultura de cana

<sup>2</sup>Veja detalhes sobre os dados de referência para validação em 3.2.2 e sobre o conteúdo da matriz de confusão em 3.2.3

de açúcar no estado de São Paulo; [Abdelli et al. 2013], para explicar e prever o comportamento do mercado financeiro marroquino; [Pascucci et al. 2018], na clusterização de uma área de agricultura próxima à cidade de Roma, Itália e [Guyet e Nicolas 2016], em seu estudo sobre a resiliência da vegetação ao longo dos anos no Senegal.

**Algoritmo de clusterização *k-medoids*** Este é um outro membro da família de algoritmos por particionamento, no qual o protótipo de cada cluster, denominado de *medoid*, é o objeto do grupo mais próximo do centro do cluster [Gandhi e Srivastava 2014]. Ele é mais robusto quanto à presença de ruídos e *outliers* que o *k-means*, pois minimiza uma soma de dissimilaridades entre pares de pontos. Por sua robustez, [Huy e Anh 2016] optaram pela versão denominada de particionamento em torno de medóides (*Partitioning Around Medoids – PAM*) em sua proposta de um método de clusterização de séries temporais. O PAM também foi o escolhido por [Hitchcock, Booth e Casella 2007] como algoritmo de clusterização no estudo que conduziram sobre os efeitos que a pré-suavização de dados exerce nos resultados de uma clusterização.

A versão do *k-medoids* denominada de *Partitioning Around Medoids – PAM*, foi implementada por Leonard Kaufman e Peter J. Rousseeuw [Kaufmann e Rousseeuw 1987], e sua formulação está representada pela equação 3-13.

$$\begin{aligned}
 F(x) &= \text{minimize} \sum_{i=1}^n \sum_{j=1}^n d(i, j) z_{ij} \\
 &\text{sujeito a} \\
 &\sum_{i=1}^n z_{ij} = 1, \quad j = 1, \dots, n \\
 &z_{ij} \leq y_i, \quad i, j = 1, \dots, n \\
 &\sum_{i=1}^n y_i = k, \quad k = \text{numero de clusters} \\
 &y_i, z_{ij} \in \{0, 1\}, \quad i, j = 1, \dots, n
 \end{aligned} \tag{3-13}$$

$F(x)$  é a função a ser minimizada;  $d(i, j)$  é a medida de dessemelhança entre os pontos  $i$  e  $j$ ;  $z_{ij}$  é uma variável que garante que apenas a dissimilaridade entre pontos do mesmo cluster será computada.

O algoritmo PAM recebe como parâmetro de entrada a matriz de pontos a serem clusterizados e o número  $k$  de *clusters* a serem formados. Opcionalmente, o PAM também pode receber a matriz de dessemelhança já calculada. O algoritmo 3.2 desenvolve-se conforme descrito a seguir:

**Algoritmo 3.2:** Algoritmo PAM**Entrada:**  $k$  – número de *clusters*; $D$  – conjunto de dados com  $n$  objetos;**Saída:** Um conjunto de  $k$  *clusters*;**início**Escolha arbitrariamente  $k$  objetos em  $D$  como os representantes iniciais;**repita**Atribua cada objeto restante ao cluster com o representativo mais próximo ( $R_j$ );Aleatoriamente selecione um objeto que não seja representante ( $R_r$ );Calcule o custo total da mudança de  $R_j$  para  $R_r$ ;**se** o custo da mudança é  $\leq 0$  **então**    Substitua o  $R_j$  por  $R_r$  para formar o novo conjunto de  $k$  representantes**até** que não haja mudanças;

**Comparação entre os algoritmos de clusterização *k-means* e *k-medoids*** Antes de efetuar a comparação dos resultados das clusterizações feitas por estes dois algoritmos, foi aplicado o filtro de suavização dos dados MaxMin. A tabela 3.3 mostra as matrizes de confusão<sup>3</sup> resultantes dos dois algoritmos de clusterização, tendo o *k-medoids* (PAM) apresentado o menor erro total entre eles. Além disso, os medóides, que são os centros dos *clusters* da classificação *k-medoids*, são pontos que já pertencem ao conjunto de dados – ao contrário do que acontece na classificação *k-means* – e podem ser utilizados diretamente como membros finais para o cálculo das frações de abundância.

**Tabela 3.3:** Matrizes de confusão das clusterizações com os algoritmos *k-means* e *k-medoids*

| Recorte:        |   | SLMB       |        |            |          |
|-----------------|---|------------|--------|------------|----------|
| Classificação:  |   | K-MEANS    |        |            |          |
| Pastagens       |   | Referência |        | Total      | Erro de  |
|                 |   | S          | N      | Linhas     | Comissão |
| Previsto        | S | 4.432      | 761    | 5.193      | 14,65%   |
|                 | N | 230        | 777    | 1.007      | 22,84%   |
| Total Colunas   |   | 4.662      | 1.538  | 6.200      | 15,98%   |
| Erro de Omissão |   | 4,93%      | 49,48% | Erro Total |          |

| Recorte:        |   | SLMB       |        |            |          |
|-----------------|---|------------|--------|------------|----------|
| Classificação:  |   | K-MEDOIDS  |        |            |          |
| Pastagens       |   | Referência |        | Total      | Erro de  |
|                 |   | S          | N      | Linhas     | Comissão |
| Previsto        | S | 4.380      | 613    | 4.993      | 12,28%   |
|                 | N | 282        | 925    | 1.207      | 23,36%   |
| Total Colunas   |   | 4.662      | 1.538  | 6.200      | 14,44%   |
| Erro de Omissão |   | 6,05%      | 39,86% | Erro Total |          |

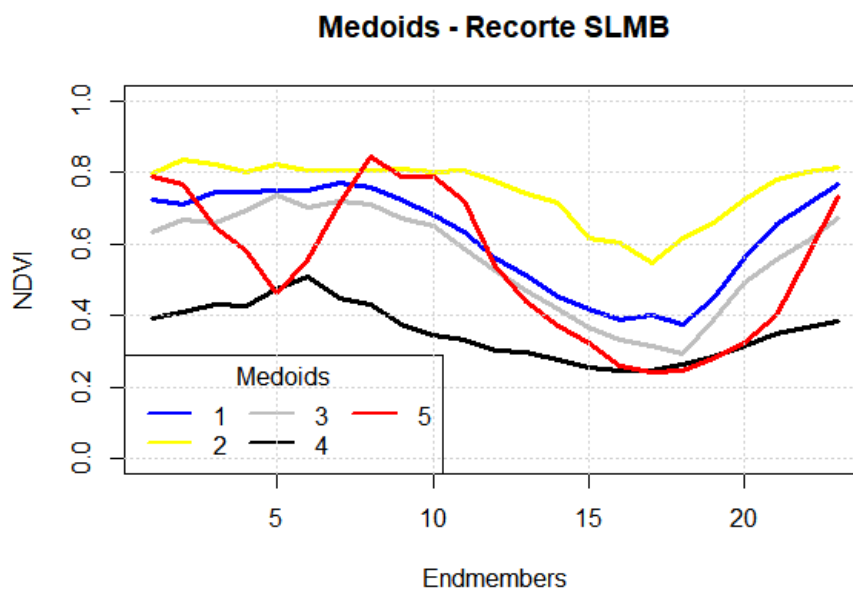
A figura 3.12 mostra que, na clusterização feita pelo PAM, dois dentre os cinco membros finais (1 e 3) estão localizados em áreas de pastagens.

<sup>3</sup>Veja detalhes sobre os dados de referência para validação em 3.2.2 e sobre o conteúdo da matriz de confusão em 3.2.3



**Figura 3.12:** Recorte SLMB - Localização dos membros finais gerados pelo PAM

A figura 3.13 mostra os perfis NDVI dos membros finais confirmando os comportamentos fenológicos esperados de cada um dos tipos de feições.

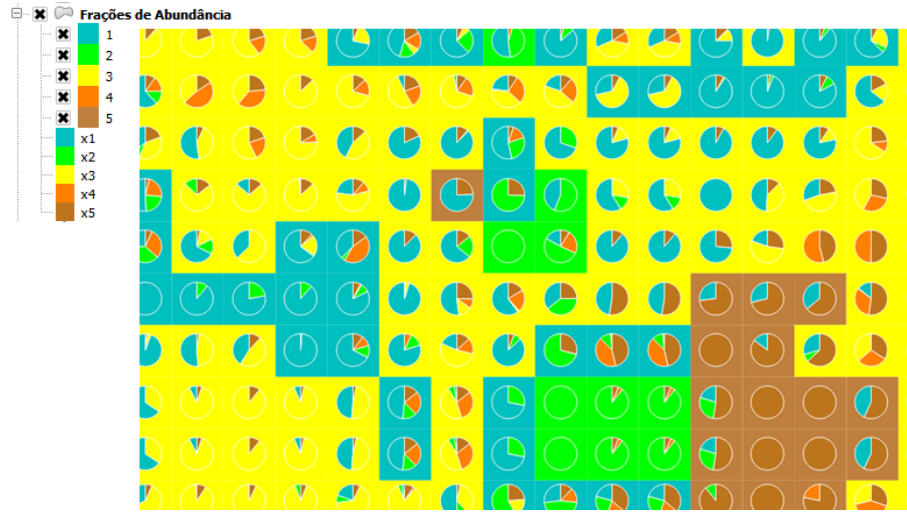


**Figura 3.13:** Recorte SLMB - perfil NDVI dos membros finais gerados pelo PAM



### 3.1.4 Cálculo das Frações de Abundância

Fração de abundância é a proporção com que cada pixel puro (membro final) de um conjunto de dados contribui na composição dos demais *pixels*, ditos misturados. Nesta metodologia os *pixels* puros correspondem aos medóides gerados no processo de clusterização pelo algoritmo PAM. A figura 3.14 mostra gráficos no estilo "torta" que representam a decomposição dos valores de *pixels* misturados em suas frações de abundância.



**Figura 3.14:** Exemplo das frações de abundância de um conjunto de pixels - Recorte SLMB

A análise linear de mistura temporal é uma técnica muito utilizada para estimar as frações de abundância dos materiais presentes nos *pixels* de uma imagem [Heinz e Chein-I-Chang 2001]. Segundo [Keshava 2003], uma das alternativas para resolver modelos de mistura linear são os algoritmos baseados em mínimos quadrados. A equação 3-14 mostra a solução para os casos mais triviais, onde não há restrições (*Unconstrained Least Squares* - UCS).

$$\hat{x}^U = (S^T S)^{-1} S^T p \quad (3-14)$$

Onde:

$x$  é a abundância de um membro final

$S$  é a matriz de membros finais

$p$  é o valor de um pixel

Para que a estimativa das frações de abundância seja consistente há duas restrições que devem ser seguidas. A primeira restrição (equação 3-15) requer que a soma das frações de abundância seja igual a um (*Abundance Sum-to-one Constraint* - ASC) e a

equação 3-16 mostra a sua solução.

$$\sum_{j=1}^M x_j = 1 \quad (3-15)$$

$$\hat{x}^F = \hat{x}^U - (S^T S)^{-1} Z^T [Z (S^T S)^{-1} Z^T]^{-1} (Z \hat{x}^U - b) \quad (3-16)$$

Onde:

$x$  é a abundância de um membro final

$S$  é a matriz de membros finais

$p$  é o valor de um pixel

$Z$  é um vetor com valores iguais a 1, de tamanho  $M$

$M$  é a quantidade de membros finais

$b$  é igual a 1

A segunda restrição (equação 3-17) requer que as frações de abundância tenham valores maiores ou iguais a zero (*Abundance Non-negativity Constraint - ANC*). Esta restrição não tem uma forma fechada de solução e enquadra-se no domínio da programação quadrática com inequações lineares como restrições [Keshava 2003].

Neste trabalho adotamos a implementação feita na linguagem R por [Mullen e Stokkum 2012] para o algoritmo *NNLS – Non-Negative Least Squares* proposto por [Lawson e Hanson 1995]. Como este algoritmo obedece apenas à restrição de não negatividade, depois de sua aplicação foi feita a normalização dos valores para que a soma destes fosse igual a um.

$$x_j \geq 0, \quad \text{para todo} \quad 1 \leq j \leq M \quad (3-17)$$

Onde:

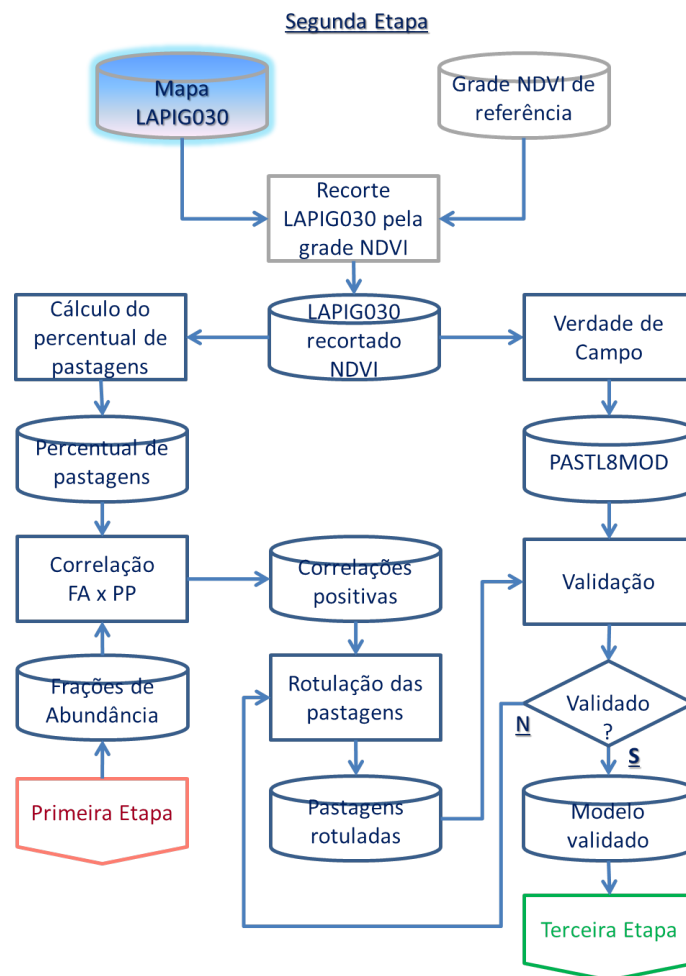
$x$  é a abundância de um membro final

$M$  é a quantidade de membros finais

O resultado final do processo de cálculo das frações de abundância é uma matriz com o número de linhas igual à quantidade de *pixels* do recorte de referência e com o número de colunas igual ao número de membros finais.

## 3.2 Modelo de classificação supervisionada

Esta é a segunda etapa da metodologia na qual é construído o modelo de classificação supervisionada. Os objetivos são a definição dos critérios para identificar e rotular os *pixels* que fazem parte da classe de pastagens e a validação do modelo de classificação, conforme detalhado no diagrama da figura 3.15.



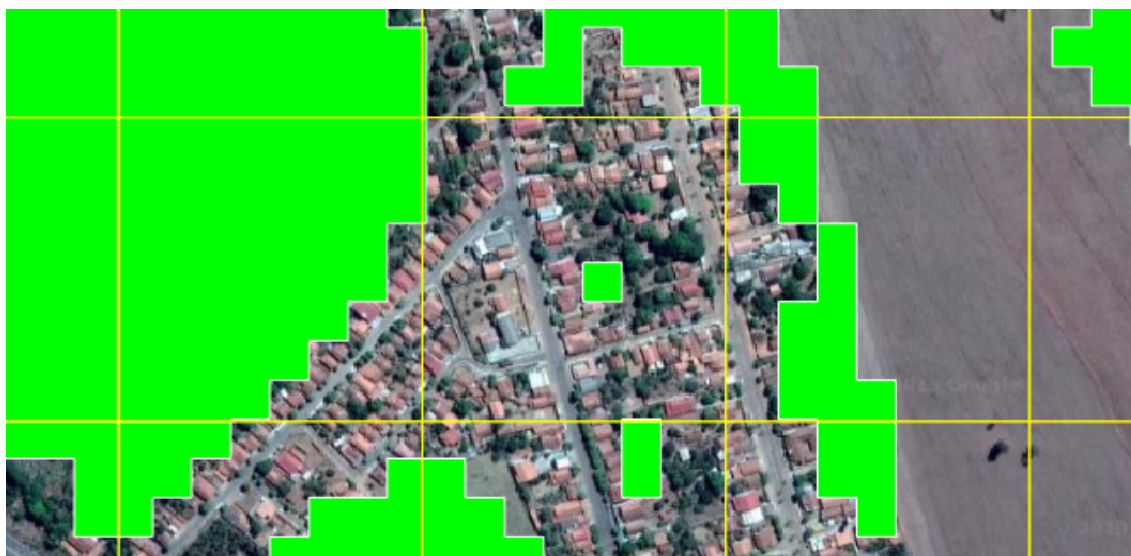
**Figura 3.15:** Segunda etapa da metodologia

### 3.2.1 Definição de parâmetros para rotulação das pastagens

Nesta etapa são definidos os percentuais de pastagem para cada pixel do recorte de referência e, em seguida, é feita uma análise da correlação entre o percentual de pastagem dos *pixels* e os seus respectivos membros finais. Por fim, são definidos os coeficientes a serem aplicados às frações de abundância para os *pixels* que apresentarem correlação positiva com o percentual de pastagens.

### 3.2.1.1 Percentuais de pastagens dos *pixels* NDVI

Para se calcular o percentual de pastagens em um pixel do recorte de referência é necessário determinar a área de pastagens que porventura esteja nele contida. É possível extrair esta informação do mapa de pastagens LAPIG030, por meio de uma sobreposição entre este mapa e a grade das imagens NDVI. Contudo, o mapa LAPIG030 tem resolução espacial de 30 m enquanto que a da grade NDVI é de 231,656 m, ou seja, quase 60 vezes maior. Como a maior resolução não é múltipla da menor, há um desencontro dos *pixels* dos dois produtos. A figura 3.16 mostra a grade NDVI na cor amarela e as áreas de pastagens do LAPIG030 na cor verde.



**Figura 3.16:** Área abrangida pelos pixels do mapa de pastagens e da série NDVI

A solução para a associação entre os dois dados é fazer uma operação espacial na qual a grade NDVI "recorta" o mapa LAPIG030, resultando em tantos polígonos quantos são aqueles que formam a grade NDVI, cada um deles contendo, ou não, áreas de pastagens em seu interior. Em seguida, encontra-se o percentual de pastagens dividindo-se a área de pastagens pela área total do pixel, com percentuais variando de 0% a 100%. Por último, são acrescidas a cada pixel as respectivas frações de abundância, calculadas na primeira etapa. Para facilitar sua identificação, este mapa será denominado neste documento de PASTL8MOD, em referência à composição do mapa de pastagens originado do Landsat 8 (LAPIG030) com a grade NDVI, originada do sensor MODIS.

### 3.2.1.2 Correlação entre percentuais de pastagens e frações de abundância

O próximo passo é o cálculo da correlação entre o percentual de pastagens e os valores de cada uma das frações de abundância de cada pixel para identificar as correlações positivas, ou seja, quais são os membros finais cujas frações de abundância

aumentam seus valores ao mesmo tempo em que os percentuais de pastagens também aumentam. Ocorrendo este fato para um ou mais membros finais, há o indicativo de que estes podem ser utilizados no critério de seleção e classificação dos *pixels*.

Dado um recorte de referência com  $n$  *pixels*, seja  $PP = \{pp_1, \dots, pp_n\}$  o conjunto dos percentuais de pastagens destes *pixels*; seja  $X_i = \{x_{i,1}, \dots, x_{i,m}\}$  (sendo  $1 \leq i \leq n$ ) o conjunto das frações de abundância do  $i$ -ésimo pixel; seja  $X_j = \{x_{1,j}, \dots, x_{n,j}\}$  (sendo  $1 \leq j \leq m$ ) o conjunto das  $j$ -ésimas frações de abundância de cada pixel, o cálculo da correlação entre a fração de abundância  $X_j$  e o percentual de pastagens  $PP$  é dado pela equação 3-18.

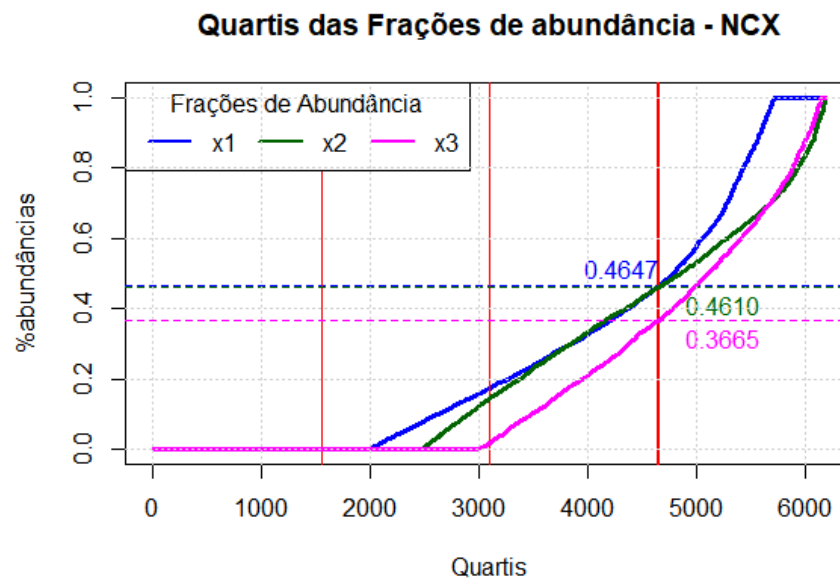
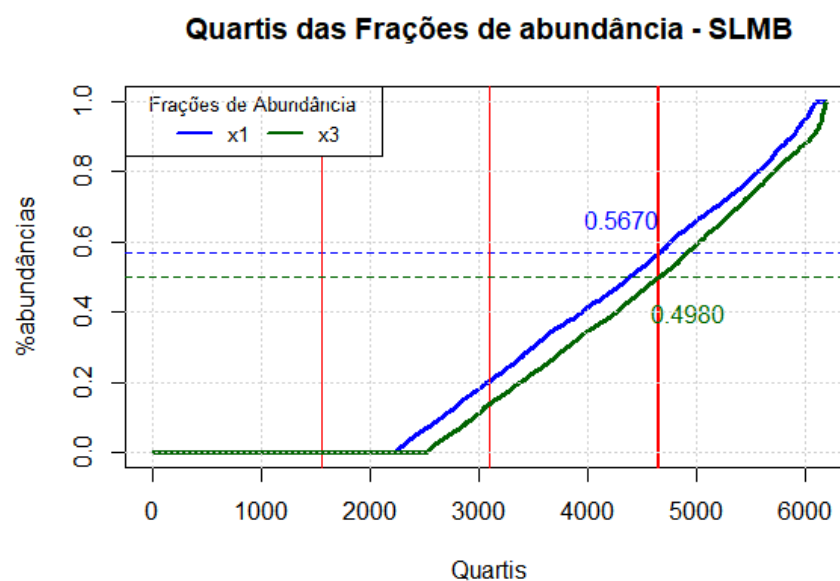
$$\rho(X_j, PP) = \text{corr}(X_j, PP) = \frac{\text{cov}(X_j, PP)}{\sigma_{X_j} \cdot \sigma_{PP}} \quad (3-18)$$

### 3.2.1.3 Definição do coeficiente de multiplicação para seleção das frações de abundância

As frações de abundância que têm correlação positiva com o percentual de pastagens são aquelas utilizadas para seleção dos *pixels* com pastagens. Uma fração de abundância corresponde ao percentual de presença de um determinado tipo de feição no pixel e tem valores no intervalo [0,1]. Devido ao fato de que um pixel de uma área de pastagens pode abranger diferentes feições, é necessário estabelecer um valor limite a partir do qual aquela fração representa tal feição.

Nesta metodologia o processo de determinação do limite para uma determinada fração de abundância foi feito analisando-se inicialmente os seus quartis em ambos os recortes de referência (NCX e SLMB). A figura 3.17 mostra que, até por volta do segundo quartil, os valores são iguais ou próximos a zero e depois há um crescimento expressivo no intervalo entre o segundo e o terceiro quartil. Isto permitiu pressupor que o valor limite situa-se entre estes dois quartis.

Os valores de ambos os quartis foram testados como limiar de seleção de *pixels* de pastagens e o valor do terceiro quartil apresentou melhores resultados, mas ainda bem acima do esperado. Para encontrar o valor ideal, os valores dos terceiros quartis das  $m$  frações de abundância relativas às áreas de pastagens ( $qx_1, \dots, qx_m$ ) foram usados como limiares iniciais e foi estabelecido um coeficiente multiplicativo  $cm$ , com valor inicial de 1,05, para ser aplicado às frações de abundância. Num processo de repetição, o coeficiente  $cm$  foi sendo decrementado em 0,05, multiplicado pelos valores iniciais das frações de abundância, feita a seleção dos *pixels*, efetuados os cálculos dos indicadores a serem analisados, até que a condição de parada fosse alcançada, conforme descrito no algoritmo 3.3. As variáveis citadas no algoritmo estão descritas no item 3.2.3.

(a) *Quartis das frações de abundância - recorte NCX*(b) *Quartis das frações de abundância - recorte SLMB***Figura 3.17:** *Quartis das frações de abundância*

**Algoritmo 3.3:** Definição do coeficiente de multiplicação  $cm$ **Entrada:** 3º quartil das  $m$  frações de abundância selecionadas  $\{qx_1, \dots, qx_m\}$ **Saída:** Coeficiente de multiplicação das frações de abundância – ( $cm$ )**início**

Inicialização;

 $cm = 1,05$  (inicia coeficiente com valor 1,05); $eq = N$  (variável para controlar ponto de equilíbrio); $s = NULL$  (sentença de seleção); $sPPPN = 0$  (sinal de  $PP - PN$ ); $sCSOS = 0$  (sinal de  $CS - OS$ ); $sCNON = 1$  (sinal de  $CN - ON$ );**repita** $cm = cm - 0,05$ ;**para**  $i = 1$  até  $i = m$  **faça****se**  $i = 1$  **então**|  $s = "x_i \geq cm \cdot qx_i"$ **senão**|  $s = s + " \text{ OR } x_i \geq cm \cdot qx_i"$ Seleciona pixels de acordo com  $s$ ;Calula  $PP, PN, CS, OS, CN$  e  $ON$ ;Identifica o sinal de  $sPPPN, sCSOS$  e  $sCNON$ ;**se um dos sinais mudou então**|  $eq = S$ **até** até  $eq = S$ ;

Os resultados da aplicação do algoritmo para o recorte NCX estão demonstrados na tabela 3.4 e no gráfico da figura 3.18.

**Tabela 3.4:** Tabela de seleção do coeficiente  $cm$  – recorte NCX

|                      |      |      |      |     |         |     |         |         |        |        |        |        |        |         |        |         |
|----------------------|------|------|------|-----|---------|-----|---------|---------|--------|--------|--------|--------|--------|---------|--------|---------|
| PX - Total de pixels |      |      |      |     | 6200    |     |         |         |        |        |        |        |        |         |        |         |
| RP - Pastagens       |      |      |      |     | 4970    |     |         |         |        |        |        |        |        |         |        |         |
| RN - Não pastagens   |      |      |      |     | 1230    |     |         |         |        |        |        |        |        |         |        |         |
| Fração abundância x1 |      |      |      |     | 0,46471 |     |         |         |        |        |        |        |        |         |        |         |
| Fração abundância x2 |      |      |      |     | 0,46101 |     |         |         |        |        |        |        |        |         |        |         |
| Fração abundância x3 |      |      |      |     | 0,36650 |     |         |         |        |        |        |        |        |         |        |         |
| cm                   | PC   | NC   | SS   | SN  | NN      | NS  | PP      | PN      | CS     | OS     | CN     | ON     | ET     | PP-PN   | CS-OS  | CN-ON   |
| 1,00                 | 4454 | 1746 | 4124 | 330 | 900     | 846 | 89,62%  | 141,95% | 7,41%  | 17,02% | 48,45% | 26,83% | 18,97% | -52,33% | -9,61% | 21,62%  |
| 0,95                 | 4620 | 1580 | 4256 | 364 | 866     | 714 | 92,96%  | 128,46% | 7,88%  | 14,37% | 45,19% | 29,59% | 17,39% | -35,50% | -6,49% | 15,60%  |
| 0,90                 | 4790 | 1410 | 4390 | 400 | 830     | 580 | 96,38%  | 114,63% | 8,35%  | 11,67% | 41,13% | 32,52% | 15,81% | -18,26% | -3,32% | 8,61%   |
| 0,85                 | 4909 | 1291 | 4486 | 423 | 807     | 484 | 98,77%  | 104,96% | 8,62%  | 9,74%  | 37,49% | 34,39% | 14,63% | -6,19%  | -1,12% | 3,10%   |
| 0,80                 | 5020 | 1180 | 4569 | 451 | 779     | 401 | 101,01% | 95,93%  | 8,98%  | 8,07%  | 33,98% | 36,67% | 13,74% | 5,07%   | 0,92%  | -2,68%  |
| 0,75                 | 5102 | 1098 | 4622 | 480 | 750     | 348 | 102,66% | 89,27%  | 9,41%  | 7,00%  | 31,69% | 39,02% | 13,35% | 13,39%  | 2,41%  | -7,33%  |
| 0,70                 | 5185 | 1015 | 4681 | 504 | 726     | 289 | 104,33% | 82,52%  | 9,72%  | 5,81%  | 28,47% | 40,98% | 12,79% | 21,81%  | 3,91%  | -12,50% |
| 0,65                 | 5259 | 941  | 4735 | 524 | 706     | 235 | 105,81% | 76,50%  | 9,96%  | 4,73%  | 24,97% | 42,60% | 12,24% | 29,31%  | 5,24%  | -17,63% |
| 0,60                 | 5339 | 861  | 4784 | 555 | 675     | 186 | 107,42% | 70,00%  | 10,40% | 3,74%  | 21,60% | 45,12% | 11,95% | 37,42%  | 6,65%  | -23,52% |

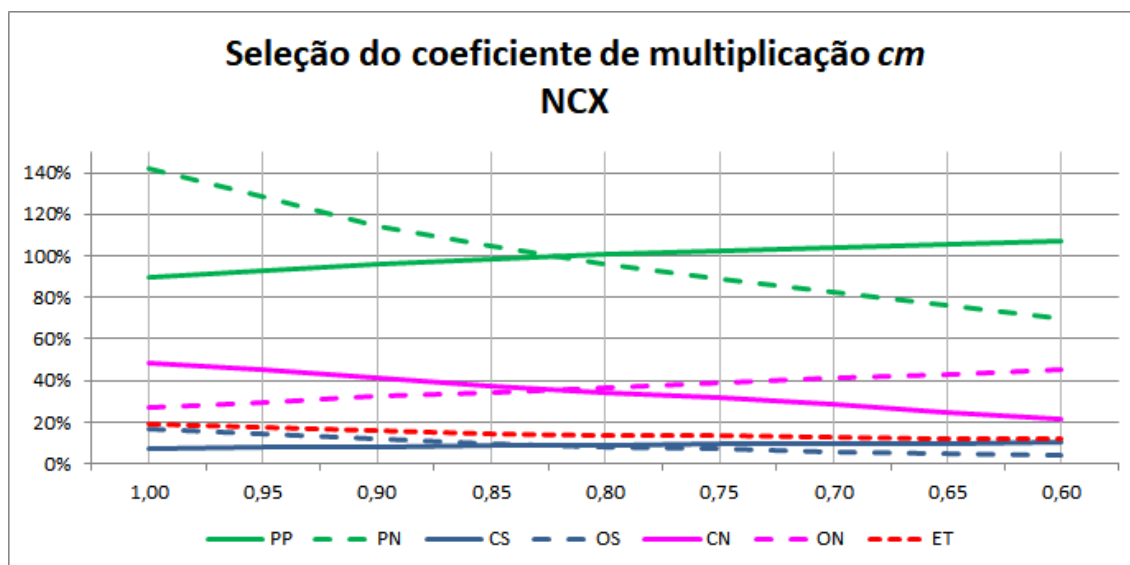


Figura 3.18: Gráfico da seleção do coeficiente cm - recorte NCX

### 3.2.2 Preparação do mapa verdade de campo

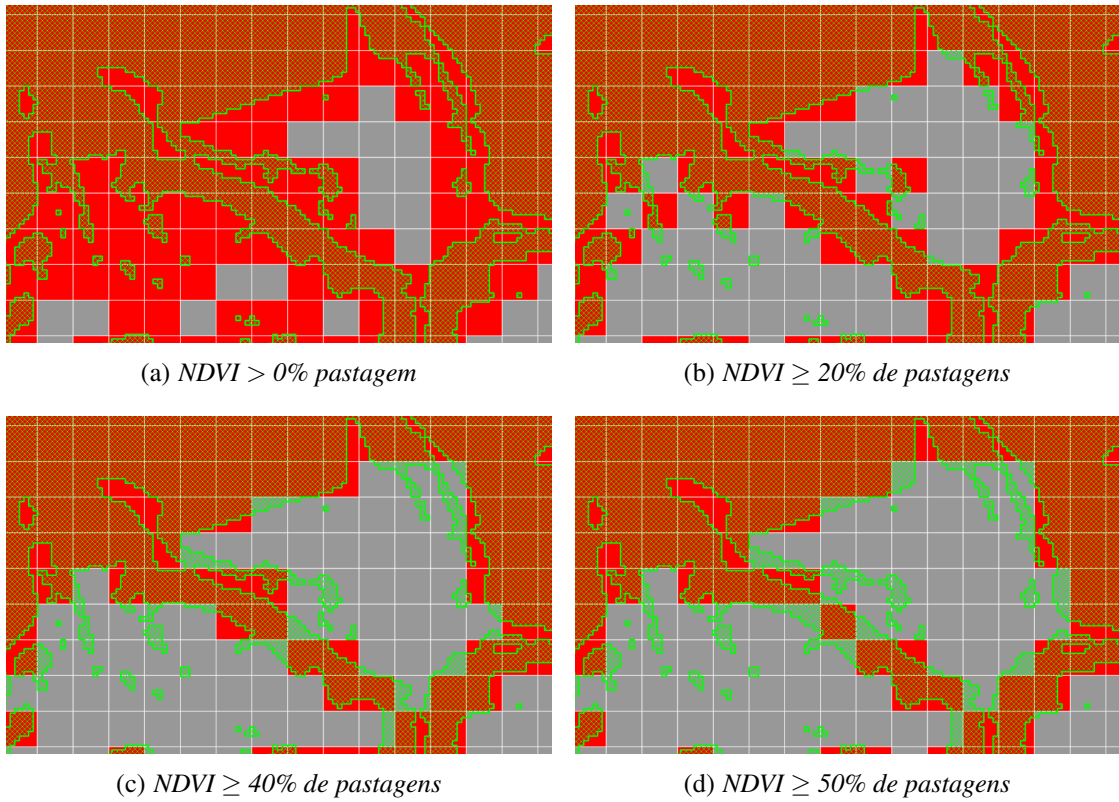
O mapa PASTL8MOD (item 3.2.1.1) foi considerado a melhor opção disponível para ser usado como verdade de campo<sup>4</sup> no processo de validação dos modelos de classificação. Essa escolha se deveu ao fato de que não possuíamos nem a *expertise* nem os recursos necessários para a comparação com imagens de alta resolução ou levantamentos em campo. Por outro lado, o LAPIG030 já havia sido validado por vários especialistas, que utilizaram diversas imagens de alta resolução e outras informações auxiliares.

Enquanto alguns polígonos do PASTL8MOD possuem 100% de sua área coberta por pastagens, muitos outros possuem percentuais bem pequenos, considerando que um pixel Landsat 8 de 30 m representa somente 1,7% de um item da grade NDVI. É preciso definir um limite mínimo do percentual de pastagens para que o pixel possa ser considerado como "Pastagem" ou "Não pastagem". A figura 3.19 mostra quatro alternativas para a escolha deste limiar ( $> 0\%$ ,  $\geq 20\%$ ,  $\geq 40\%$  e  $\geq 50\%$ ). Os *pixels* com a cor cinza representam aqueles que possuem um percentual de pastagens menor do que um limite escolhido, enquanto que a cor vermelha identifica os *pixels* que possuem um percentual de pastagens maior ou igual ao mesmo limite. Em hachura na cor verde, estão as regiões onde há pastagens.

Analisando as quatro situações, optamos pelo percentual mínimo de 40% de área de pastagens (figura 3.19(c)) para compor o mapa PASTL8MOD, sendo que os *pixels* com área de pastagens abaixo deste percentual são isolados, ou estão em regiões de transição e são pouco representativos.

<sup>4</sup>Uma das expressões em português para o termo em inglês *ground truth* e que foi adotada neste documento





**Figura 3.19:** Percentual de pastagens nos pixels NDVI

### 3.2.3 Matriz de Confusão

A análise dos erros e acertos das classificações é feita tomando-se como base de comparação o mapa PASTL8MOD e seu produto é uma matriz de confusão. Os valores e índices apurados na verificação de cada execução e que compõem essa matriz estão descritos a seguir.

Os dados relativos ao recorte de referência que foram retirados do mapa PASTL8MOD são os seguintes:

- PX – Total de *pixels* do recorte.
- RP – Total de pixels rotulados como pastagens.
- RN – Total de pixels rotulados como não pastagens.
- $x_1, \dots, x_m$  – valor do 3º quartil para a  $m$  – *sima* fração de abundância.

A seguir, estão descritos os valores que são apurados a partir da classificação dos *pixels* e de sua comparação com o mapa PASTL8MOD. Para facilitar a referência aos componentes das matrizes de confusão, a letra "S" foi usada para identificar a classe "Pastagens" e a letra "N" para a classe "Não pastagens".

- PC – quantidade de *pixels* que foram classificados como pastagens.
- NC – quantidade de *pixels* que foram classificados como não pastagens.

- SS – quantidade de *pixels* que foram classificados corretamente como pastagens.
- SN – quantidade de *pixels* que foram classificados como pastagens mas não eram.
- NS – quantidade de *pixels* que foram classificados como não sendo pastagens mas eram.
- NN – quantidade de *pixels* que foram classificados corretamente como não sendo pastagens.
- PP – quantidade de *pixels* classificados como pastagens sobre a quantidade real de *pixels* de pastagens ( $PP = PC / RP$ ).
- PN – quantidade de *pixels* classificados como não pastagens sobre a quantidade real de *pixels* que não são de pastagens ( $PN = NC / RN$ ).

A partir de tais valores, são calculados os erros de comissão, omissão e erro total, cujos significados e fórmulas de cálculo estão descritos a seguir:

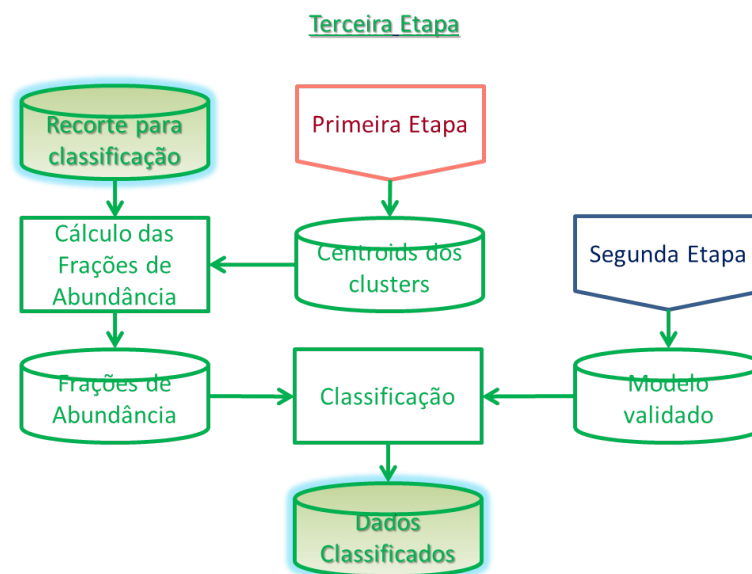
- Erros de comissão - indicam o número de elementos que não pertencem à categoria em análise mas foram incluídos nela.
  - CS – erros de comissão da classe "Pastagens- quantidade de *pixels* que foram classificados como pastagens mas não eram pastagens, dividida pela quantidade total de *pixels* classificados como pastagens. ( $CS = SN / (SS + SN)$ ).
  - CN – erros de comissão da classe "Não Pastagens- quantidade de *pixels* que foram classificados como não pastagens mas eram pastagens, dividida pela quantidade total de *pixels* classificados como não pastagens. ( $CN = NS / (NS + NN)$ ).
- Erros de omissão - indicam o número de elementos que pertencem à categoria em análise mas não foram incluídos nela.
  - OS – erros de omissão da classe "Pastagens- quantidade de *pixels* que foram classificados como não pastagens mas eram pastagens, dividida pela quantidade total de *pixels* que realmente eram pastagens. ( $OS = NS / (SS + NS)$ ).
  - ON – erros de omissão da classe "Não Pastagens- quantidade de *pixels* que foram classificados como pastagens mas não eram pastagens, dividida pela quantidade total de *pixels* que realmente não eram pastagens. ( $ON = SN / (SN + NN)$ ).
- Erro total – quantidade de *pixels* que foram classificados erroneamente sobre o total de *pixels* classificados. ( $ET = (SN + NS) / PX$ ).

Associados aos erros de comissão, omissão e total estão as suas métricas complementares, as acurácias.

- Acurácia do usuário – é o complemento do erro de comissão e estas métricas estão relacionados à confiabilidade do classificador.
- Acurácia do produtor – é o complemento do erro de omissão e estas métricas estão relacionados à sensibilidade do classificador.
- Acurácia total – é o complemento do erro total.

### 3.3 Execução da classificação supervisionada

Na terceira etapa da metodologia é executada a classificação de uma área para a qual se deseja conhecer a cobertura de pastagens. É feito o cálculo das frações de abundâncias do recorte a ser classificado, a partir dos membros finais obtidos na primeira etapa. O passo seguinte é a classificação das áreas de pastagens, ou seja, os *pixels* são selecionados de acordo com os critérios estabelecidos na segunda etapa. O procedimento referente a esta terceira etapa está representado no diagrama da figura 3.20.



**Figura 3.20:** Terceira etapa da metodologia

## Resultados e Discussões

---

Este Capítulo apresenta os resultados obtidos nas três etapas da metodologia e as discussões que surgiram a partir deles.

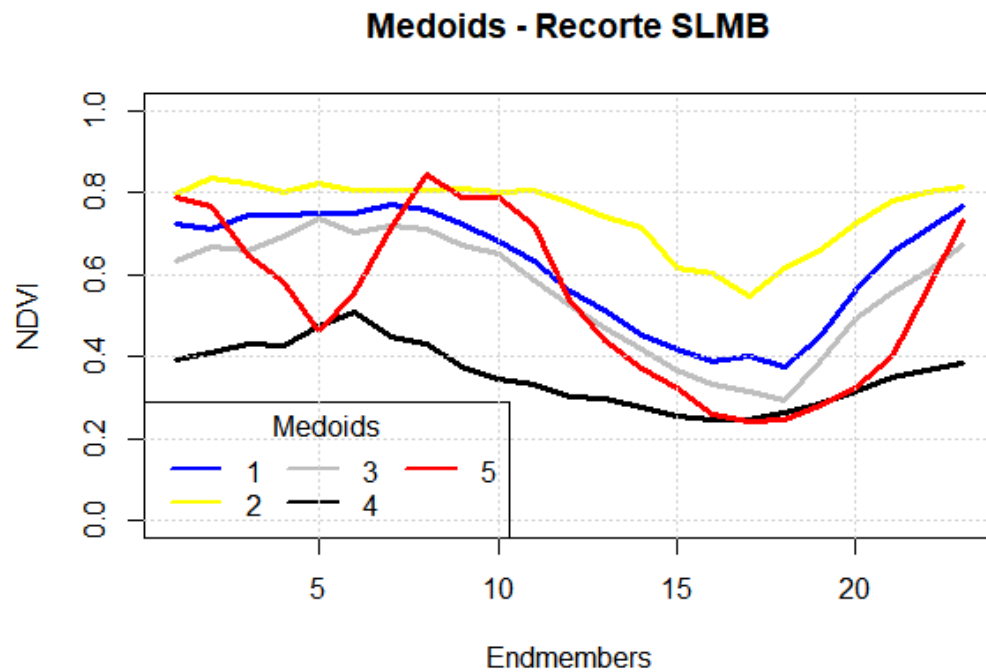
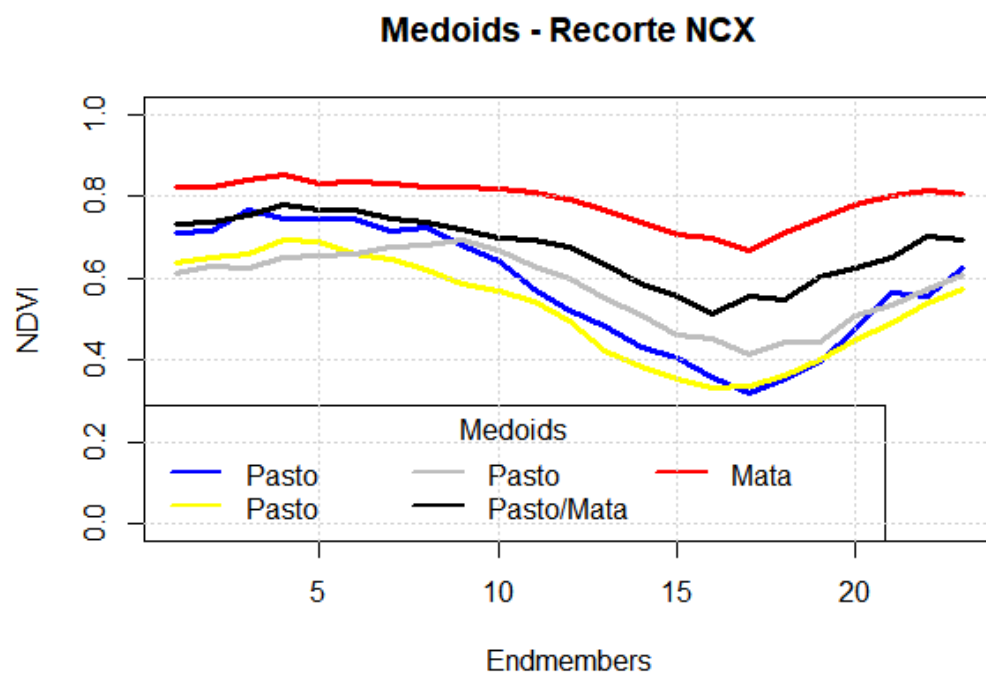
### 4.1 Primeira etapa

A primeira etapa consistiu na obtenção dos membros finais e no cálculo das frações de abundância. Os membros finais foram obtidos a partir dos medóides gerados pela clusterização das áreas de referência, que são os recortes SLMB e NCX, descritos no Capítulo 2, Seção 2.1.

#### 4.1.1 Obtenção dos membros finais

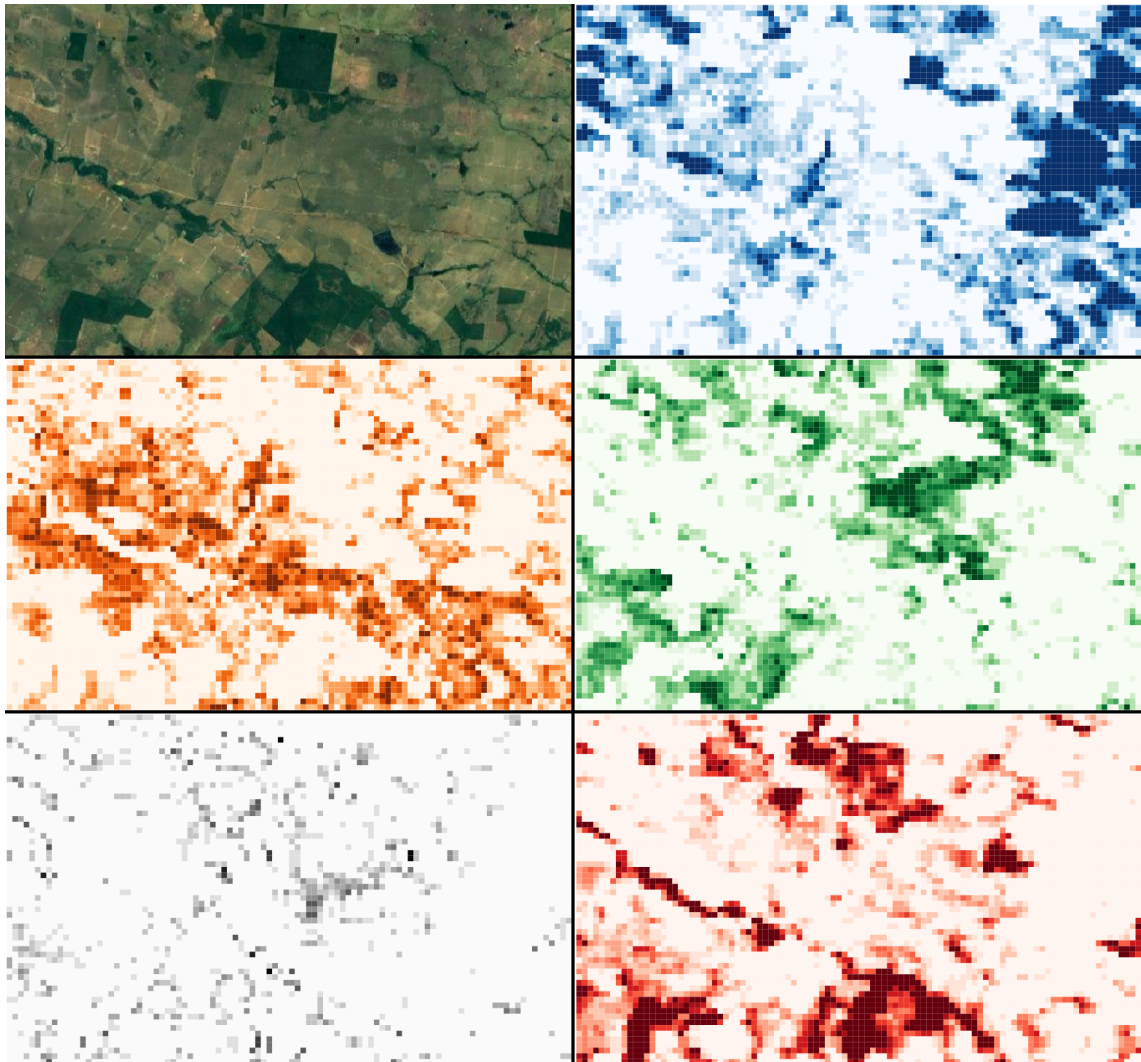
Os membros finais foram obtidos pelo processo de clusterização descrito na Seção 3.1.3.3 e, para cada um dos recortes de referência, foram gerados cinco clusters. Os perfis de NDVI dos medóides destes clusters, representados graficamente na figura 4.1, correspondem ao comportamento fenológico típico das coberturas presentes nos respectivos recortes. No caso do recorte SLMB, os perfis de seus medóides representam a diversidade de coberturas nele presentes, tais como áreas urbanas, mata, agricultura e pastagens. O recorte NCX, por ser mais homogêneo, gerou medóides com perfis semelhantes uns aos outros que representam, majoritariamente, as áreas de pastagens.

A escolha de duas áreas de referência com características diferentes teve como objetivo a avaliação do desempenho do classificador usando cada um dos modelos produzidos e aplicados em outras áreas também distintas entre si.

(a) *Membros finais - recorte SLMB*(b) *Membros finais - recorte NCX***Figura 4.1:** *Membros finais dos recortes de referência*

### 4.1.2 Frações de abundância

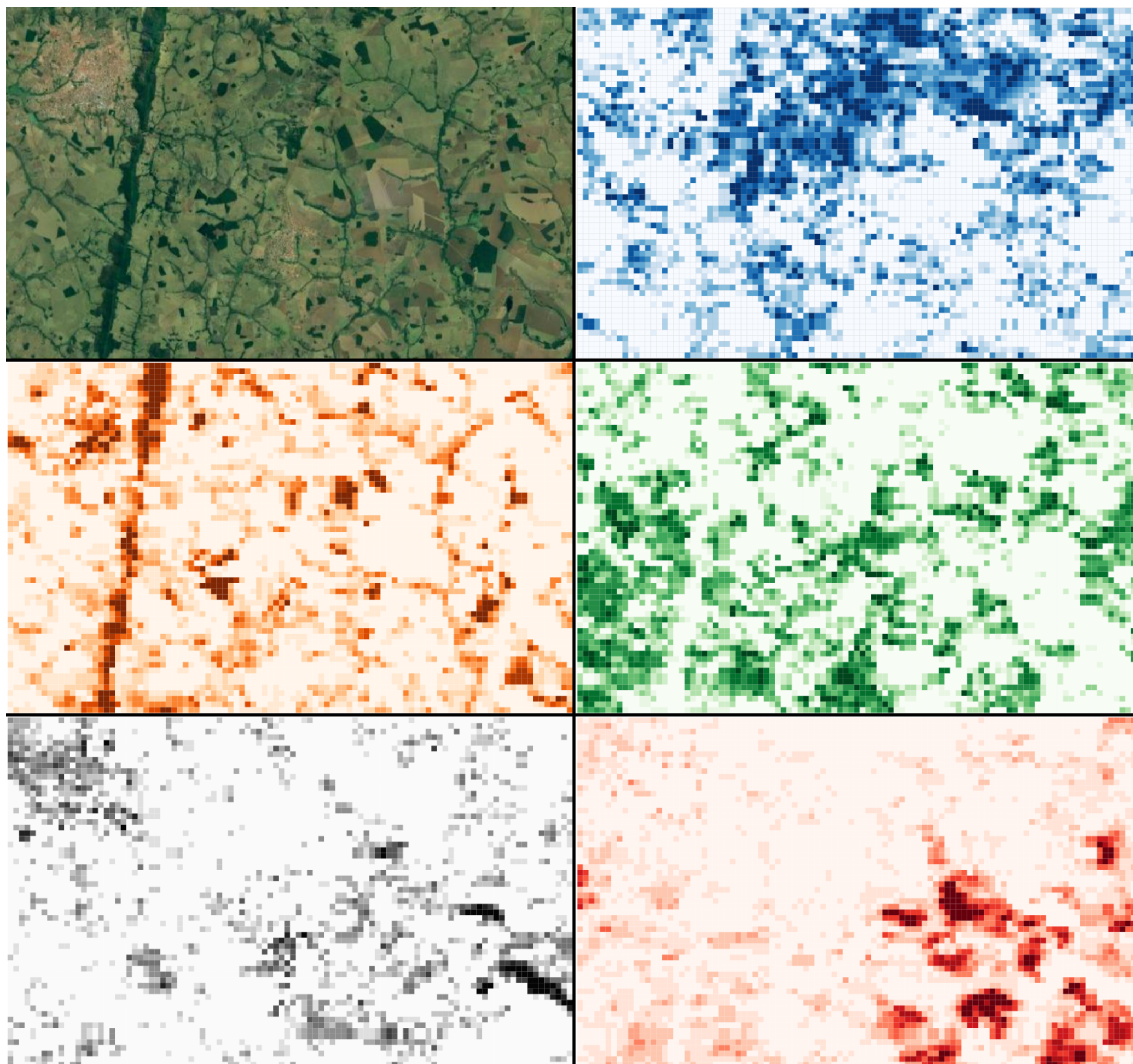
Depois de obtidos os membros finais foi feito o cálculo das frações de abundância, para apurar a proporção com que cada membro final participou da composição dos *pixels* de cada recorte de referência. As figuras 4.2 e 4.3 mostram o comportamento de cada uma das cinco frações de abundância nos dois recortes de referência.



**Figura 4.2:** Frações de abundância - recorte NCX  
x11(azul) - x22(laranja) - x3(verde) - x4(cinza) -  
x5(vermelho)

Para o recorte NCX, as frações de abundância x1, x2 e x3 são as que melhor representam as áreas de pastagens. A fração x5 corresponde às áreas de matas enquanto que a fração x4 tem pouco destaque e não representa nenhuma feição específica.

No caso do recorte SLMB, as frações de abundância x1 e x3 são as que melhor representam as áreas de pastagens. A fração x2 representa melhor as áreas de mata, a x4 as áreas urbanas e a x5 as áreas agrícolas.



**Figura 4.3:** *Frações de abundância - recorte SLMB*  
*x1(azul) - x2(laranja) - x3(verde) - x4(cinza) -*  
*x5(vermelho)*

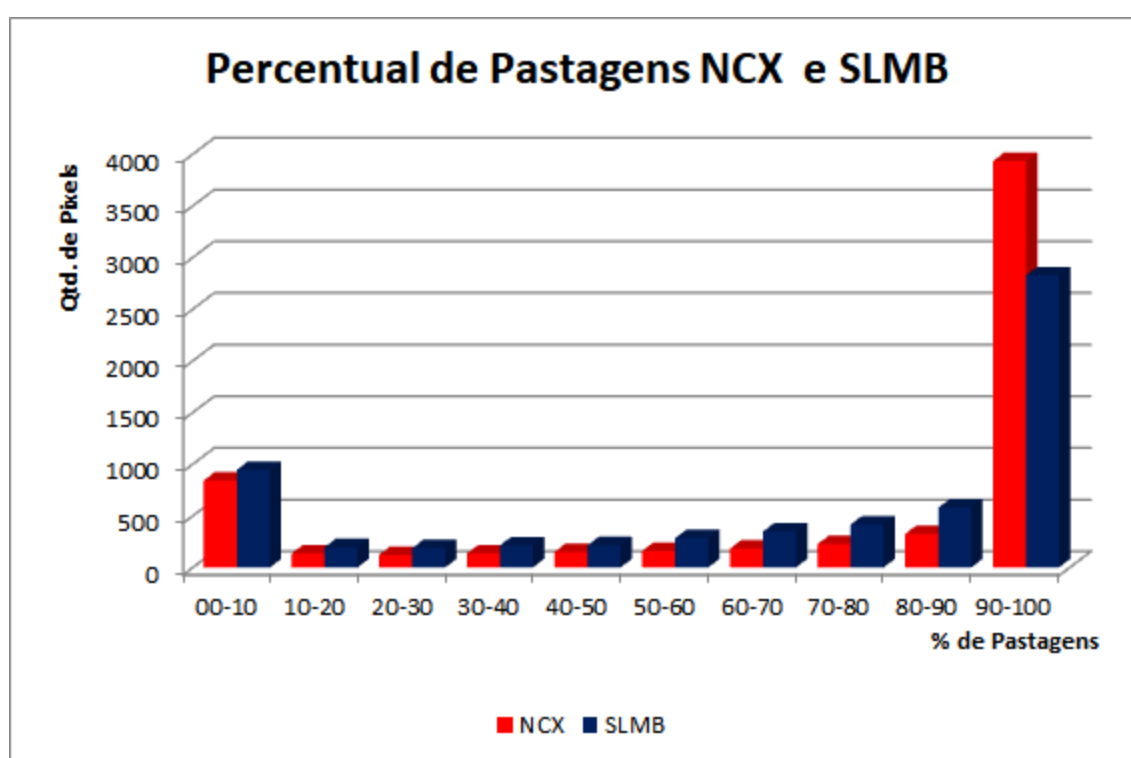
## 4.2 Segunda etapa

Conforme descrito na metodologia, nesta etapa foi construído o modelo de classificação supervisionada, que compreendeu o cálculo do percentual de pastagens para cada pixel da grade NDVI; a correlação entre estes percentuais e as frações de abundância; o estabelecimento dos limiares das frações de abundância para a seleção dos *pixels* de pastagens e a validação do modelo de classificação por meio de comparação com a verdade de campo.

Para a correta identificação da correspondência entre os valores dos *pixels* dos recortes de referência e as feições neles presentes, é necessária a análise de um especialista. Em nossa metodologia assumimos que o mapa LAPIG030 supriu esta função, considerando que ele já tem corretamente identificados todos os *pixels* que pertencem à classe de pastagens.

### 4.2.1 Cálculo do percentual de pastagens

O percentual de pastagens de cada pixel de cada um dos recortes de referência, foi calculado de acordo com o item 3.2.1.1, ou seja, o mapa PASTL8MOD resultou das seguintes etapas: o mapa LAPIG030 foi recortado pela grade NDVI do recorte; foi calculada a área das pastagens contidas em cada polígono da grade; o valor da área das pastagens foi dividido pela área total da polígono, resultando no seu percentual de pastagens; por fim, foram acrescidas as frações de abundância de cada pixel, calculadas na primeira etapa. A figura 4.4 apresenta o histograma dos percentuais de pastagens nos dois recortes de referência, na qual é possível confirmar que o recorte NCX possui maior quantidade de pastagens do que o recorte SLMB na faixa entre 90% e 100%.



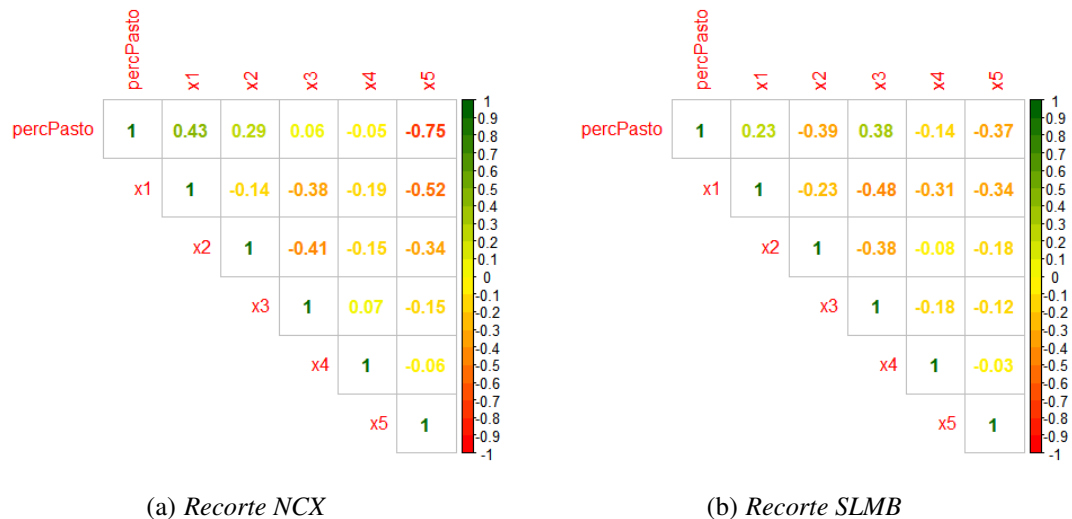
**Figura 4.4:** Percentual de pastagens nos recortes de referência NCX e SLMB

### 4.2.2 Correlação entre o percentual de pastagens e as frações de abundância

Foi calculada a correlação entre o percentual de pastagens e as frações de abundância com o objetivo de identificar os casos de correlação positiva, ou seja, quais frações de abundância tinham seus valores crescendo ao mesmo tempo em que cresciam os percentuais de pastagens. A figura 4.5 mostra a correlação existente entre o percentual de pastagens e as frações de abundância dos recortes de referência NCX e SLMB. No



recorte NCX as frações x1, x2 e x3 apresentam correlação positiva enquanto que no recorte SLMB, são as frações x1 e x3 que aumentam ao mesmo tempo que o percentual de pastagens.



**Figura 4.5:** Correlação entre as frações de abundância e os percentuais de pastagens

### 4.2.3 Coeficiente de multiplicação para seleção das frações de abundância

O coeficiente de multiplicação  $cm$  para seleção das frações de abundância foi calculado de acordo com o algoritmo 3.3, descrito no item 3.2.1.3. No caso do recorte NCX, o coeficiente de multiplicação  $cm$  foi aplicado às frações de abundância x1, x2 e x3 e foi definido com o valor 0,80, conforme tabela 4.1 e gráfico contido na figura 4.6, resultando na sentença de seleção  $sql$  descrita na equação 4-1.

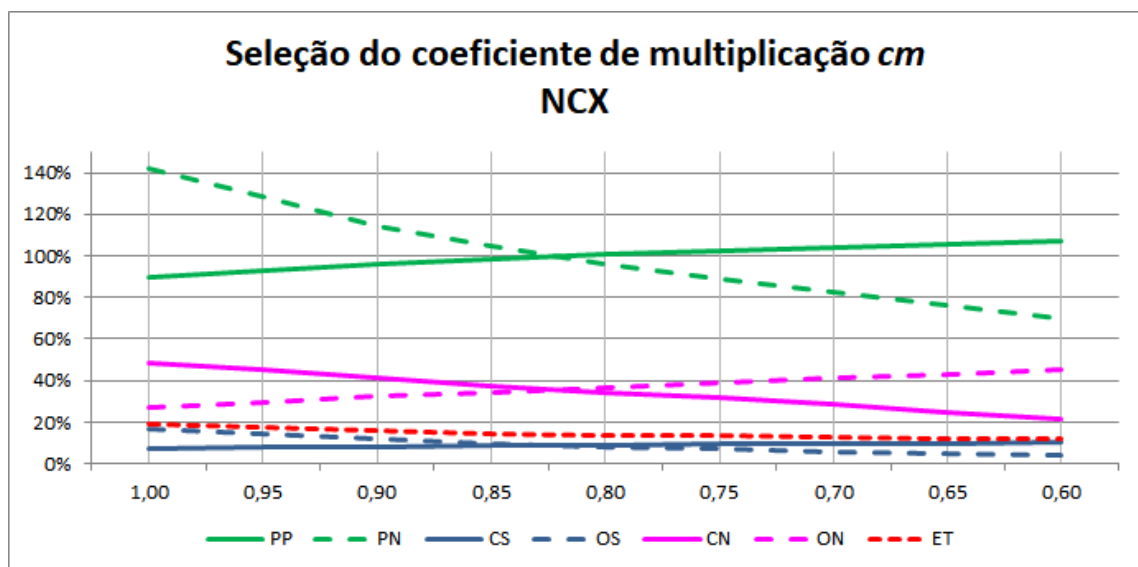
$$sql = "x1" \geq (0.80 \cdot 0.4647) \quad OR \quad "x2" \geq (0.80 \cdot 0.4610) \quad OR \quad "x3" \geq (0.8 \geq 0.3665) \quad (4-1)$$

**Tabela 4.1:** Tabela de seleção do coeficiente *cm* - recorte NCX

|                      |         |
|----------------------|---------|
| PX - Total de pixels | 6200    |
| RP - Pastagens       | 4970    |
| RN - Não pastagens   | 1230    |
| Fração abundância x1 | 0,46471 |
| Fração abundância x2 | 0,46101 |
| Fração abundância x3 | 0,36650 |

| <i>m</i> | PC   | NC   | SS   | SN  | NN  | NS  | PP      | PN      | CS     | OS     | CN     | ON     | ET     | PP-PN   | CS-OS  | CN-ON   |
|----------|------|------|------|-----|-----|-----|---------|---------|--------|--------|--------|--------|--------|---------|--------|---------|
| 1,00     | 4454 | 1746 | 4124 | 330 | 900 | 846 | 89,62%  | 141,95% | 7,41%  | 17,02% | 48,45% | 26,83% | 18,97% | -52,33% | -9,61% | 21,62%  |
| 0,95     | 4620 | 1580 | 4256 | 364 | 866 | 714 | 92,96%  | 128,46% | 7,88%  | 14,37% | 45,19% | 29,59% | 17,39% | -35,50% | -6,49% | 15,60%  |
| 0,90     | 4790 | 1410 | 4390 | 400 | 830 | 580 | 96,38%  | 114,63% | 8,35%  | 11,67% | 41,13% | 32,52% | 15,81% | -18,26% | -3,32% | 8,61%   |
| 0,85     | 4909 | 1291 | 4486 | 423 | 807 | 484 | 98,77%  | 104,96% | 8,62%  | 9,74%  | 37,49% | 34,39% | 14,63% | -6,19%  | -1,12% | 3,10%   |
| 0,80     | 5020 | 1180 | 4569 | 451 | 779 | 401 | 101,01% | 95,93%  | 8,98%  | 8,07%  | 33,98% | 36,67% | 13,74% | 5,07%   | 0,92%  | -2,68%  |
| 0,75     | 5102 | 1098 | 4622 | 480 | 750 | 348 | 102,66% | 89,27%  | 9,41%  | 7,00%  | 31,69% | 39,02% | 13,35% | 13,39%  | 2,41%  | -7,33%  |
| 0,70     | 5185 | 1015 | 4681 | 504 | 726 | 289 | 104,33% | 82,52%  | 9,72%  | 5,81%  | 28,47% | 40,98% | 12,79% | 21,81%  | 3,91%  | -12,50% |
| 0,65     | 5259 | 941  | 4735 | 524 | 706 | 235 | 105,81% | 76,50%  | 9,96%  | 4,73%  | 24,97% | 42,60% | 12,24% | 29,31%  | 5,24%  | -17,63% |
| 0,60     | 5339 | 861  | 4784 | 555 | 675 | 186 | 107,42% | 70,00%  | 10,40% | 3,74%  | 21,60% | 45,12% | 11,95% | 37,42%  | 6,65%  | -23,52% |

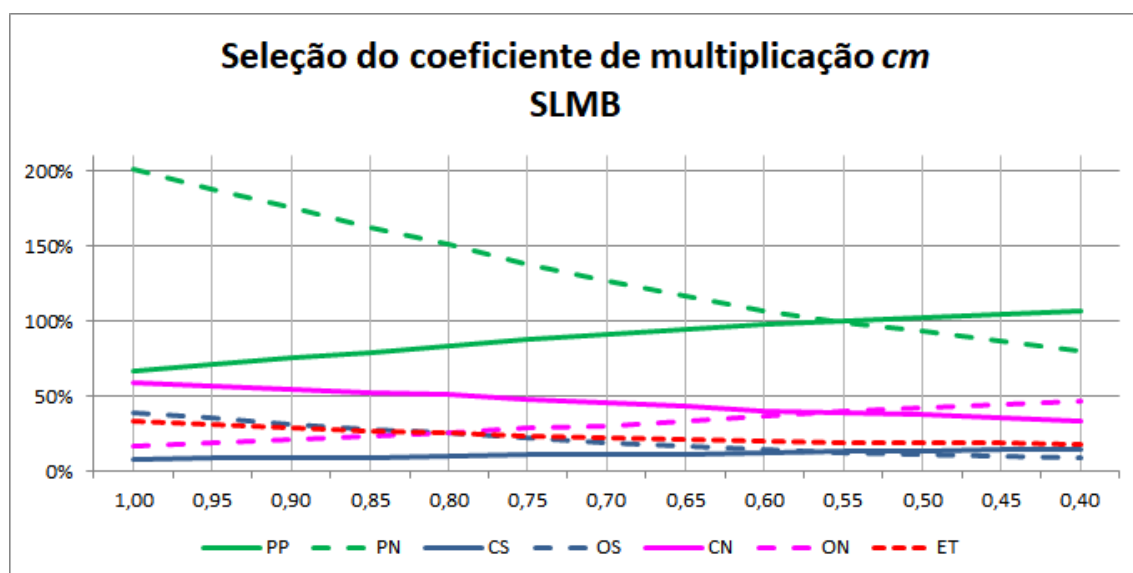
**Figura 4.6:** Gráfico da seleção do coeficiente *cm* - recorte NCX

Para o recorte SLMB, o coeficiente de multiplicação *cm* foi aplicado às frações de abundância x1 e x3 e foi definido com o valor 0,55, conforme tabela 4.2 e gráfico contido na figura 4.7, resultando na sentença de seleção descrita na equação 4-2.

$$sql = "x1" \geq (0,55 \cdot 0,5670) \quad OR \quad "x3" \geq (0,8 \geq 0,4980) \quad (4-2)$$

**Tabela 4.2:** Tabela de seleção do coeficiente *cm* - recorte SLMB

| PX - Total de pixels |      |      |      |     |      |      | 6200    |         |        |        |        |        |        |          |         |         |
|----------------------|------|------|------|-----|------|------|---------|---------|--------|--------|--------|--------|--------|----------|---------|---------|
| RP - Pastagens       |      |      |      |     |      |      | 4662    |         |        |        |        |        |        |          |         |         |
| RN - Não pastagens   |      |      |      |     |      |      | 1538    |         |        |        |        |        |        |          |         |         |
| Fração abundância x1 |      |      |      |     |      |      | 0,56696 |         |        |        |        |        |        |          |         |         |
| Fração abundância x3 |      |      |      |     |      |      | 0,49801 |         |        |        |        |        |        |          |         |         |
| cm                   | PC   | NC   | SS   | SN  | NN   | NS   | PP      | PN      | CS     | OS     | CN     | ON     | ET     | PP-PN    | CS-OS   | CN-ON   |
| 1,00                 | 3100 | 3100 | 2841 | 259 | 1279 | 1821 | 66,50%  | 201,56% | 8,35%  | 39,06% | 58,74% | 16,84% | 33,55% | -135,07% | -30,71% | 41,90%  |
| 0,95                 | 3307 | 2893 | 3018 | 289 | 1249 | 1644 | 70,94%  | 188,10% | 8,74%  | 35,26% | 56,83% | 18,79% | 31,18% | -117,17% | -26,52% | 38,04%  |
| 0,90                 | 3507 | 2693 | 3183 | 324 | 1214 | 1479 | 75,23%  | 175,10% | 9,24%  | 31,72% | 54,92% | 21,07% | 29,08% | -99,87%  | -22,49% | 33,85%  |
| 0,85                 | 3701 | 2499 | 3347 | 354 | 1184 | 1315 | 79,39%  | 162,48% | 9,56%  | 28,21% | 52,62% | 23,02% | 26,92% | -83,10%  | -18,64% | 29,60%  |
| 0,80                 | 3875 | 2325 | 3482 | 393 | 1145 | 1180 | 83,12%  | 151,17% | 10,14% | 25,31% | 50,75% | 25,55% | 25,37% | -68,05%  | -15,17% | 25,20%  |
| 0,75                 | 4077 | 2123 | 3637 | 440 | 1098 | 1025 | 87,45%  | 138,04% | 10,79% | 21,99% | 48,28% | 28,61% | 23,63% | -50,58%  | -11,19% | 19,67%  |
| 0,70                 | 4244 | 1956 | 3774 | 470 | 1068 | 888  | 91,03%  | 127,18% | 11,07% | 19,05% | 45,40% | 30,56% | 21,90% | -36,14%  | -7,97%  | 14,84%  |
| 0,65                 | 4408 | 1792 | 3891 | 517 | 1021 | 771  | 94,55%  | 116,51% | 11,73% | 16,54% | 43,02% | 33,62% | 20,77% | -21,96%  | -4,81%  | 9,41%   |
| 0,60                 | 4557 | 1643 | 3996 | 561 | 977  | 666  | 97,75%  | 106,83% | 12,31% | 14,29% | 40,54% | 36,48% | 19,79% | -9,08%   | -1,97%  | 4,06%   |
| 0,55                 | 4672 | 1528 | 4063 | 609 | 929  | 599  | 100,21% | 99,35%  | 13,04% | 12,85% | 39,20% | 39,60% | 19,48% | 0,86%    | 0,19%   | -0,40%  |
| 0,50                 | 4769 | 1431 | 4117 | 652 | 886  | 545  | 102,30% | 93,04%  | 13,67% | 11,69% | 38,09% | 42,39% | 19,31% | 9,25%    | 1,98%   | -4,31%  |
| 0,45                 | 4871 | 1329 | 4185 | 686 | 852  | 477  | 104,48% | 86,41%  | 14,08% | 10,23% | 35,89% | 44,60% | 18,76% | 18,07%   | 3,85%   | -8,71%  |
| 0,40                 | 4967 | 1233 | 4248 | 719 | 819  | 414  | 106,54% | 80,17%  | 14,48% | 8,88%  | 33,58% | 46,75% | 18,27% | 26,37%   | 5,60%   | -13,17% |

**Figura 4.7:** Gráfico da seleção do coeficiente *cm* - recorte SLMB

#### 4.2.4 Validação em relação ao mapa LAPIG250

Foi feita uma verificação adicional da qualidade do modelo de classificação gerado nesta metodologia, por meio da comparação de seus resultados com aqueles obtidos quando verificamos a exatidão do mapa LAPIG250 em relação ao mapa LAPIG030.

##### 4.2.4.1 Geração de grade compatível com o LAPIG250

Conforme descrito na Seção 2.2, os *pixels* da grade NDVI identificados por 250 m têm, na verdade, resolução espacial de 231,656 m (veja figura 4.8(a)). Por outro lado, o mapa LAPIG250 tem a resolução espacial efetiva de 250 m e isto faz com que os pixels deste mapa não se encaixem corretamente nos recortes que foram feitos com base na grade

NDVI (veja figura 4.8(b)). A solução para este problema foi a criação, para os recortes de referência, de uma grade com resolução espacial de 250 m, compatível com o mapa LAPIG250 (veja figura 4.8(c)). Por ter uma área um pouco maior do que a da grade NDVI, esta grade de 250 m possui um número de quadriculas também um pouco menor para cada recorte.

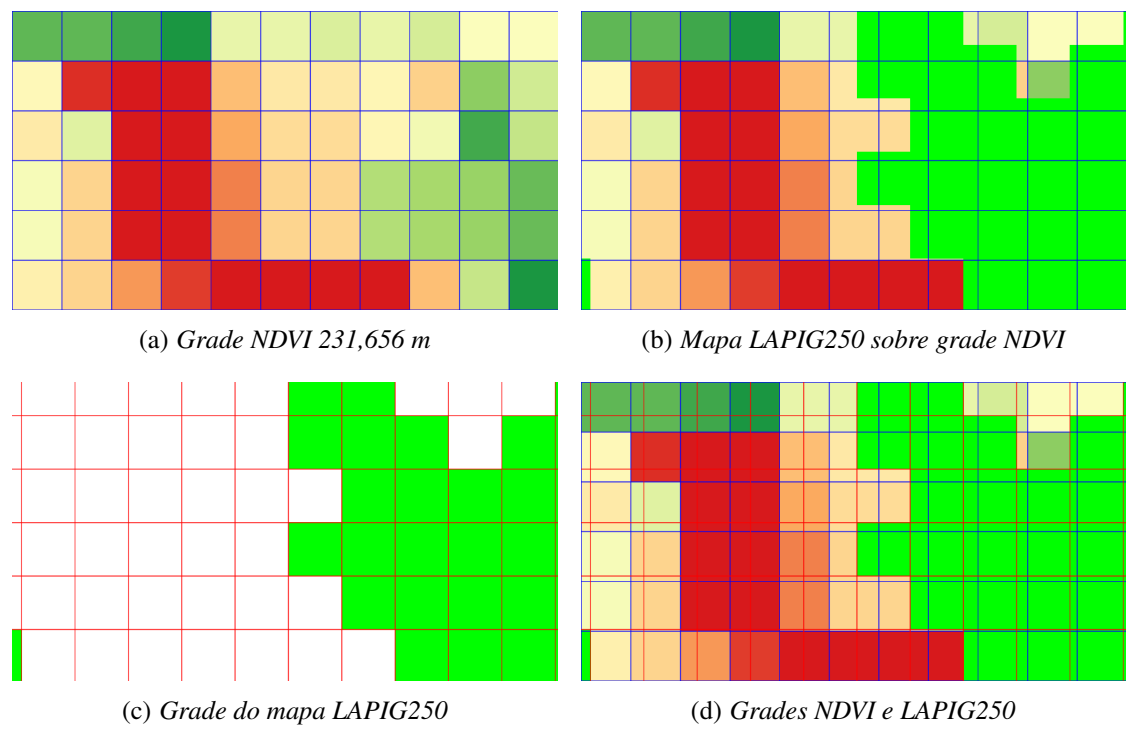


Figura 4.8: Desencontro entre as grades NDVI x LAPIG250

4.2.4.2 Comparação das classificações dos recortes de referência com o mapa LAPIG250

A classificação do recorte SLMB pelo modelo SLMB produziu um erro total de 19,48% enquanto que o mapa LAPIG250 apresentou um erro total de 18,99% em comparação com o mapa LAPIG030 para o mesmo recorte, conforme tabela 4.3.

Tabela 4.3: Matriz de confusão - recorte SLMB

| Recorte:        |   | SLMB       |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | SLMB       |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 4.063      | 609    | 4.672        | 13,04%           |
|                 | N | 599        | 929    | 1.528        | 39,20%           |
| Total Colunas   |   | 4.662      | 1.538  | 6.200        | 19,48%           |
| Erro de Omissão |   | 12,85%     | 39,60% | Erro Total   |                  |

| Recorte:        |   | SLMB       |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | LAPIG      |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 3.316      | 304    | 3.620        | 8,40%            |
|                 | N | 692        | 932    | 1.624        | 42,61%           |
| Total Colunas   |   | 4.008      | 1.236  | 5.244        | 18,99%           |
| Erro de Omissão |   | 17,27%     | 24,60% | Erro Total   |                  |

A classificação do recorte NCX pelo modelo NCX produziu um erro total de 13,74% enquanto que o mapa LAPIG250 apresentou um erro total de 20,10% em comparação com o mapa LAPIG030 para o mesmo recorte, conforme tabela 4.4.

**Tabela 4.4:** Matriz de confusão - recorte NCX

| Recorte:        |   | NCX        |        |              |                  | Recorte:        |   | NCX        |       |              |                  |
|-----------------|---|------------|--------|--------------|------------------|-----------------|---|------------|-------|--------------|------------------|
| Classificação:  |   | NCX        |        |              |                  | Classificação:  |   | LAPIG      |       |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão | Pastagens       |   | Referência |       | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |                 |   | S          | N     |              |                  |
| Previsto        | S | 4.569      | 451    | 5.020        | 8,98%            | Previsto        | S | 3.217      | 43    | 3.260        | 1,32%            |
|                 | N | 401        | 779    | 1.180        | 33,98%           |                 | N | 1.011      | 973   | 1.984        | 50,96%           |
| Total Colunas   |   | 4.970      | 1.230  | 6.200        | 13,74%           | Total Colunas   |   | 4.228      | 1.016 | 5.244        | 20,10%           |
| Erro de Omissão |   | 8,07%      | 36,67% | Erro Total   |                  | Erro de Omissão |   | 23,91%     | 4,23% | Erro Total   |                  |

O desempenho dos modelos de classificação nos recortes de referência, quando comparado com os índices do mapa LAPIG250 em relação ao mapa LAPIG030, foi considerado satisfatório tendo em vista que o modelo SLMB foi apenas 2,58% pior do que o LAPIG250 que, por sua vez, foi 46,29% pior do que o modelo NCX.

## 4.3 Terceira etapa

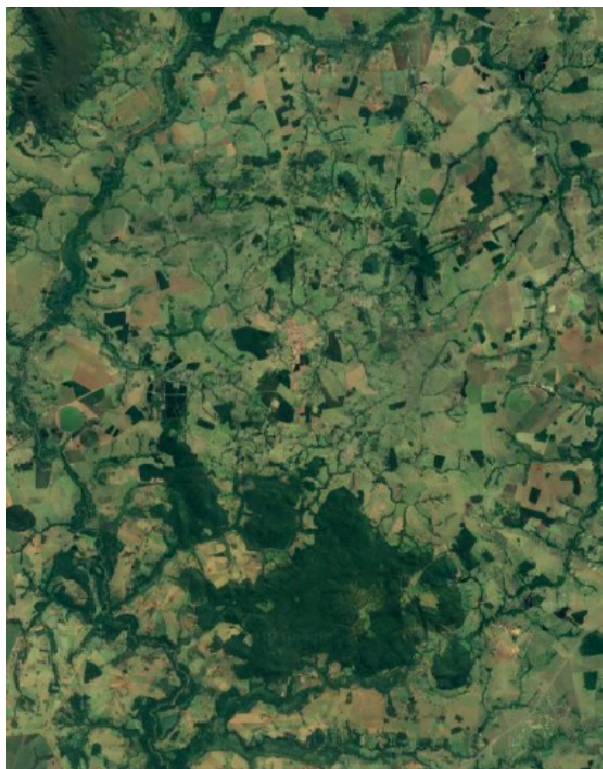
### 4.3.1 Classificação dos recortes de classificação

Uma vez que os modelos foram aprovados na classificação dos recortes de referência, a terceira etapa da metodologia consistiu na aplicação de tais modelos aos recortes de classificação.

#### 4.3.1.1 Recorte CDG

Este recorte abrange todo o Município de Campestre de Goiás e apresenta uma boa diversidade de feições terrestres tais como áreas de pastagens, agricultura, hidrografia de pequeno porte e áreas urbanas, conforme figura 4.9.

A tabela 4.5 apresenta os resultados produzidos pela classificação do recorte CDG pelos modelos de referência. O erro total de 21,88% na classificação com o modelo SLMB foi próximo ao percentual encontrado quando este modelo foi aplicado ao seu próprio recorte de referência (19,48%). Isto se deve ao fato de que as áreas destes dois recortes apresentam semelhança em suas características. Porém, no caso do modelo NCX, cujo recorte de referência mostra uma predominância de áreas de pastagens e uma pequena quantidade de matas, o erro total subiu para 24,13%, o que reflete as diferenças das feições terrestres existentes nos dois recortes.



**Figura 4.9:** Recorte CDG

**Tabela 4.5:** Matriz de confusão - recorte CDG

| Recorte:        |   | CDG        |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | NCX        |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 6.438      | 2.053  | 8.491        | 24,18%           |
|                 | N | 522        | 1.659  | 2.181        | 23,93%           |
| Total Colunas   |   | 6.960      | 3.712  | 10.672       | 24,13%           |
| Erro de Omissão |   | 7,50%      | 55,31% | Erro Total   |                  |

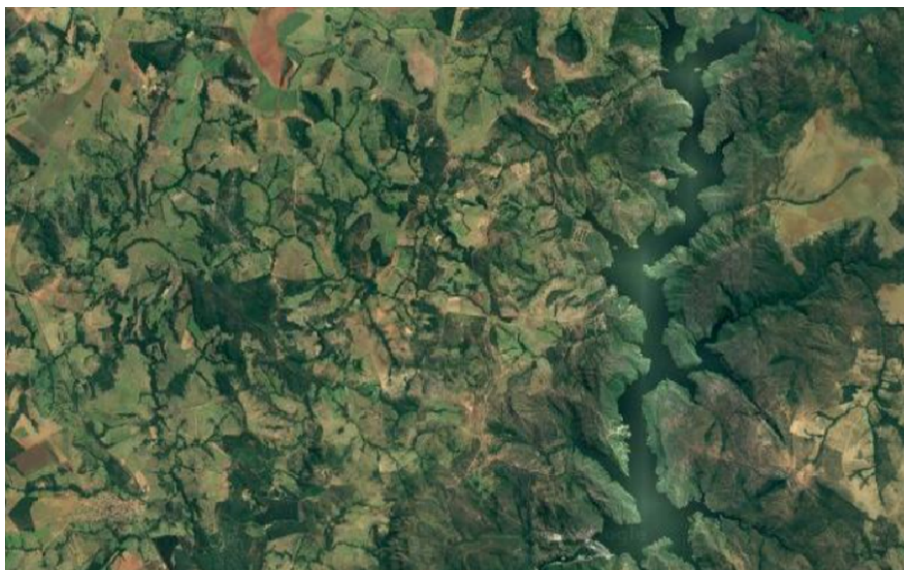
| Recorte:        |   | CDG        |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | SLMB       |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 6.025      | 1.400  | 7.425        | 18,86%           |
|                 | N | 935        | 2.312  | 3.247        | 28,80%           |
| Total Colunas   |   | 6.960      | 3.712  | 10.672       | 21,88%           |
| Erro de Omissão |   | 13,43%     | 37,72% | Erro Total   |                  |

#### 4.3.1.2 Recorte CLV

O recorte CLV abrange parte dos municípios de Marzagão e Caldas Novas e tem como características poucas áreas de pastagens e abundância de hidrografia de grande porte, conforme figura 4.10.

A tabela 4.6 mostra as matrizes de confusão resultantes da classificação do recorte CLV pelos dois modelos de referência. O erro total de 26,16% na classificação com o modelo SLMB aumentou em relação ao recorte CDG pelo fato de que também aumentaram as diferenças nas características dos dois recortes. O erro aumentou significativamente na classificação pelo modelo NCX que, no caso, foi de 38,15%, considerando que o recorte CLV difere ainda mais do recorte NCX do que difere do recorte SLMB.





**Figura 4.10:** Recorte CLV

**Tabela 4.6:** Matriz de confusão - recorte CLV

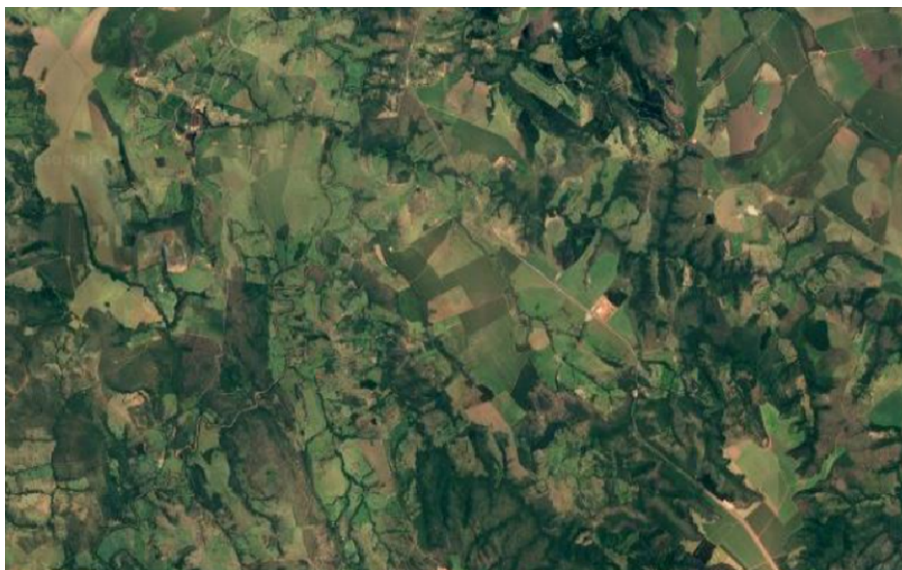
| Recorte:        |   | CLV        |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | NCX        |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 2.156      | 2.050  | 4.206        | 48,74%           |
|                 | N | 315        | 1.679  | 1.994        | 15,80%           |
| Total Colunas   |   | 2.471      | 3.729  | 6.200        | 38,15%           |
| Erro de Omissão |   | 12,75%     | 54,97% | Erro Total   |                  |

| Recorte:        |   | CLV        |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | SLMB       |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 1.564      | 715    | 2.279        | 31,37%           |
|                 | N | 907        | 3.014  | 3.921        | 23,13%           |
| Total Colunas   |   | 2.471      | 3.729  | 6.200        | 26,16%           |
| Erro de Omissão |   | 36,71%     | 19,17% | Erro Total   |                  |

#### 4.3.1.3 Recorte LUZ

O recorte LUZ fica no Município de Luziânia e tem como características poucas áreas de pastagens, um grande número de áreas de agricultura e muita hidrografia de pequeno porte, conforme figura 4.11.

A tabela 4.7 mostra as matrizes de confusão resultantes da classificação do recorte LUZ pelos dois modelos de referência. As características deste recorte diferem bastante daquelas do recorte SLMB e muito mais ainda daquelas do recorte NCX. Em decorrência, as suas classificações pelos modelos gerados por estes dois recortes de referência apresentaram erros totais bastantes expressivos, indicando uma inadequação desta metodologia nestes casos. O erro total na classificação com o modelo SLMB foi de 33,47%, enquanto que na classificação com o modelo NCX foi de 60,71%.



**Figura 4.11:** Recorte LUZ

**Tabela 4.7:** Matriz de confusão - recorte LUZ

| Recorte:        |   | LUZ        |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | NCX        |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 1.846      | 3.697  | 5.543        | 66,70%           |
|                 | N | 67         | 590    | 657          | 10,20%           |
| Total Colunas   |   | 1.913      | 4.287  | 6.200        | 60,71%           |
| Erro de Omissão |   | 3,50%      | 86,24% | Erro Total   |                  |

| Recorte:        |   | LUZ        |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | SLMB       |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 1.347      | 1.509  | 2.856        | 52,84%           |
|                 | N | 566        | 2.778  | 3.344        | 16,93%           |
| Total Colunas   |   | 1.913      | 4.287  | 6.200        | 33,47%           |
| Erro de Omissão |   | 29,59%     | 35,20% | Erro Total   |                  |

### 4.3.2 Definição de um novo recorte - REC39K

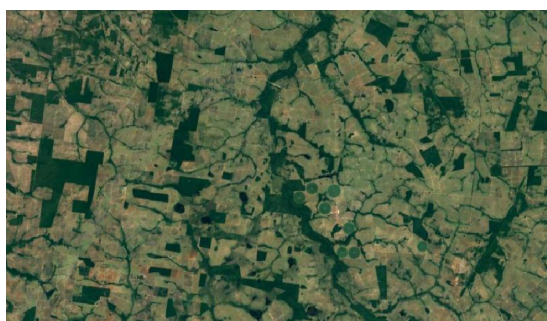
De acordo com os resultados das classificações dos recortes selecionados inicialmente, ficou clara a tendência de que quanto maiores fossem as diferenças entre as características das feições dos recortes de referência e dos de classificação, pior seriam os resultados. Partindo do pressuposto de que áreas próximas e com semelhanças visuais implicam em características fenológicas semelhantes, foi selecionado um novo recorte para classificação, o REC39K, dentro do Município de Nova Crixás, com localização próxima ao recorte de referência NCX e com características semelhantes. O recorte REC39K foi classificado usando apenas o modelo NCX e o erro total foi de 14,26% (tabela 4.8), bem próximo aos 13,74% apurados quando o recorte NCX foi classificado pelo modelo NCX.

A figura 4.12 mostra uma imagem do recorte REC39K (4.12(a)), sua cobertura de pastagens segundo o mapa LAPIG030 (4.12(b)), as pastagens classificadas pelo modelo NCX (4.12(c)) e o principal erro encontrado, que foi o de omissão para a classe de "Não pastagens" em uma região de agricultura irrigada (figura 4.12(d)) e que não estava bem representada no modelo. Este resultado comprovou que a semelhança de características entre as feições que originam um modelo e aquelas a serem classificadas, são fundamentais para que os resultados tenham um baixo índice de erro.

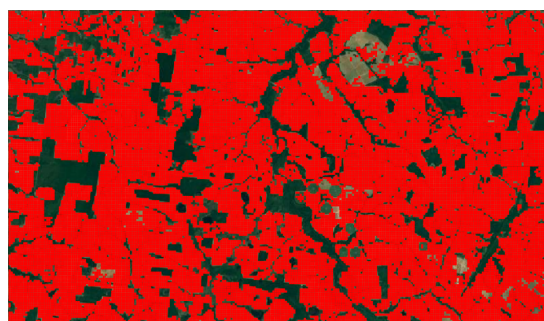


**Tabela 4.8:** Matriz de confusão - recorte REC39K

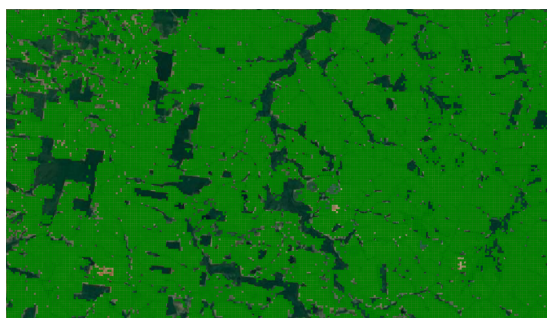
| Recorte:        |   | REC39K     |        |              |                  |
|-----------------|---|------------|--------|--------------|------------------|
| Classificação:  |   | NCX        |        |              |                  |
| Pastagens       |   | Referência |        | Total Linhas | Erro de Comissão |
|                 |   | S          | N      |              |                  |
| Previsto        | S | 27.618     | 4.024  | 31.642       | 12,72%           |
|                 | N | 1.537      | 5.821  | 7.358        | 20,89%           |
| Total Colunas   |   | 29.155     | 9.845  | 39.000       | 14,26%           |
| Erro de Omissão |   | 5,27%      | 40,87% | Erro Total   |                  |



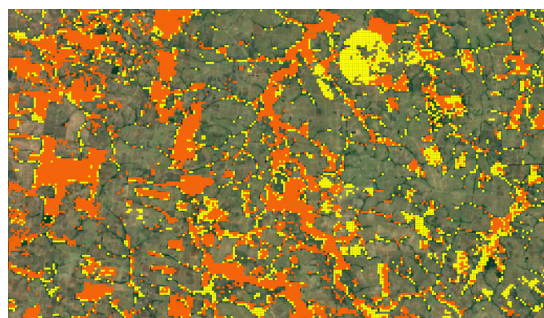
(a) Recorte REC39K



(b) Mapa LAPIG030 sobre recorte REC39K



(c) Pastagens classificadas pelo modelo NCX



(d) Erros de Omissão - Não Pastagens

**Figura 4.12:** Classificador NCX aplicado ao recorte REC39K

### 4.3.3 Comparação com os resultados do mapa LAPIG250

Considerando que os modelos de classificação foram criados com base no produto MOD13Q1 e, considerando que o LAPIG250 tem sua origem no mesmo produto, avaliamos o desempenho dos modelos de classificação e do LAPIG250 mediante uma mesma fonte de comparação. Para cada recorte, seja de referência ou de classificação, elaboramos uma matriz de confusão com os resultados da aplicação dos modelos de classificação. Para os mesmos recortes, também criamos matrizes de confusão contendo os resultados da validação do LAPIG250. Conforme o item 3.2.2, o mapa PASTL8MOD foi utilizado como verdade de campo, ou seja, assumiu-se que seus dados estão corretos. A tabela 4.9 mostra um resumo das matrizes de confusão geradas na aplicação dos modelos de classificações e também na aferição de desempenho do mapa LAPIG250.

**Tabela 4.9:** *Desempenho dos modelos de classificação*

| RECORTE | MODELO | SS    | SN   | NN   | NS   | TOTAL | CS     | CN     | OS     | ON     | ET     |
|---------|--------|-------|------|------|------|-------|--------|--------|--------|--------|--------|
| CDG     | NCX    | 6438  | 2053 | 1659 | 522  | 10672 | 24,18% | 7,50%  | 23,93% | 55,31% | 24,13% |
|         | SLMB   | 6025  | 1400 | 2312 | 935  | 10672 | 18,86% | 13,43% | 28,80% | 37,72% | 21,88% |
|         | LAPIG  | 3562  | 270  | 2890 | 2373 | 9095  | 7,05%  | 45,09% | 39,98% | 8,54%  | 29,06% |
| CLV     | NCX    | 2156  | 2050 | 1679 | 315  | 6200  | 48,74% | 12,75% | 15,80% | 54,97% | 38,15% |
|         | SLMB   | 2132  | 2113 | 1616 | 339  | 6200  | 31,37% | 36,71% | 23,13% | 19,17% | 26,16% |
|         | LAPIG  | 1608  | 172  | 2644 | 877  | 5301  | 9,66%  | 24,91% | 35,29% | 6,11%  | 19,79% |
| LUZ     | NCX    | 1846  | 3697 | 590  | 67   | 6200  | 66,70% | 3,50%  | 10,20% | 86,24% | 60,71% |
|         | SLMB   | 1347  | 1509 | 2778 | 566  | 6200  | 52,84% | 29,59% | 16,93% | 35,20% | 33,47% |
|         | LAPIG  | 1062  | 372  | 3123 | 687  | 5244  | 25,94% | 18,03% | 39,28% | 10,64% | 20,19% |
| NCX     | NCX    | 4569  | 451  | 779  | 401  | 6200  | 8,98%  | 8,07%  | 33,98% | 36,67% | 13,74% |
|         | LAPIG  | 3217  | 43   | 973  | 1011 | 5244  | 1,32%  | 50,96% | 23,91% | 4,23%  | 20,10% |
| REC39K  | NCX    | 27618 | 4024 | 5821 | 1537 | 39000 | 12,72% | 5,27%  | 20,89% | 40,87% | 14,26% |
|         | LAPIG  | 21058 | 1292 | 6963 | 4186 | 33499 | 5,78%  | 37,55% | 16,58% | 15,65% | 16,35% |
| SLMB    | SLMB   | 4063  | 609  | 929  | 599  | 6200  | 13,04% | 12,85% | 39,20% | 39,60% | 19,48% |
|         | LAPIG  | 3316  | 304  | 932  | 692  | 5244  | 8,40%  | 42,61% | 17,27% | 24,60% | 18,99% |

Pelos indicadores de erro das matrizes de confusão é possível verificar que a metodologia adotada neste trabalho, quando foi aplicada em áreas com características semelhantes às dos recortes de referência, apresentou resultados compatíveis com os apresentados pelo mapa LAPIG250 para as mesmas áreas.

---

## Conclusões

---

Os principais objetivos deste trabalho foram a apresentação de uma metodologia de classificação de pastagens com baixo custo de recursos computacionais, materiais e humanos, produzindo resultados com qualidade compatível com outras iniciativas que usaram os mesmos dados, mas que aplicaram métodos distintos. As séries temporais de NDVI que foram utilizadas neste trabalho estão disponíveis para *download* de forma gratuita [NASA/EARTHDATA] e os *softwares* utilizados no processamento são todos *Open Source* ([RStudio], [QGIS] e [PostGIS]). A classificação não supervisionada trabalha com alguns algoritmos que são custosos computacionalmente, mas as áreas clusterizadas são pequenas e não impactam de maneira relevante no processo. Já a classificação supervisionada exige apenas a filtragem dos dados antes de serem submetidos ao cálculo das frações de abundância. Considerando que o mapa LAPIG030 foi usado para suprir a função do especialista, é necessária a iteração humana apenas na escolha dos recortes de referência e de classificação, sendo que o restante do processo pode ser implementado de forma automática.

### 5.1 Sobre a metodologia aplicada

Neste trabalho foi comprovado que o sensoriamento remoto é capaz de gerar informações adequadas para a detecção das áreas de pastagens, sendo que as séries temporais de NDVI, obtidas a partir do sensor MODIS, são uma das principais fontes de informações para o monitoramento destas áreas. A existência de um índice, disponibilizado gratuitamente a cada 16 dias, que mede a quantidade de vegetação em cada pixel de uma determinada área, permite a geração de uma assinatura temporal para cada tipo de feição terrestre. Por outro lado, a baixa resolução espacial do produto MOD13Q1 impõe algumas limitações em seu uso na detecção das áreas de pastagens. Estas áreas não são homogêneas e é possível encontrar vários elementos terrestres contidos no espaço de um único pixel, gerando assinaturas distintas para *pixels* vizinhos.

Por causa da mistura de feições em um mesmo pixel, esta metodologia adotou o modelo linear de mistura temporal, que é uma adaptação do modelo linear de mistura

espectral, na qual se usa séries temporais no lugar das bandas espectrais. No caso de uma imagem multiespectral ou hiperespectral os *pixels* com valores espectrais extremos são considerados puros e os demais são representados como uma combinação linear dos primeiros. Nas séries temporais, para um pixel ser considerado puro e escolhido como membro final deve-se levar em conta, além dos valores espectrais extremos de cada ocorrência da série, a frequência e a sequência em que isto acontece.

Segundo [Liu et al. 2019], a análise de mistura temporal depende de uma boa precisão dos membros finais e o algoritmo PPI, um dos mais utilizados na obtenção de membros finais, não foi eficiente no caso das pastagens. Os *pixels* com valores extremos encontrados pelo PPI representavam áreas urbanas, matas densas, áreas agrícolas etc e a combinação linear dos valores destas feições não conseguiu identificar corretamente as áreas de pastagens. Nas regiões classificadas, os valores de NDVI das pastagens mantiveram-se em níveis intermediários durante quase todo o período e apresentaram apenas uma pequena redução na época de seca. A alternativa ao algoritmo PPI foi a classificação não supervisionada com o algoritmo PAM, sendo que os centros dos *clusters*, denominados de medóides, foram escolhidos como membros finais, por representarem as características de grupos de *pixels* de mesma natureza. A combinação linear destes medóides gerou frações de abundância que foram capazes de representar todas as feições da área clusterizada.

Para a construção do modelo de classificação era necessário o conhecimento de um especialista para fazer a rotulação dos dados. O mapa LAPIG030 foi usado para fazer o papel deste especialista e identificar quais *pixels* estavam localizados em áreas de pastagens. A partir da correlação entre as frações de abundância e o percentual de pastagens de cada pixel, foram estabelecidos os parâmetros do modelo que foi utilizado para a classificação de novas áreas. As pastagens mostraram-se um tipo de feição de difícil caracterização tendo em vista que, num mesmo recorte de referência, dois membros finais apresentaram correlação positiva com o percentual de pastagens ao mesmo tempo em que apresentaram correlação negativa entre si.

Após a aplicação dos modelos em outras áreas, as matrizes de confusão foram construídas a partir da comparação dos resultados das classificações com o mapa PASTL8MOD, produto da fusão entre o mapa LAPIG030 e a grade NDVI. É importante ressaltar, mais uma vez, que o mapa LAPIG030 foi construído a partir de uma fonte de dados cujo pixel cobre uma área cerca de 60 vezes menor do que a área do pixel das séries temporais de NDVI. Foi constatado que os modelos são sensíveis às mudanças nas feições terrestres, ou seja, quanto maiores as diferenças entre as características da área sendo classificada e as da área de referência, menos assertivos foram os resultados. Isto ficou evidente quando o modelo gerado a partir do recorte NCX, que contém basicamente áreas de pastagens, foi aplicado aos demais recortes e, em dois deles, o erro total apresentou

valores muito acima do esperado. O recorte CLV possui uma grande área de hidrografia enquanto que o recorte LUZ possui uma grande quantidade de áreas de agricultura irrigada, características que não estão presentes no recorte NCX e que causaram o alto índice de erro na classificação.

Levando em consideração este aspecto da semelhança nas características, foi escolhido um novo recorte, o REC39K, situado próximo ao recorte de referência NCX e com uma área aproximadamente seis vezes maior que os demais recortes. O maior erro cometido pelo classificador usando o modelo NCX no recorte REC39K foi o de omissão na classe "Não pastagens" numa região de agricultura irrigada que não estava representada no recorte NCX, o que comprovou a sensibilidade do modelo às mudanças nas feições. Contudo, esta classificação apresentou o menor erro total entre todas as classificações que foram efetuadas ao longo deste trabalho, ficando em 14,95%.

## 5.2 Comparação com outros trabalhos

O mapa LAPI250 foi gerado num processo de classificação supervisionada cujos dados de treinamento foram obtidos a partir do mapa LAPIG030, sendo que os seus *pixels* foram reamostrados da resolução espacial de 30 m do Landsat 8 para a resolução espacial de 250 m do MODIS e sua acurácia geral foi de cerca de 80% [Parente e Ferreira 2018]. Considerando as semelhanças entre as fontes de dados e os resultados produzidos, foi feita uma comparação entre os erros totais de classificação dos dois produtos, tomando como base as mesmas áreas. O erro total do LAPIG250 ficou em torno dos 20% esperados, variando de 16,35% a 29,06% em todos os recortes. Quando os modelos NCX e SLMB foram aplicados em suas respectivas áreas, os erros totais foram de 13,74% e 19,48% respectivamente; contudo, ao serem aplicados aos demais recortes, o erro total destes modelos chegou a atingir o patamar de 61,13%. Finalmente, na classificação do recorte REC39K usando o modelo NCX, ambos situados em regiões próximas e com características semelhantes, o erro total foi de apenas 14,26%, ligeiramente inferior ao erro aferido na validação do LAPIG250, que foi de 16,35%.

## 5.3 Trabalhos futuros

Nesta metodologia, a rotulagem dos dados usados para a construção do modelo de classificação, bem como sua validação, foram feitas usando um mapa de pastagens existentes, o LAPIG030. Uma alternativa que não foi testada e que poderá ser desenvolvida em trabalhos futuros, parte da premissa de que um especialista faria a rotulagem dos dados e a validação dos resultados.

Inicialmente, esse especialista faria a escolha de uma área a ser classificada, cujos limites seriam definidos em função da homogeneidade de suas feições. Em seguida, seriam selecionadas pequenas áreas, dentro da área escolhida, que fossem representativas da diversidade de seus tipos de cobertura e que passariam por um processo de clusteração conjunta. Os passos seguintes seriam a rotulagem dos medóides, o cálculo das frações de abundância e a definição dos parâmetros de seleção. Este último passo levaria em consideração as frações de abundância relativas aos membros finais rotulados como "Pastagens". Haveria um processo cíclico de fixação dos parâmetros, seleção de *pixels* e validação, até que houvesse equilíbrio entre os erros de comissão, omissão e erro total. O último passo seria a aplicação dos parâmetros finais na área escolhida para ser classificada.

---

## Referências Bibliográficas

---

- [Abdelli et al. 2013]ABDELLI, A. et al. A classification framework of issuers in the moroccan financial market. In: *2013 8th International Conference on Intelligent Systems: Theories and Applications (SITA)*. [S.l.: s.n.], 2013. p. 1–5.
- [Abdou et al. 1996]ABDOU, B. et al. A review of vegetation indices. *Remote Sensing Reviews*, v. 13, p. 95–120, 01 1996.
- [Aghabozorgi, Shirkhorshidi e Teh 2015]AGHABOZORGI, S. R.; SHIRKHORSHIDI, A. S.; TEH, Y. W. Time-series clustering - a decade review. *Inf. Syst.*, v. 53, p. 16–38, 2015.
- [Alexandratos e Bruinsma 2012]ALEXANDRATOS, N.; BRUINSMA, J. *World agriculture towards 2030/2050: the 2012 revision*. [S.l.], 2012.
- [ALMEIDA et al. 2016]ALMEIDA, C. A. d. et al. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazonica*, scielo, v. 46, p. 291 – 302, 09 2016. ISSN 0044-5967.
- [Batista et al. 2013]BATISTA, G. et al. Cid: An efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery*, v. 28, 04 2013.
- [Boardman J. W. e Green 1995]BOARDMAN J. W., K. F. A.; GREEN, R. O. Mapping target signatures via partial unmixing of aviris data. In: *Fifth JPL Airborne Earth Science Workshop*. [S.l.: s.n.], 1995. v. 1, p. 23–26.
- [Breiman e Cutler 2001]BREIMAN, L.; CUTLER, A. Random forests. *Machine Learning*, p. 5–32, 2001.
- [Chang et al. 2008]CHANG, C.-I. et al. A pyramid-based block of skewers for pixel purity index for endmember extraction in hyperspectral imagery. *International Journal of High Speed Electronics and Systems*, v. 18, p. 469–482, 06 2008.
- [Chi, Kim e Kang 2016]CHI, J.; KIM, H.-C.; KANG, S.-H. Machine learning based temporal mixture analysis of hypertemporal antarctic sea ice data. *Remote Sensing Letters*, v. 7, n. 2, p. 190–199, 2016.

- [Didan 2015]DIDAN, K. *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006*. NASA EOSDIS Land Processes DAAC. 2015. Disponível em: <<https://doi.org/10.5067/MODIS/MOD13Q1.006>>. Acesso em: 29/dez/2018.
- [FAO 2018]FAO. *World food and agriculture - Statistical pocketbook 2018*. [S.l.]: FAO - Food and Agriculture Organization, 2018. ISBN 978-92-5-131012-0.
- [Ferreira 2003]FERREIRA, M. E. *Análise do Modelo Linear de Mistura Espectral na Discriminação de Fitofisionomias do Parque Nacional de Brasília (Bioma Cerrado)*. 143 p. Dissertação (Mestrado) — Universidade de Brasília, Brasília - DF, 2003.
- [Foerster et al. 2014]FOERSTER, M. et al. Timespec - a software tool for analyzing time-series of spectral data. In: *2014 IEEE Geoscience and Remote Sensing Symposium*. [S.l.: s.n.], 2014. p. 3941–3944. ISSN 2153-6996.
- [Gandhi e Srivastava 2014]GANDHI, G.; SRIVASTAVA, R. Review paper: A comparative study on partitioning techniques of clustering algorithms. *International Journal of Computer Applications*, v. 87, 01 2014.
- [Giglio L.]GIGLIO L., J. C. B. L. R. D. *Collection 6 MODIS Burned Area Product User's Guide*. Disponível em: <<https://modis-land.gsfc.nasa.gov>>. Acesso em: 25/nov/2019.
- [Gonçalves et al. 2017]GONÇALVES, R. R. do V. et al. Agricultural monitoring using clustering techniques on satellite image time series of low spatial resolution. In: *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)*. [S.l.: s.n.], 2017. p. 1–4.
- [Gonçalves et al. 2014]GONÇALVES, R. R. V. et al. Land use temporal analysis through clustering techniques on satellite image time series. In: *2014 IEEE Geoscience and Remote Sensing Symposium*. [S.l.: s.n.], 2014. p. 2173–2176. ISSN 2153-6996.
- [Guyet e Nicolas 2016]GUYET, T.; NICOLAS, H. Long term analysis of time series of satellite images. *Pattern Recognition Letters*, v. 70, p. 17 – 23, 2016. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865515003918>>.
- [Heinz e Chein-I-Chang 2001]Heinz, D. C.; Chein-I-Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, v. 39, n. 3, p. 529–545, March 2001. ISSN 1558-0644.
- [Hird e McDermid 2009]HIRD, J. N.; MCDERMID, G. J. Noise reduction of ndvi time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, v. 113, n. 1, p. 248 – 258, 2009. ISSN 0034-4257. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0034425708002745>>.



- [Hitchcock, Booth e Casella 2007]HITCHCOCK, D. B.; BOOTH, J. G.; CASELLA, G. The effect of pre-smoothing functional data on cluster analysis. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 77, n. 12, p. 1043–1055, 2007. Disponível em: <<https://doi.org/10.1080/10629360600880684>>.
- [Huy e Anh 2016]HUY, V. T.; ANH, D. T. An efficient implementation of anytime k-medoids clustering for time series under dynamic time warping. In: *Proceedings of the Seventh Symposium on Information and Communication Technology*. New York, NY, USA: ACM, 2016. (SoICT '16), p. 22–29. ISBN 978-1-4503-4815-7. Disponível em: <<http://doi.acm.org/10.1145/3011077.3011128>>.
- [Kaufmann e Rousseeuw 1987]KAUFMANN, L.; ROUSSEEUW, P. Clustering by means of medoids. *Data Analysis based on the L1-Norm and Related Methods*, p. 405–416, 01 1987.
- [Keshava 2003]KESHAVA, N. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, v. 14, p. 55–78, 01 2003.
- [LAPIG-UFG 2017]LAPIG-UFG. *Mapa de Pastagens Brasileiras*. 2017. Disponível em: <<http://maps.lapig.iesa.ufg.br/lapig.html>>. Acesso em: 06/jan/2019.
- [Lawson e Hanson 1995]LAWSON, C. L.; HANSON, R. J. *Solving least squares problems*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1995. xii+337 p. (Classics in Applied Mathematics, v. 15). ISBN 0-89871-356-0.
- [Leinenkugel et al. 2013]LEINENKUGEL, P. et al. Characterisation of land surface phenology and land cover based on moderate resolution satellite data in cloud prone areas - a novel product for the mekong basin. *Remote Sensing of Environment*, v. 136, p. 180 – 198, 2013. ISSN 0034-4257. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S003442571300148X>>.
- [Liu et al. 2019]LIU, Z. et al. Regional scale mapping of fractional rice cropping change using a phenology-based temporal mixture analysis. *International Journal of Remote Sensing*, v. 40, n. 7, p. 2703–2716, 2019.
- [Mercado et al. 2016]MERCADO, J. M. et al. A method based on satellite imagery to identify spatial units for eutrophication management. *Remote Sensing of Environment*, v. 186, p. 123 – 134, 2016. ISSN 0034-4257. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0034425716303145>>.
- [Mullen e Stokkum 2012]MULLEN, K. M.; STOKKUM, I. H. M. van. *The Lawson-Hanson algorithm for non-negative least squares*. [S.l.], 2012. Disponível em: <<https://cran.r-project.org/web/packages/nnls/nnls.pdf>>.

- [NASA/EARTHDATA]NASA/EARTHDATA. *Search Earth Data*. Disponível em: <<https://search.earthdata.nasa.gov/search>>. Acesso em: 06/jan/2019.
- [NASA/Landsat8]NASA/LANDSAT8. *Landsat 8*. Disponível em: <<https://eospso.nasa.gov/missions/landsat-8>>. Acesso em: 29/dez/2018.
- [NASA/MODIS]NASA/MODIS. *MODIS - Moderate Resolution Imaging Spectroradiometer*. Disponível em: <<https://modis.gsfc.nasa.gov/data/>>. Acesso em: 26/dez/2018.
- [NASA/NDVI]NASA/NDVI. *Normalized Difference Vegetation Index (NDVI)*. Disponível em: <<https://earthobservatory.nasa.gov/features/MeasuringVegetation>>. Acesso em: 26/dez/2018.
- [Parente e Ferreira 2018]PARENTE, L.; FERREIRA, L. Assessing the spatial and occupation dynamics of the brazilian pasturelands based on the automated classification of modis images from 2000 to 2016. *Remote Sensing*, v. 10, n. 4, 2018. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/10/4/606>>.
- [Parente et al. 2017]Parente, L. et al. Monitoring the brazilian pasturelands: A new mapping approach based on the landsat 8 spectral and temporal domains. *International Journal of Applied Earth Observation and Geoinformation*, v. 62, p. 135–143, out. 2017.
- [Pascucci et al. 2018]PASCUCCI, S. et al. A comparison between standard and functional clustering methodologies: Application to agricultural fields for yield pattern assessment. *Remote Sensing*, v. 10, n. 4, 2018. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/10/4/585>>.
- [Pereira et al. 2018]PEREIRA, O. J. R. et al. Assessing pasture degradation in the brazilian cerrado based on the analysis of modis ndvi time-series. *Remote Sensing*, v. 10, n. 11, 2018. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/10/11/1761>>.
- [PostGIS]POSTGIS. *The PostgreSQL Global Development Group*. Disponível em: <<https://postgis.net/>>. Acesso em: 26/dez/2018.
- [QGIS]QGIS. *QGIS Association*. Disponível em: <[https://www.qgis.org/pt\\_BR/site/](https://www.qgis.org/pt_BR/site/)>. Acesso em: 26/dez/2018.
- [RStudio]RSTUDIO. *RStudio, Inc*. Disponível em: <<https://www.rstudio.com/>>. Acesso em: 26/dez/2018.
- [Savitzky e Golay 1964]SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, v. 36, p. 1627–1639, 1964.

- [Senin 2009]SENIN, P. Dynamic time warping algorithm review. 01 2009.
- [Shimabukuro e Ponzoni 2019]SHIMABUKURO, Y. E.; PONZONI, F. J. *Spectral Mixture for Remote Sensing: Linear Model and Applications*. Switzerland: Springer Nature Switzerland AG, 2019. ISBN 978-3-030-02017-0.
- [Spruce et al. 2018]SPRUCE, J. et al. Developing land use land cover maps for the lower mekong basin to aid hydrologic modeling and basin planning. *Remote Sensing*, v. 10, n. 12, 2018. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/10/12/1910>>.
- [Theodoridis e Koutroumbas 2003]THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern recognition*. [S.l.]: Academic Press, 2003.
- [Zhuo et al. 2018]ZHUO, L. et al. An improved temporal mixture analysis unmixing method for estimating impervious surface area based on modis and dmsp-ols data. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 142, p. 64 – 77, 2018. ISSN 0924-2716. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0924271618301497>>.