

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ – CAMPUS DE CASCAVEL
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

**MAPEAMENTO SEMIAUTOMÁTICO POR MEIO DE PADRÃO ESPECTRO-TEMPORAL DE
ÁREAS AGRÍCOLAS E ALVOS PERMANENTES COM EVI/MODIS NO PARANÁ**

WEVERTON RODRIGO VERICA

Cascavel – Paraná – Brasil
FEVEREIRO – 2018

WEVERTON RODRIGO VERICA

**MAPEAMENTO SEMIAUTOMÁTICO POR MEIO DE PADRÃO ESPECTRO-TEMPORAL DE
ÁREAS AGRÍCOLAS E ALVOS PERMANENTES COM EVI/MODIS NO PARANÁ**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola para obtenção do título de mestre em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, com a temática Geoprocessamento, Estatística Espacial e Agricultura de Precisão.

Orientador: Dr. Jerry Adriani Johann

**Cascavel – Paraná – Brasil
FEVEREIRO - 2018**

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Verica, Weverton Rodrigo

Mapeamento semiautomático por meio de padrão espectro-temporal de áreas agrícolas e alvos permanentes com EVI/MODIS no Paraná / Weverton Rodrigo Verica; orientador(a), Jerry Adriani Johann, 2018.

116 f.

Dissertação (mestrado), Universidade Estadual do Oeste do Paraná, Campus de Cascavel, Centro de Ciências Exatas e Tecnológicas, Programa de Pós-Graduação em Engenharia Agrícola, 2018.

1. KDD. 2. Random forest. 3. Classificação. I. Johann, Jerry Adriani. II. Título.

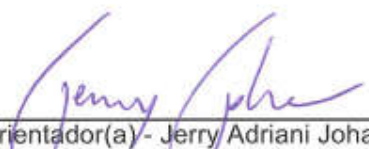
Revisora de Português, inglês e normas: Dhandara Capitani

Data da revisão 08/maio/2018

WEVERTON RODRIGO VERICA

Mapeamento semiautomático por meio de padrão espectro-temporal de áreas agrícolas e alvos permanentes com EVI/MODIS no Paraná

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO(A) pela seguinte banca examinadora:


Orientador(a) - Jerry Adriani Johann

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)


Carlos Antonio da Silva Júnior

Universidade do Estado de Mato Grosso (UNEMAT)


Erivelto Mercante

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)


Flávio Gurgacz

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Cascavel, 16 de fevereiro de 2017

BIOGRAFIA RESUMIDA

Brasileiro, nasceu em Cafelândia, PR no ano de 1993. Graduado em Matemática pela Universidade Estadual do Oeste do Paraná – UNIOESTE, campus de Cascavel (2015). Aluno do Programa de Pós-Graduação em Engenharia Agrícola – nível mestrado (2016), na área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão.

AGRADECIMENTO

Primeiramente a Deus, pela vida;

À Universidade Estadual do Oeste do Paraná – UNIOESTE e ao Programa de Pós-Graduação em Engenharia Agrícola – PGEAGRI, pela oportunidade oferecida em fazer o mestrado;

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro;

Ao meu orientador Jerry Adriani Johann, pelo apoio, a dedicação e o companheirismo;

Aos colegas e amigos do Laboratório de Estatística Aplicada – LEA. Em especial ao Alex Paludo, Jonathan Richetti, Kleber R. Nascimento, Laíza Cavalcante, Leticia Dal' Canton, Tamara Cantú Maltauro e Willyan Becker;

A minha família, em especial a minha irmã Eliane Cristina Verica, meu primo Anderson Rodrigo Pereira e principalmente a meus pais, Almir Verica e Euza de Fátima Pereira Verica, que sempre me incentivaram a estudar e me apoiaram;

Aos colegas de pós-graduação, pela amizade, o convívio e o apoio;

Aos amigos, pela compreensão e pelos momentos de felicidade;

E a todos que de alguma forma contribuíram na realização deste trabalho.

Obrigado!

MAPEAMENTO SEMIAUTOMÁTICO DE ÁREAS AGRÍCOLAS E ALVOS PERMANENTES POR MEIO DE PADRÃO ESPECTRO-TEMPORAL DE EVI/MODIS NO PARANÁ

RESUMO

O conhecimento da localização e da quantidade de áreas destinadas a agricultura ou a florestas nativas ou plantadas é relevante para que os gestores públicos tomem suas decisões pautadas em dados fidedignos com a realidade. Além disso, parte das receitas de ICMS advindas do Fundo de Participação aos Municípios (FPM) depende de dados de produção agropecuária, número de propriedades rurais e fator ambiental. Diante disso, esta dissertação teve como objetivo elaborar uma metodologia objetiva e semiautomática para mapear áreas agrícolas e alvos permanente e posteriormente identificar áreas de soja, milho 1ª e 2ª safras, culturas de inverno, agricultura semi-perene, florestas e demais alvos permanentes no estado do Paraná para os anos-safra (2013/14 a 2016/17), utilizando séries temporais de índices de vegetação EVI/Modis. A metodologia proposta segue os passos do Processo de descoberta de conhecimento em base de dados – KDD, sendo que para isso foram elaboradas métricas extraídas do perfil espectro temporal de cada pixel e foi empregada a tarefa de classificação, realizada pelo algoritmo Random Forest. Para a validação dos mapeamentos utilizaram-se amostras extraídas de imagens Landsat-8, obtendo-se os índices de exatidão global maior que 84,37% e um índice kappa variando entre 0,63 e 0,98, sendo, portanto, considerados mapeamentos com boa ou excelente acurácia espacial. Os dados municipais da área de soja, milho 1ª safra, milho 2ª safra e culturas de inverno mapeada foram confrontados com as estatísticas oficiais obtendo-se coeficientes de correlação linear entre 0,61 a 0,9, indicando moderada ou forte correlação com os dados oficiais. Desse modo, a metodologia semiautomática proposta obteve êxito na realização do mapeamento, bem como a automatização do processo de elaboração das métricas, gerando, com isso um script no software R de maneira a facilitar mapeamentos futuros com baixo tempo de processamento.

PALAVRAS-CHAVE: KDD, Random Forest, Classificação.

SEMI-AUTOMATIC MAPPING OF AGRICULTURAL AREAS AND TARGETS PERMANENT BY PROFILE SPECTRUM-TEMPORARY OF EVI / MODIS IN PARANA

ABSTRACT

Knowledge of location and quantity of areas for agriculture or either native or planted forests is relevant for public managers to make their decisions based on reliable data. In addition, part of ICMS revenues from the Municipal Participation Fund (FPM) depends on agricultural production data, number of rural properties and the environmental factor. The objective of this research was to design an objective and semiautomatic methodology to map agricultural areas and targets permanent, and later to identify areas of soybean, corn 1st and 2nd crops, winter crops, semi-perennial agriculture, forests and other permanent targets in the state of Paraná for the harvest years (2013/14 to 2016/17), using temporal series of EVI/Modis vegetation indexes. The proposed methodology follows the steps of the Knowledge Discovery Process in Database – KDD, in which the classification task was performed by the Random Forest algorithm. For the validation of the mappings, samples extracted from Landsat-8 images were used, obtaining the global accuracy indices greater than 84.37% and a kappa index ranging from 0.63 to 0.98, hence considered mappings with good or excellent spatial accuracy. The municipal data of the area of soybean, corn 1st crop, corn 2nd crop and winter crops mapped were confronted with the official statistics obtaining coefficients of linear correlation between 0.61 to 0.9, indicating moderate or strong correlation with the data officials. In this way, the proposed semi-automatic methodology was successful in the mapping, as well as the automation of the process of elaboration of the metrics, thus generating a script in the software R in order to facilitate future mappings with low processing time.

KEY-WORDS: KDD, Random forest, Classification.

SUMÁRIO

RESUMO	vii
ABSTRACT	viii
LISTA DE FIGURAS	xi
LISTA DE TABELAS	xiv
1. INTRODUÇÃO	15
2. OBJETIVOS	17
3. REVISÃO BIBLIOGRÁFICA	18
3.1 Caracterização do estado do Paraná	18
3.1.1 Solo.....	18
3.1.2 Relevo.....	19
3.1.3 Clima.....	19
3.1.4 Economia do estado e de seus municípios.....	21
3.2 Uso e ocupação do solo no Paraná	24
3.2.1 Alvos permanentes.....	24
3.2.2 Culturas Perenes.....	24
3.2.3 Culturas semi-perenes.....	24
3.2.3.1 Cana-de-açúcar.....	24
3.2.3.2 Pastagem.....	25
3.2.4 Culturas anuais.....	26
3.2.4.1 Soja.....	26
3.2.4.2 Milho.....	27
3.2.4.3 Cultura de inverno.....	29
3.2.4.4 Feijão.....	31
3.3 Sensoriamento remoto	33
3.3.1 Sensoriamento remoto aplicado a agricultura.....	33
3.3.2 Índice de vegetação.....	34
3.3.3 Sensores.....	35
3.3.3.1 Modis.....	35
3.3.3.2 Landsat-8.....	36
3.3.4 Séries temporais.....	37
3.3.5 Mapeamento.....	40
3.4 KDD	41
3.4.1 Seleção de dados.....	42
3.4.2 Pré-processamento.....	42
3.4.3 Formatação.....	42
3.4.4 Mineração de dados.....	43

3.4.5	Avaliação e Interpretação.....	44
4.	METODOLOGIA.....	45
4.1	Mineração de dados	48
5.	REFERÊNCIAS	51
6.	ARTIGOS.....	56
6.1	Mapeamento semiautomático de áreas agrícolas e permanentes com métricas do perfil espectro-temporal de EVI/MODIS	56
6.1.1	Introdução.....	56
6.1.2	Material e métodos.....	58
6.1.2.1	Área de estudo	58
6.1.2.2	Processo de Descoberta de conhecimento em base de dados.....	58
6.1.3	Resultados	67
6.1.3.1	Método de verificação do conjunto de dados para os mapeamentos.....	67
6.1.3.2	Mapeamentos dos alvos para os anos-safra 2013/14 a 2016/17	67
6.1.3.3	Mapeamento único dos anos-safra.....	71
6.1.4	Conclusões	75
6.1.5	Referências.....	75
6.2	Uso da dinâmica espectro-temporal de EVI/MODIS para mapeamento semiautomático de florestas e culturas semi-perenes e anuais no Paraná.....	78
6.2.1	Introdução.....	78
6.2.2	Material e métodos.....	80
6.2.2.1	Área de estudo	80
6.2.2.2	Processo de descoberto de conhecimento em base de dados – KDD	81
6.2.3	Resultados	88
6.2.3.1	Validação dos mapeamentos com os dados dos conjuntos de amostras de pixels puros	88
6.2.3.2	Comparação com dados oficiais.....	92
6.2.4	Conclusões	94
6.2.5	Referências.....	95
7	CONSIDERAÇÕES FINAIS	98

LISTA DE FIGURAS

Figura 1 Mapa dos Solos do estado do Paraná	18
Figura 2 Relevo do estado do Paraná.....	19
Figura 3 Temperatura média anual no estado do Paraná.	20
Figura 4 Precipitação anual no estado do Paraná.....	20
Figura 5 Classificação do clima no estado do Paraná.....	21
Figura 6 Produção média de cana-de-açúcar nos municípios paranaenses entre os anos de 2012 e 2016.....	25
Figura 7 Produção média de soja nos municípios paranaenses entre os anos de 2012 e 2016.	27
Figura 8 Média da produção de milho (2012 a 2016) [a] primeira safra; b) segunda safra] nos municípios paranaenses.	28
Figura 9 Produção anual média (2012 a 2016) de: a) trigo; e b) culturas de outono-inverno, nos municípios paranaenses.	30
Figura 10 Produção média (2012 a 2016) de feijão [a] primeira safra; b) segunda safra, c) terceira safra] nos municípios paranaenses.....	33
Figura 11 Padrão espectro-temporal do índice de vegetação EVI de uma cultura agrícola compreendendo o período do dia juliano (c) 225 ao 113, com as datas DS (data de semeadura), DMDV (data de máximo desenvolvimento vegetativo) e DC (data de colheita).....	38
Figura 12 Perfil temporal do índice de vegetação NDVI de pixels de cana-de-açúcar, soja, mata, cidade e pastagem.....	39
Figura 13 Processo Descoberta de Conhecimento em Base de Dados (KDD).	41
Figura 14 Fluxograma para a elaboração do mapeamento de um ano-safra seguindo o processo KDD	45
Figura 15 Representação esquemática das métricas obtidas de uma série espectro-temporal de índice e vegetação (a) com ênfase na curva referente a taxa de variação de uma série temporal (b).....	46
Figura 16 Perfil espectro-temporal médio de cada alvo.	47
Figura 17 Método manual de seleção de alvos de culturas agrícolas e alvos permanentes na composição falsa cor RGB564 de Landsat-8 com o detalhamento da malha Modis. ...	49
Figura 18 Localização do Estado do Paraná com as 10 mesorregiões e path-row do satélite Landsat-8.....	58
Figura 19 Fluxograma para a separação de alvos agrícolas e alvos permanentes, com as divisões do processo KDD.....	59
Figura 20 Representação do cubo de imagem (a) utilizada para geração do padrão espectro-temporal do índice de vegetação (b).	60
Figura 21 Forma de extração das métricas de uma série temporal.....	61
Figura 22 Método de seleção de alvos de culturas agrícolas e alvos permanentes na composição RGB564 com o detalhamento da malha MODIS.....	62
Figura 23 Fluxograma para reclassificação das classes com erro de omissão superior a 40%. a) processo de segmentação em cinco classes; b) Verificação dos alvos de cada classe; c) Processo de decisão de quais classes seriam reclassificadas e qual classe seria utilizada como base para elaborar o intervalo interquartilício; d) comparação dos	

pixels das classes reclassificadas com o intervalo interquartilico e definição da nova classe.	66
Figura 24 Mapeamento de áreas agrícolas e alvos permanentes para os anos-safra 2013/14 (a), 2014/15 (b), 2015/16 (c) e 2016/17 (d).	68
Figura 25 Erro de omissão (EO) e Erro de comissão (EC) para alvo agrícola (Agri) e alvos permanentes (Perm) nos anos-safra 2013/14 a 2016/17.	69
Figura 26 Gráficos de dispersão entre dados oficiais de área agrícola de cada município e os mapeamentos realizados para os anos-safra 2013/2014 (a), 2014/2015 (b) e 2015/2016 (c).	70
Figura 27 Boxplot das diferenças dos dados oficiais e o mapeamento para os anos-safra 2013/14, 2014/15 e 2015/16.	70
Figura 28 Área total agrícola (ha) do mapeamento elaborado e dos dados oficiais para o Paraná nos anos-safra 2013/14 a 2015/16.	71
Figura 29 Mapeamento de áreas agrícolas e alvos permanentes do estado do Paraná entre os anos-safra 2013/14 e 2016/17 (a) com ênfase nas áreas com maior concentração da classe 3: Parque nacional de Ilha Grande (b); e litoral paranaense (c).	72
Figura 30 Exemplo da classificação de cada uma das classes conforme composição colorida RGB-564 para cenas Landsta-8/OLI. a) Imagem RGB-564 original; b) Imagem RGB-564 com destaque para os pixels classificados na classe 1; c) Imagem RGB-564 com destaque para os pixels classificados na classe 2; d) Imagem RGB-564 com destaque para os pixels classificados na classe 3; e) Imagem RGB-564 com destaque para os pixels classificados na classe 4; f) Imagem RGB-564 com destaque para os pixels classificados na classe 5.	73
Figura 31 Mapeamento final de áreas agrícolas e alvos permanentes no Paraná para o período de 2013 a 2017.	74
Figura 32 Localização do Estado do Paraná com as 10 mesorregiões, 399 municípios e <i>tiles</i> do satélite Landsat-8.	80
Figura 33 Fluxograma para a separação floresta, demais alvos permanentes, cultura semi-perene, soja/feijão, milho 1ª e 2ª safra, culturas de inverno e outros, com as divisões do processo KDD.	81
Figura 34 Representação esquemática das métricas obtidas de uma série espectro-temporal de índice e vegetação (a) com ênfase na curva referente a taxa de variação de uma série temporal (b).	83
Figura 35 Método de Seleção de alvos de culturas agrícolas e alvos permanentes na composição falsa cor RGB564 com o detalhamento da malha Modis.	85
Figura 36 Procedimento de classificação com algoritmo Random Forest dos pixels EVI/Modis no Paraná em alvos permanentes (C1), alvos com culturas agrícolas anuais ou semi-perenes (C2) e culturas anuais de primeira safra (C3) e segunda safra (C4).	86
Figura 37 Mapeamento de alvos permanentes, floresta, culturas semi-perenes soja/feijão, milho 1ª e 2ª safra, culturas de inverno e outros, para os anos-safra 2013/14 a 2016/17, sendo que para 2013/14 tem-se o mapeamento 1ª safra (a) e 2ª safra (b), para o ano-safra 2014/15 tem-se 1ª safra (c) 2ª safra (d) no ano-safra 2015/16 tem-se o mapeamento da 1ª safra (e) e 2ª safra (f) e por fim o ano-safra 2016/17 possui apenas mapeamento da a 1ª safra (g).	90
Figura 38 Erro de Omissão (EO) e Erro de comissão (EC) para os alvos floresta/reflorestamento (flo), demais alvos permanente (dap), agricultura semi-perene (asp), soja/feijão (s/f), milho 1ª safra (mil1), outros (out), milho 2ª safra (mil2) e culturas de inverno (ci) para os anos-safra 2013/14 (a); 2014/15 (b); 2015/16 (c) e 2016/17(d).	91

Figura 39 Comparação das áreas (ha) do mapeamento com os dados oficiais para os alvos soja/feijão (s/f), milho 1ª safra (mil1), milho 2ª safra (mil2) e culturas de inverno (ci).	92
Figura 40 Gráficos de dispersão para os alvos agrícolas (Soja/feijão, milho 1ª e 2ª safra e culturas de inverno) para os anos-safra 2013/14 a 2015/16.	93
Figura 41 Tela inicial do software Cymp	102
Figura 42 Caminho para utiliza o filtro Savitzck-Golay no Cymp	102
Figura 43 Tela de seleção dos dados de entrada e dados de saída	103
Figura 44 Configurações do filtro	103
Figura 45 Informações do cubo de imagem.	104
Figura 46 Método de extração dos perfil espectro-temporal de cada um dos pixels do cubo de imagens	105
Figura 47 Plot de cada uma das 15 métricas	112
Figura 48 Conjunto de treinamento	113

LISTA DE TABELAS

Tabela 1 Critérios de Participação dos municípios na Cota-parte do ICMS	22
Tabela 2 Porcentagem mensal de semeadura e colheita da soja no estado do Paraná.....	27
Tabela 3 Porcentagem mensal de semeadura e colheita do milho no estado do Paraná para as duas safras	29
Tabela 4 Porcentagem mensal de semeadura e colheita do trigo no estado do Paraná	31
Tabela 5 Porcentagem mensal de semeadura e colheita do feijão no estado do Paraná para as três safras	32
Tabela 6 Características do sensor Modis abordo dos satélites Terra e Aqua	36
Tabela 7 Características do sensor OLI abordo dos satélites Landsat-8.....	37
Tabela 8 Métricas extraídas de uma série espectro-temporal de índice de vegetação (EVI)...	47
Tabela 9 Métricas extraídas de uma série espectro-temporal de índice de vegetação (EVI)...	61
Tabela 11 Matriz de confusão das áreas agrícolas e dos alvos permanentes.....	74
Tabela 12 Métricas extraídas de uma série espectro-temporal de índice de vegetação (EVI)	84
Tabela 13 Valores de Exatidão global (EG) e índice Kappa (IK) para os quatro anos-safra....	91
Tabela 14 Média dos Indicadores estatísticos: Erro médio (ME), Erro médio quadrático (RMSE) Índice de concordância aprimorado de Willmott (dr) e coeficiente de correlação de Pearson (r) para soja/feijão, milho 1ª safra, milho 2ª safra e culturas de inverno.....	94

1 INTRODUÇÃO

Historicamente a agricultura desempenha um papel importante na economia brasileira devido a sua participação nos mercados externo e interno. Além disso, o país também se destaca na pecuária, sendo o maior exportador de carne bovina do mundo (PORTAL BRASIL, 2016).

Essa importância do setor agropecuário no Brasil pode ser identificada analisando a participação do PIB agropecuário no PIB total do país; segundo Portal Brasil (2016), essa porcentagem foi de 23% em 2016.

O setor agropecuário do estado do Paraná também tem grande importância para sua economia, sendo que a porcentagem do PIB agropecuário corresponde a 10,4%, de acordo com o Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES, 2018). O estado também ocupa o segundo lugar em produção de soja e milho, com, respectivamente, 17,68 e 21,65% da produção total do ano de 2016, ficando atrás apenas do estado do Mato Grosso para ambas as culturas (IBGE, 2017).

Diante disto, é fundamental conhecer o uso e a ocupação do solo paranaense, distinguindo as áreas destinadas a agricultura (culturas semi-perenes e anuais) de áreas permanentes (floresta, reflorestamento, corpos hídricos, alvos urbanos). Estes dados, fidedignos com a realidade, passam a ter maior importância aos municípios quando se avalia o cenário de repasse de ICMS, que é realizado por meio do Fundo de Participação aos Municípios (FPM). Parte da receita anual, obtida pelo FPM, de cada município deriva de dados de produção agropecuária (8%), número de propriedades rurais (6%) e fator ambiental (5%) – conhecido como ICMS Ecológico (PARANÁ, 2017). Além disto, o mercado de commodities agrícolas e os governos se utilizam de dados da produção agrícola para tomadas de decisão referentes a época de semeadura e colheita (JOHANN et al., 2016), armazenamento de grãos, escoamento das safras e oferta dos produtos agrícolas.

Atualmente, as estimativas de área plantada são realizadas anualmente pela Secretaria da Agricultura e do Abastecimento (SEAB) e pela Companhia Nacional de Abastecimento (CONAB).

Entretanto, com o desenvolvimento de técnicas de sensoriamento remoto, a elaboração de mapeamentos vem se aprimorando tanto na qualidade quanto no tempo de elaboração. Essa melhora ocorre devido à captação de imagem orbitais com melhores resolução espacial, temporal e espectral, além de softwares gratuitos que permitem processar essas imagens de maneira rápida e eficaz.

Embora as técnicas de sensoriamento remoto tenham facilitado a execução de mapeamentos, com elas também surgiram novos desafios relacionados à análise das imagens orbitais, como a distinção de alvos terrestres similares (ZHONG et al., 2014;

CECHIM JUNIOR et al., 2017; SON et al., 2013; MENGUE; FONTANA, 2015) ou o tempo e o custo do processamento dos dados (ZHANG et al., 2017; ZHONG et al., 2016).

Para auxiliar na distinção de alvos similares é recomendável a utilização de série temporal, pois as culturas agrícolas são caracterizadas por apresentar alta dinâmica temporal, sendo recomendável um acompanhamento dos alvos por meio de séries temporais de imagens de satélite.

Outro motivo relevante para utilizar séries temporais em imagens orbitais é que a partir delas é possível extrair métricas fenológicas dos alvos, sendo que estas podem revelar características de grande importância para a identificação de culturas agrícolas no campo. Entretanto, com a utilização de séries temporais a dificuldade atrelada ao tempo necessário para o processamento dos dados é aumentada. Uma solução para contornar isto é por meio da automatização dos processos, de modo que este possa ser realizado várias vezes ao longo dos anos, tornando-o mais dinâmico (ZHONG et al., 2016).

Mesmo com a automatização dos processos, o uso de séries temporais para grandes áreas gera uma enorme quantidade de dados, tornando imprescindível a utilização de técnicas e ferramentas computacionais para obter informações relevantes e em menor tempo. Dessa forma, utilizou-se o processo de Descoberta de Conhecimento em Base de Dados ou “Knowledge Discovery in Databases – KDD”, para a descoberta de padrões válidos, úteis, novos e acessíveis em grandes bases de dados. O objetivo do processo KDD foi extrair conhecimento a partir de um banco de dados de imagens de satélite, ou seja, a partir de séries espectro-temporais do índice de vegetação Enhanced Vegetation Index (EVI) do sensor Moderate Resolution Imaging Spectroradiometer (MODIS), levantar o uso e a ocupação do solo no estado do Paraná.

2 OBJETIVOS

O objetivo geral foi elaborar uma metodologia semiautomática para criar o mapa de ocupação do solo do estado do Paraná por meio de séries espectro-temporais do índice de vegetação EVI/Modis, para os anos-safra 2013/14 a 2016/17.

Como objetivos específicos elencaram-se:

- Elaborar métricas baseadas no perfil espectro-temporal dos pixels de imagens EVI/MODIS para realizar o mapeamento de ocupação do solo do Paraná;
- Desenvolver rotinas no software R para automatizar este processo de mapeamento;
- Mapear e estimar a área destinada a agricultura e alvos permanentes no estado;
- Realizar no mapeamento de agricultura a separação entre agricultura semi-perene e culturas anuais. Posteriormente, identificar quais eram as culturas dentro do mapeamento de culturas anuais (soja/feijão, milho e culturas de inverno);
- Estimar as áreas mapeadas com culturas anuais (soja, milho e culturas de inverno) e compará-las com os dados oficiais;
- Realizar no mapeamento de alvos permanentes a separação em florestas/reflorestamento e demais alvos permanentes no estado (corpos hídricos, alvos urbanos e demais áreas não utilizadas para agricultura).

3 REVISÃO BIBLIOGRÁFICA

3.1 Caracterização do estado do Paraná

O estado do Paraná está localizado entre as coordenadas 22°30'58" e 26°43'00" de latitude Sul e 48°05'37" e 54°37'08" de longitude Oeste, encontra-se no Planalto Meridional e na Região Sul do Brasil, sendo subdividido em 10 mesorregiões.

Segundo o IPARDES (2018), o estado do Paraná ocupa uma área de 199.880 km², com densidade demográfica (medida em 2016) de 56,25 habitantes por Km², distribuídos em 399 municípios.

Na sequência serão apresentadas as características do estado em relação a solo, relevo, clima, economia do estado e de seus municípios.

3.1.1 Solo

De acordo com ITCG (2006) o Paraná possui 10 categorias de solos (Figura 1), sendo que as principais são Latossolo com 29,65% da área do Paraná, Neossolo com 24,91 e Nitossolo com 15,61%.

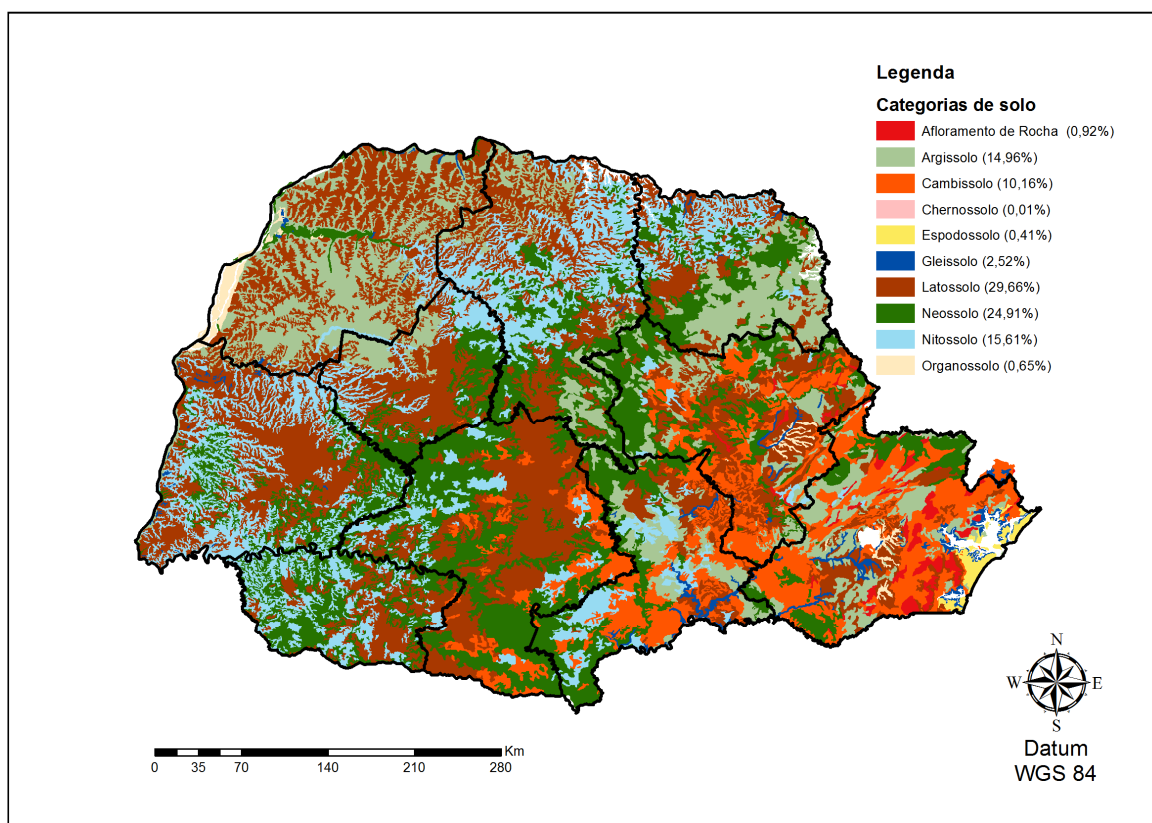


Figura 1 Mapa dos Solos do estado do Paraná
Fonte: ITCG (2006)

3.1.2 Relevo

O Estado do Paraná apresenta uma grande variedade nas formas de relevo, sendo que o estado possui a Planície Litorânea a Serra do Mar e três grandes planaltos conhecidos como: Primeiro Planalto ou Planalto de Curitiba, cuja altitude varia entre 850 e 950 metros; Segundo Planalto ou Planalto de Ponta Grossa, formando a região dos Campos Gerais; Terceiro Planalto, ou o planalto de Guarapuava que possui sua altitude média igual a 170 metros (ESPÍRITO SANTO JÚNIOR; SILVEIRA, 2007). O relevo do estado pode ser observado na Figura 2.

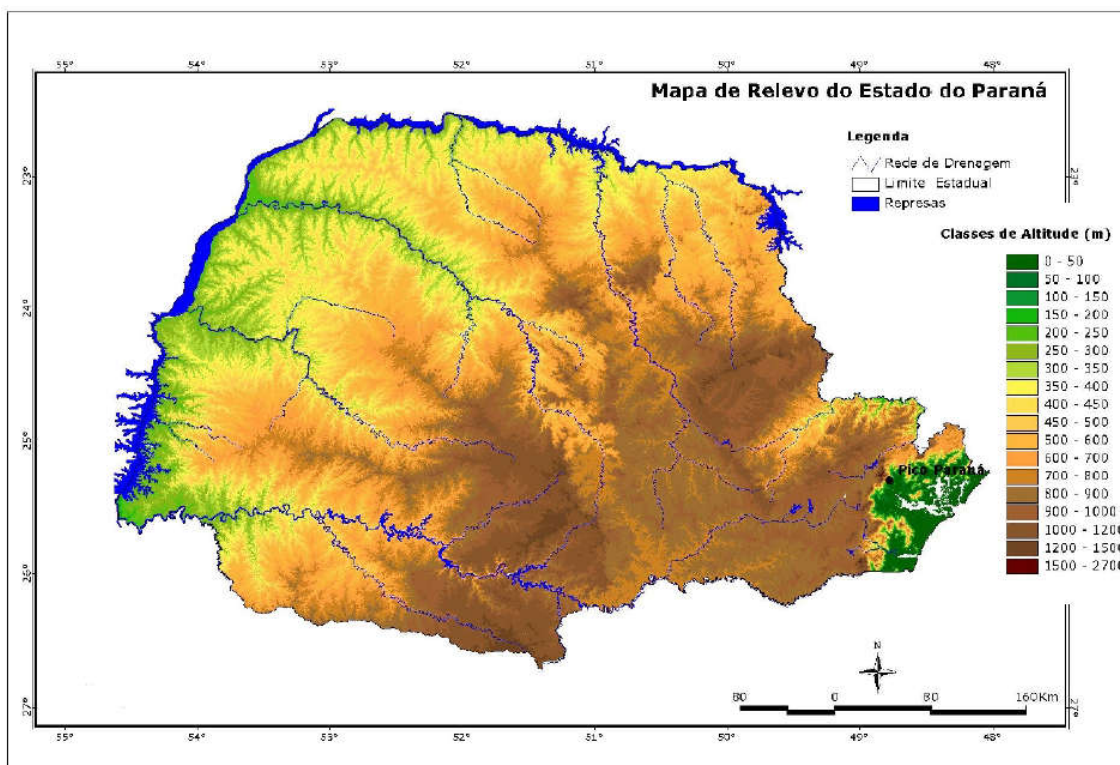


Figura 2 Relevo do estado do Paraná.

Fonte: SRTM (2012)

3.1.3 Clima

Identificar o clima de uma região permite conhecer condições médias de precipitação pluvial e de temperaturas esperadas. Esse é o primeiro passo para se planejar todas as atividades humanas e explorações vegetais e animais (DEMARCHI, 2010).

O Paraná se localiza em uma transição climática com diversos microclimas e com diferentes situações de temperatura e de precipitação, sendo essa variação associada com variações de latitude e altitude (DEMARCHI, 2010).

No estado do Paraná predomina o clima Subtropical, com uma pequena região com clima Tropical. A média anual de temperatura dos municípios paranaenses varia de 15 a 24

°C (Figura 3), e a amplitude pluviométrica varia de 1.100 a 1.920 mm por ano (Figura 4) (APARECIDO et al., 2016).

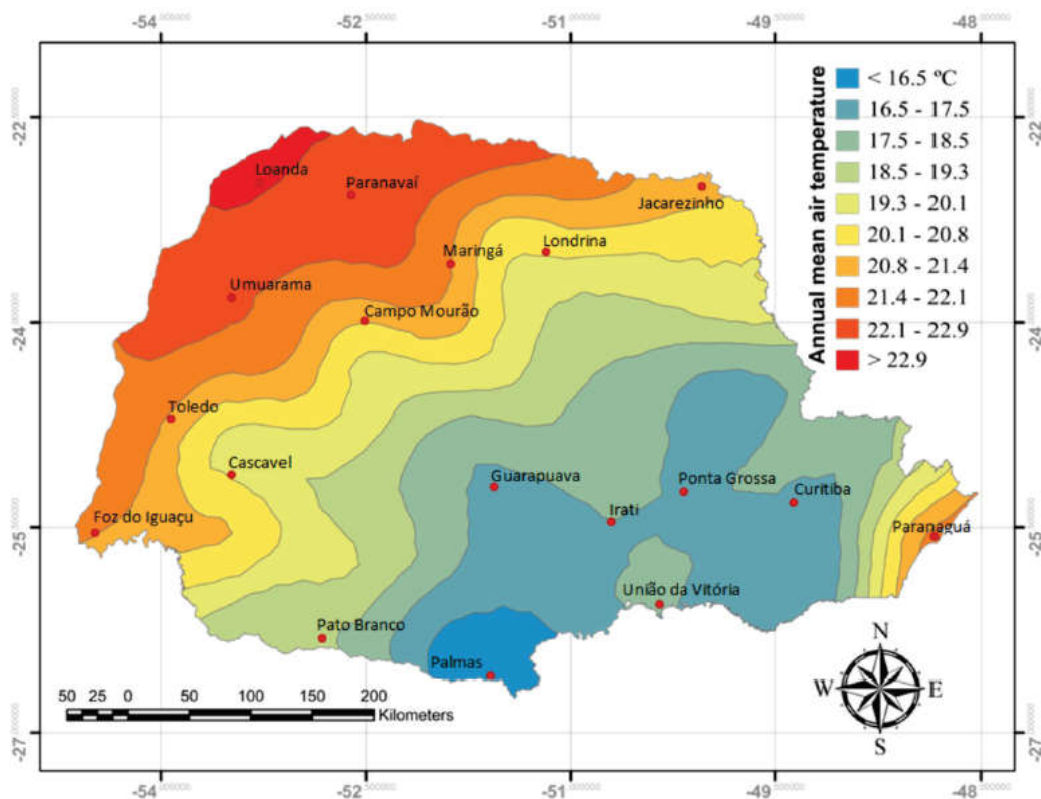


Figura 3 Temperatura média anual no estado do Paraná.
Fonte: Aparecido et al. (2016).

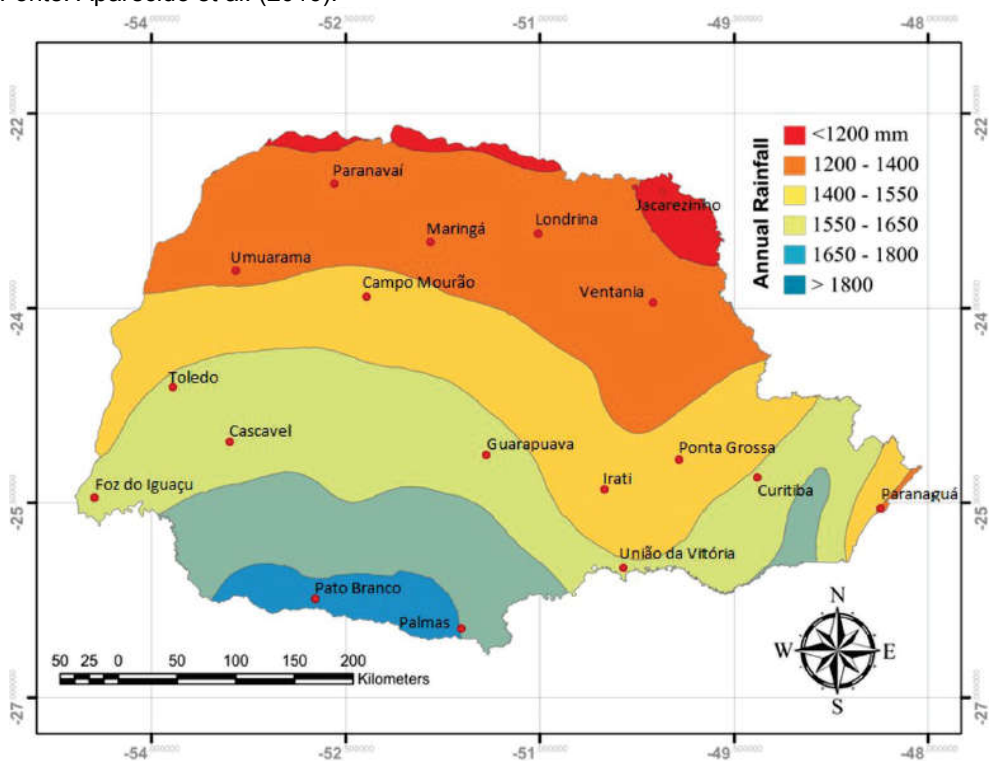


Figura 4 Precipitação anual no estado do Paraná.
Fonte: Aparecido et al. (2016).

A classe climática com maior predomínio no estado do Paraná é a Cfa (clima temperado úmido com verão quente), representando 50% do estado. Este clima não apresenta época de seca, e, neste clima, caracteriza-se por ser o local mais úmido do Paraná com precipitação anual de 1.800 mm por ano.

A classe Cfb (clima temperado úmido com verão moderadamente quente) é a segunda maior, representando 39,8% da área do estado, ocorrendo em áreas de alta altitude sem estação seca e com precipitação anual variando de 1.300 mm a 1.850 mm por ano.

Em relação ao clima Aw (savana tropical úmida), a estação mais seca coincide com o inverno e a precipitação máxima nos meses mais secos tende a ser inferior a 60 mm por mês, sendo que em alguns locais apresenta uma precipitação anual de menos de 1.250 mm por ano. Além disso, este clima ocupa 5,7% do Paraná.

Em relação ao clima Cwa (clima temperado úmido com inverno seco), o verão é quente, com temperaturas do ar no mês mais quente acima de 22 °C, precipitação média inferior a 60 mm em pelo menos um dos meses da estação e ocupa 4,3% do estado.

De acordo com a classificação de Köppen, no Paraná predominam os climas do tipo C (Mesotérmico) e o clima do tipo A (Tropical Chuvoso), conforme Figura 5.

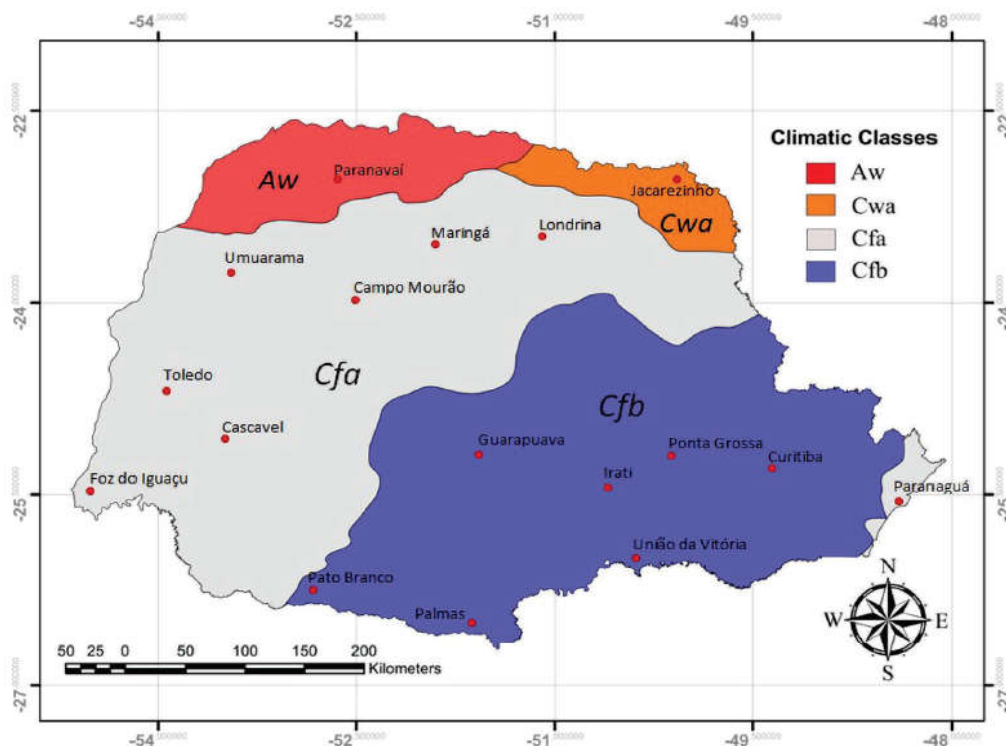


Figura 5 Classificação do clima no estado do Paraná.

Fonte: Aparecido et al. (2016).

3.1.4 Economia do estado e de seus municípios

O Paraná é um estado agrícola, que possui um setor agroindustrial predominantemente primário (PARANÁ, 2018). De acordo com IPARDES (2018), O PIB do

Paraná, em 2016, foi de aproximadamente 378 bilhões de reais, o que caracteriza o estado como a quinta maior economia do Brasil.

Em contrapartida, a maioria dos municípios paranaenses não é autossuficiente, ou seja, dependem dos repasses do governo federal (WOITKE; MOURÃO, 2015). Uma parte importante da receita dos municípios é obtida por meio do repasse do Imposto sobre operações relativas à Circulação de Mercadorias e Prestação de Serviços de Transporte de Comunicação, Intermunicipal e Interestad – ICMS.

Em 2015, a arrecadação de ICMS no estado do Paraná foi de 331,85 bilhões de reais (PARANÁ, 2017), sendo que 25% desse montante são repassados aos municípios (BRASIL, 1988).

A distribuição do ICMS aos municípios (Cota-parte do ICMS) é realizada por critérios compostos por variáveis, como: população, área, mortalidade infantil, evasão escolar, e principalmente, o valor adicionado fiscal (VAF). O VAF realiza a medição da riqueza econômica, que está ligada à produção de bens e serviço.

Tabela 1 Critérios de Participação dos municípios na Cota-parte do ICMS

Critério	% Participação no Índice	Fonte
Valor Adicionado Fiscal (VAF)	75%	Secretaria da Fazenda
Produção Agropecuária	8%	SEAB/DERAL
População	6%	IBGE
Fator Ambiental	5%	SEMA (IAP e Instituto das Águas)
Propriedades rurais	2%	INCRA
Área territorial dos municípios	2%	SEMAA/ITC
Índice Fixo	2%	Secretária da Fazenda
Total	100%	

Fonte: Adaptado de Paraná (2017).

Verifica-se na Tabela 1 que a maior parte (75%) da redistribuição do ICMS se refere ao fator de produção de bens e serviços captado através do valor adicionado fiscal, 10% utilizam-se de aspectos relacionadas a atividade rural, 10% são distribuídos através de critérios de isonomia com os municípios paranaenses e 5% relacionados a questões ambientais (ICMS ecológico).

De acordo com a Lei estadual 9.491 de 21 de dezembro de 1990, é assegurado que os municípios que possuem em seu território mananciais de abastecimento público, unidades de conservação ambiental ou que sejam diretamente influenciados por elas, cinco por cento (5%) dos recursos do Cota-parte do ICMS (Paraná, 1990).

Devido a vários municípios paranaenses depender dos repasses intergovernamentais, é importante verificar a veracidade dos dados, isto é possível por meio da estimativa de produção e estimativa de áreas florestais.

3.2 Uso e ocupação do solo no Paraná

A análise do uso e ocupação do solo de uma superfície é a primeira etapa para nortear uma gestão eficaz no que se refere à adequação do meio, seja este rural ou urbano (MARCUSI, 2010). Esta análise consiste em identificar o uso que é dado ao solo em uma determinada região; geralmente o resultado desta é apresentado em forma de mapa.

No Paraná as terras destinadas para lavouras perfazem 73,2% da área, o restante do estado se divide em pastagens plantadas (14,1%); silvicultura e pastagens naturais (4,6%); e as áreas sem aptidão agrícola (8,1%) (ESPÍRITO SANTO JÚNIOR; SILVEIRA, 2007).

Uma forma de realizar uma análise de uso e ocupação do solo é separar os alvos permanentes das áreas agricultáveis; entretanto, neste trabalho, além dessa separação também serão identificados culturas perenes, culturas semi-perenes e culturas anuais.

3.2.1 Alvos permanentes

Considera-se alvo permanente toda a cobertura florestal, corpos hídricos e cidades (PALUDO et al., 2014) ou seja, toda área que não é destinada a agricultura ou pastagem. Dentro dessa classificação se enquadra floresta, reflorestamento, cidades, rios lagos entre outros alvos que não proporcionam a prática agropecuária.

3.2.2 Culturas Perenes

São classificados como agricultura perene os cultivos agrícolas com o ciclo fenológico longo que compreende vários anos e que possuem a capacidade de produzir por várias safras, como, por exemplo: café e as árvores frutíferas (MAPBIOMAS, 2017).

A agricultura perene é a classe de uso e ocupação do solo com a menor dinâmica de mudanças em seu ciclo, se comparada com a agricultura semi-perene e agricultura anual. Além disso, esta classe possui uma pequena quantidade de área ocupada, tanto em nível nacional, quanto estadual (MAPBIOMAS, 2017).

3.2.3 Culturas semi-perenes

A agricultura semi-perene compreende os cultivos agrícolas com o ciclo fenológico maior que o período de um ano, além de possuir a capacidade de produzir por mais de uma safra (MAPBIOMAS, 2017). No estado do Paraná as culturas semi-perene de maior destaque são as de cana-de-açúcar e de pastagem.

3.2.4 Cana-de-açúcar

A cana-de-açúcar é uma planta de origem asiática e foi introduzida no cenário brasileiro pelos colonizadores portugueses (DILLON, 2007), sendo que o cultivo de cana-de-

açúcar desempenhou um papel importante para a história do Brasil, tornando-se a primeira cultura agrícola a ser exportada no país (NONATO, 2010).

Usualmente, a cana possui cinco cortes (ciclo de crescimento), sendo necessário a reforma do canavial após esses ciclos pois a cada ciclo a produtividade diminui. O primeiro ciclo da cana-de-açúcar é chamado de “cana-planta” e este possui dois intervalos de tempo que são usados. Se esta for plantada de setembro a outubro, o intervalo do ciclo é de um ano; entretanto, se a cana for plantada de janeiro a março, o ciclo possui um intervalo de 18 meses (NONATO, 2010).

Atualmente; a cana-de-açúcar é utilizada como matéria prima na produção de açúcar e combustível (etanol) (FERNANDES, 2009); porém, também é utilizada para alimentação de animais, para fabricação de cachaça e xarope de cana, dentro outros produtos.

No estado do Paraná, essa cultura é mais difundida no noroeste do estado. De maneira geral, o Paraná apresentou uma produção média entre os anos de 2012 e 2016 de 47.834.613 toneladas ao ano; além disso, a média de área colhida de cana-de-açúcar no estado foi de 651.312 ha por ano, sendo que os municípios que mais produzem pertencem a regiões Noroeste, Norte central e Norte pioneiro (Figura 6) (IPARDES, 2018).

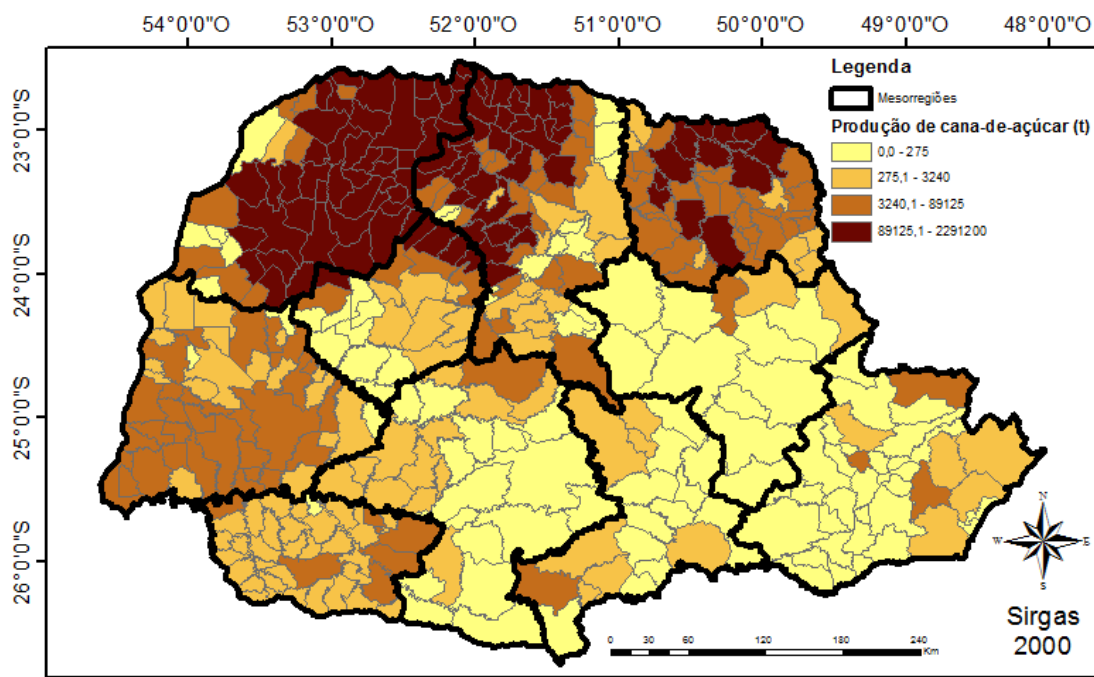


Figura 6 Produção média de cana-de-açúcar nos municípios paranaenses entre os anos de 2012 e 2016.

Fonte: IBGE (2017).

3.2.4.1 Pastagem

O estado do Paraná apresenta uma pecuária na maioria extensiva, com variedades de solos e de climas que propiciam a implantação de diversas pastagens de qualidade além da criação e adaptação de diferentes espécies bovinas e seus cruzamentos (MEZZADRI, 2007).

Os municípios com os números mais expressivas na pecuária de corte no estado são Umuarama e Paranavaí, porém existem outras regiões muito importantes na atividade, como: Londrina, Maringá, Campo Mourão, Jacarezinho e Ponta Grossa (MEZZADRI, 2007).

3.2.5 Culturas anuais

A classe de culturas anuais compreende todos os cultivos agrícolas com o ciclo fenológico menor do que o período de um ano, restrito a uma safra; além disso, esse tipo de cultura apresenta uma maior dinâmica de mudanças se comparado com os demais alvos (MAPBIOMAS, 2017).

No geral, os ciclos dessas culturas se concentram em três época de cultivo, sendo elas culturas de primavera-verão, culturas de verão-outono e cultura de outono-inverno. O ciclo das culturas de primavera-verão é considerado o período com as melhores condições de temperatura e de precipitação, sendo este compreendido entre os meses de setembro e abril. Para as culturas de verão-outono, o período de abrangência é de fevereiro a agosto. Por fim, o período da cultura de inverno tem início no mês de maio, podendo se estender até novembro (SEAB/DERAL, 2017)

Segundo IPARDES (2018), as culturas anuais mais cultivadas no estado são: soja, milho, feijão e trigo. Na sequência são descritas informações relevantes sobre as principais culturas anuais com o intuito de conhecer as informações que podem ser úteis para diferenciá-las dos demais alvos terrestres – e posteriormente diferenciá-las entre si.

3.2.5.1 Soja

A soja (*Glycine max* (L.) Merrill) é uma das principais fontes de proteína e óleo vegetal do mundo e tem sido utilizada para alimentação humana e animal por milênios, tendo como local de origem a região Leste da China (BORÉM, 1999).

No cenário brasileiro, o estado do Paraná se destaca sendo o segundo maior produtor de soja do país, considerando a média de área colhida de soja nos anos de 2012 a 2016 foi de 4.978.258 ha, sendo que os municípios com as maiores produções se concentram nas mesorregiões: Centro ocidental, Oeste e Centro oriental (Figura 7) e a média de produção para esse mesmo intervalo de data foi de 15.208.636 toneladas por ano (IPARDES, 2018).

A semeadura da soja ocorre, em grande maioria, nos meses de outubro a dezembro, enquanto que a colheita compreende os meses de janeiro a abril (Tabela 2).

Tabela 2 Porcentagem mensal de semeadura e colheita da soja no estado do Paraná

Safra	Ano-safra	Situação	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai
1º	14-15	Plantio	7	40	43	10	0	0	0	0	0
		Colheita	0	0	0	0	8	20	49	23	0
	13-14	Plantio	2	45	48	5	0	0	0	0	0
		Colheita	0	0	0	0	4	38	37	20	1
	12 13	Plantio	3	43	51	3	0	0	0	0	0
		Colheita	0	0	0	0	1	38		59	2
	11 12	Plantio	4	47	43	5	1	0	0	0	0
		Colheita	0	0	0	0	3	19	43	32	3

Fonte: Adaptado de SEAB/Deral (2017).

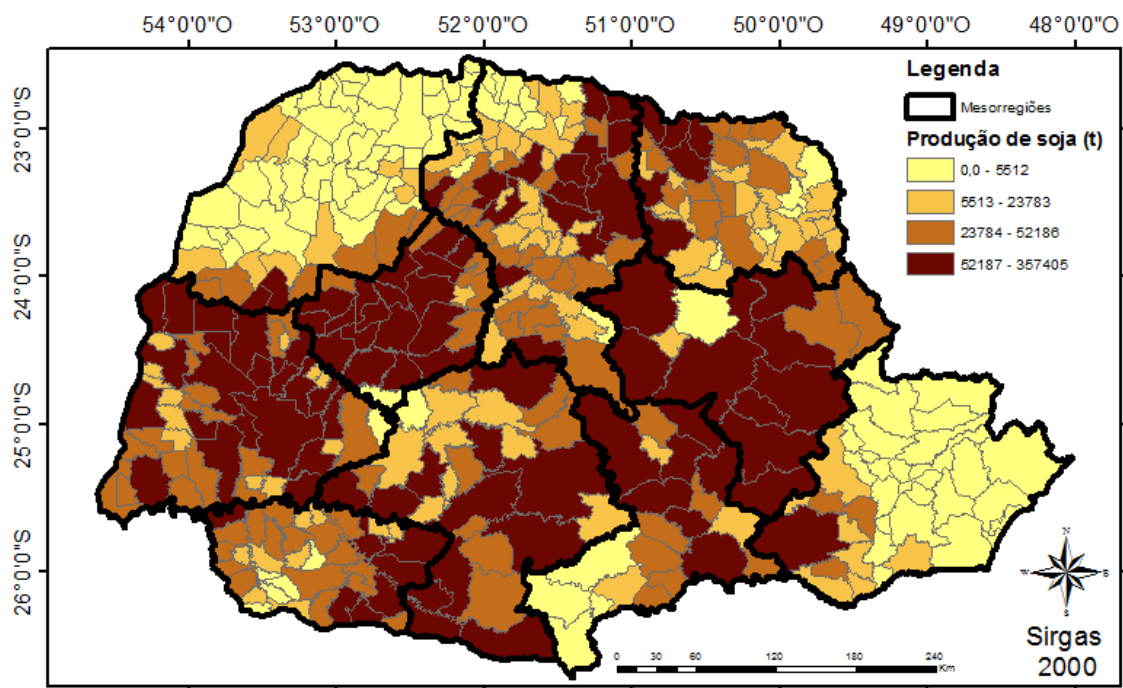


Figura 7 Produção média de soja nos municípios paranaenses entre os anos de 2012 e 2016. Fonte: IBGE (2017)

3.2.5.2 Milho

O cultivo do milho teve origem na América, porém existem pesquisadores que defendem a hipótese desta cultura ter iniciado na Ásia; entretanto, o consenso é de que há milhares de anos o milho desempenha um papel importante na história. Nos dias atuais, o milho é usado diretamente na alimentação humana e de animais, constituindo matéria prima básica para uma grande quantidade de produtos industrializados (CANÉCHIO FILHO, 1987).

O estado do Paraná, a cultura do milho apresentou a produção média de 15.877.077 ton por ano (calculada entre os anos de 2012 e 2016) (IPARDES, 2018). Entretanto, esses dados referem-se à soma das duas safras de milho, sendo que, para a segunda safra, os

municípios que mais possuem área plantada pertence, ao “cinturão da soja”, que se estende da região Oeste até o Norte pioneiro (Figura 8).

Pode-se identificar (Figura 8) que a cultura de milho não é muito difundida no litoral e na região noroeste do estado. Percebe-se também que para o milho segunda safra (Figura 8b) apresenta maior produção na região conhecida como “cinturão da soja”.

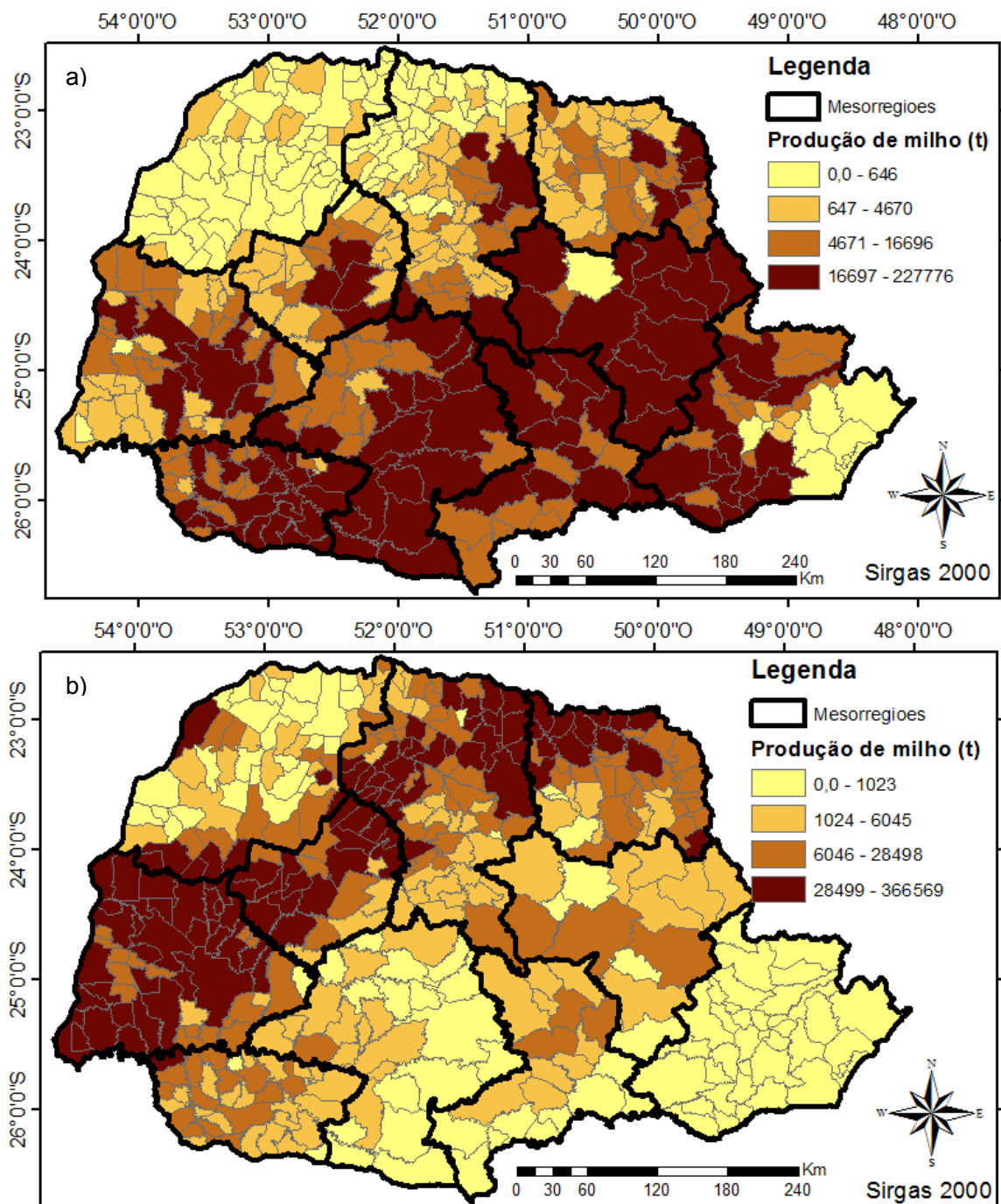


Figura 8 Média da produção de milho (2012 a 2016) [a] primeira safra; b) segunda safra] nos municípios paranaenses.

Fonte: IBGE (2017)

Informações oficiais de plantio e colheita do milho primeira e segunda safra podem ser observadas na Tabela 3, sendo que a época de plantio do milho primeira safra se concentra entre os meses de setembro e novembro, e a colheita, em sua maioria, compreende os meses de fevereiro a abril (Tabela 3).

Tabela 3 Porcentagem mensal de semeadura e colheita do milho no estado do Paraná para as duas safras

Safra	Ano-safra	Situação	Set	Out	Nov	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago
1º	14-15	Plantio	41	49	9	0	0	0	0	0	0	0	0
		Colheita	0	0	0	6	21	43	24	5	1	0	0
	13-14	Plantio	27	58	14	0	0	0	0	0	0	0	0
		Colheita	0	0	0	1	27	35	30	5	2	0	0
	12 13	Plantio	26	60	13	0	0	0	0	0	0	0	0
		Colheita	0	0	0	0	19	0	73	6	2	0	0
	11 12	Plantio	47	39	12	0	0	0	0	0	0	0	0
		Colheita	0	0	0	1	21	33	36	8	0	1	0
2º	14-15	Plantio	0	0	0	11	38	44	7	0	0	0	0
		Colheita	0	0	0	0	0	0	0	2	7	22	57
	13-14	Plantio	0	0	0	8	49	40	3	0	0	0	0
		Colheita	0	0	0	0	0	0	0	1	4	41	41
	12 13	Plantio	0	0	0	2	49	42	7	0	0	0	0
		Colheita	0	0	0	0	0	0	0	0	3	26	55
	11 12	Plantio	0	0	0	4	22	62	12	0	0	0	0
		Colheita	0	0	0	0	0	0	0	1	3	22	57

Fonte: Adaptado de SEAB/Deral (2017).

Para a segunda safra de milho, a época de plantio se inicia em janeiro e se estende até abril, sendo que nos quatro anos-safra analisados os meses de fevereiro e março responderam por mais de 80% do plantio de milho segunda safra, enquanto que a colheita ocorre, em grande maioria, entre os meses de julho e setembro.

3.2.5.3 Cultura de inverno

Dentre as culturas de inverno mais difundidas no Paraná estão as culturas de aveia, centeio, cevada, sorgo, trigo e triticale, sendo que o trigo é a principal cultura de outono-inverno do estado do Paraná, sendo que sua produção média dos anos de 2012 a 2016 representou 89,66% do total das principais culturas de inverno (aveia, centeio, cevada, sorgo, trigo e triticale) (IPARDES, 2018).

De modo geral, as mesorregiões Litorânea e Noroeste possuem os municípios com as menores áreas plantadas de culturas de outono-inverno e, conseqüentemente, possui baixa produção (Figura 9).

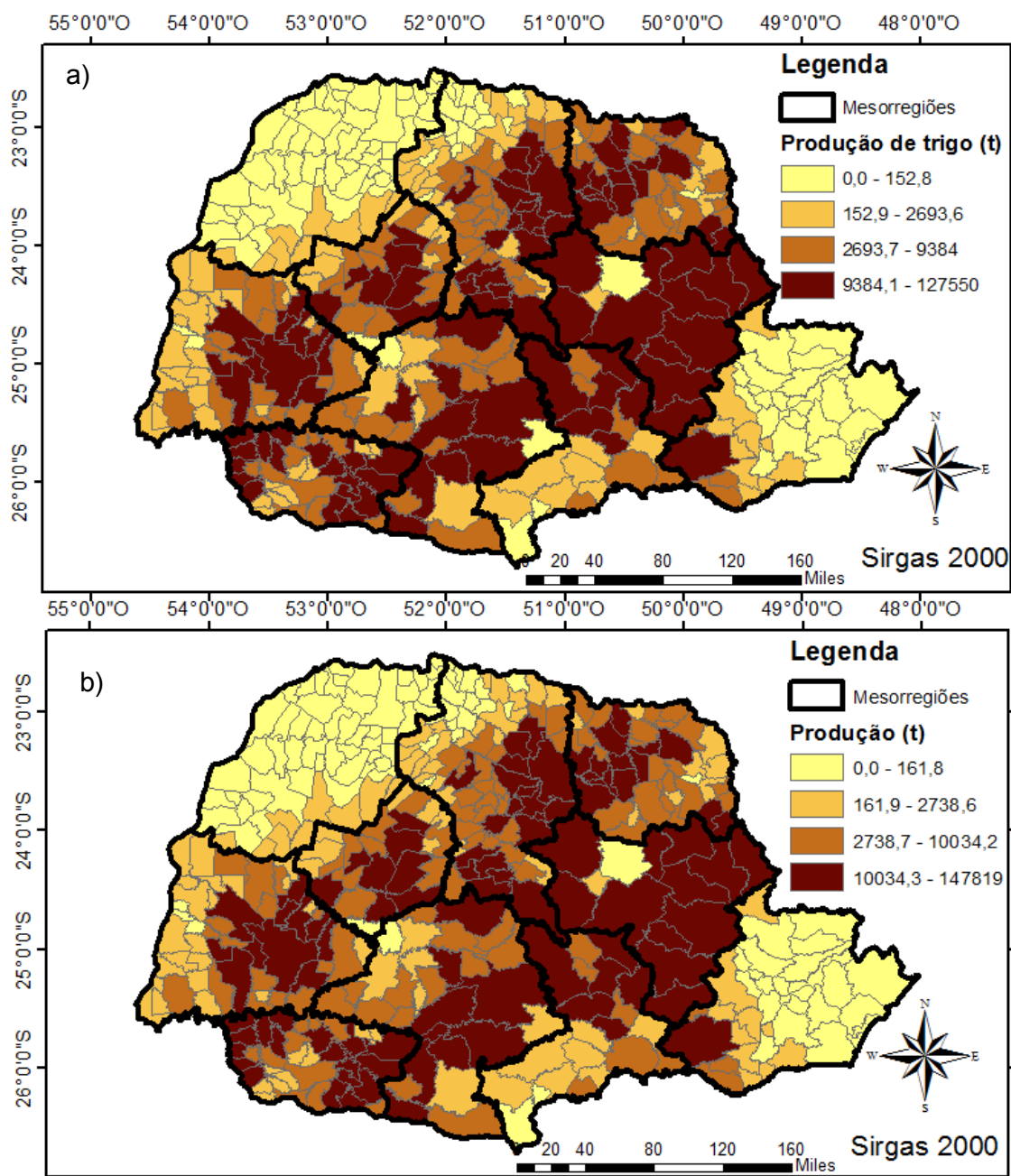


Figura 9 Produção anual média (2012 a 2016) de: a) trigo; e b) culturas de outono-inverno, nos municípios paranaenses.

Fonte: IBGE (2017)

O calendário da cultura de trigo compreende os meses de maio a dezembro, com o plantio concentrado nos meses de maio e junho, e a colheita realizada, em sua maioria, entre os meses de setembro e novembro (Tabela 4).

Tabela 4 Porcentagem mensal de semeadura e colheita do trigo no estado do Paraná

Ano-Safra	Situação	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
14-15	Plantio	61	21	18	0	0	0	0	0
	Colheita	0	0	0	4	37	36	22	1
13-14	Plantio	57	28	15	0	0	0	0	0
	Colheita	0	0	0	3	23	34	35	5
12 13	Plantio	52	29	19	0	0	0	0	0
	Colheita	0	0	0	9	47	22	22	0
11 12	Plantio	67	20	13	0	0	0	0	0
	Colheita	0	0	0	4	50	20	20	6

Fonte: Adaptado de SEAB/Deral (2017).

3.2.5.4 Feijão

O feijão é uma planta anual herbácea, pertencente à família das leguminosas. Essa planta possui raiz não muito profunda, a parte aérea consta de uma base principal ramificada e as flores apresentam um cálice protegido por bractéolas persistente. O fruto do feijão é uma vagem de comprimento que varia de 10 a 20 cm e, quando está madura, apresenta uma coloração amarelada com sementes de formatos cilíndricos ou elíptico (CANÉCHIO FILHO, 1987).

O plantio deve ser realizado, de preferência, em regiões com primaveras e outonos longos e não devem possuir verão e inverno muito rigorosos. O ideal é haver precipitação pluviométrica de cerca de 100 mm na época do plantio e do crescimento, com temperaturas entre 10 e 25 °C. As temperaturas altas são prejudiciais à cultura do feijão, por isso não é recomendável o plantio nos meses de novembro e dezembro. Para o solo, os ideais são os que possuem textura areno-argiloso e são ricos em materiais orgânicos, com pH menor que 7 (CANÉCHIO FILHO, 1987).

Entre os anos de 2012 e 2016, o estado do Paraná apresentou produção média de 701.722 toneladas por ano (IPARDES, 2018). Entretanto, essa cultura é cultivada em três safras (Figura 10 e Tabela 5); porém, a maior parte ocorre nas duas primeiras safras.

Todas as mesorregiões do estado possuem áreas com plantio de feijão (Figura 10); entretanto, para o feijão terceira safra (Figura 10c) o norte do estado concentra grande parte da produção.

De maneira geral, o plantio e a colheita do feijão podem ser realizados durante todo o ano, levando em conta que este possui três safras. O plantio do feijão primeira safra ocorre entre os meses de setembro e outubro, e a colheita desta safra ocorre, em grande maioria, nos meses de janeiro e fevereiro (Tabela 5). Para o feijão segunda safra o plantio ocorre nos três primeiros meses do ano e a colheita se concentra nos meses de abril, maio e junho (Tabela 5). Por fim, o feijão terceira safra possui a época de plantio entre os meses de março e maio, e a colheita de junho a agosto, podendo ocorrer até outubro (Tabela 5).

Tabela 5 Porcentagem mensal de semeadura e colheita do feijão no estado do Paraná para as três safras

Safra	Ano-safra	Situação	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	
1º	14-15	Plantio	49	36	14	1	0	0	0	0	0	0	0	0	
		Colheita	0	0	4	7	70	17	2	0	0	0	0	0	
	13-14	Plantio	22	62	19	1	0	0	0	0	0	0	0	0	
		Colheita	0	0	0	11	63	24	2	0	0	0	0	0	
	12 13	Plantio	20	63	16	1	0	0	0	0	0	0	0	0	
		Colheita	0	0	2	14	43	36	5	0	0	0	0	0	
	11 12	Plantio	50	32	16	2	0	0	0	0	0	0	0	0	
		Colheita	0	0	4	6	54	32	4	0	0	0	0	0	
	2º	14-15	Plantio	0	0	0	0	37	48	15	0	0	0	0	0
			Colheita	0	0	0	0	0	0	7	16	53	23	1	0
		13-14	Plantio	0	0	0	0	33	47	20	0	0	0	0	0
			Colheita	0	0	0	0	0	0	1	8	37	50	4	0
12 13		Plantio	0	0	0	4	21	51	24	0	0	0	0	0	
		Colheita	0	0	0	0	0	3	0	10	35	40	12	0	
11 12		Plantio	0	0	0	0	27	42	31	0	0	0	0	0	
		Colheita	0	0	0	0	0	0	0	17	34	32	17	0	
3º		14-15	Plantio	0	0	0	0	0	0	33	34	31	2	0	0
			Colheita	0	0	0	0	0	0	0	0	14	22	23	38
		13-14	Plantio	0	0	0	0	0	0	24	42	24	4	6	0
			Colheita	0	0	0	0	0	0	0	0	3	15	33	37
	12 13	Plantio	0	0	0	0	0	0	0	70	21	3	6	0	
		Colheita	0	0	0	0	0	0	0	0	0	37	8	46	
	11 12	Plantio	0	0	0	0	0	0	0	87	2	11	0	0	
		Colheita	0	0	0	0	0	0	0	0	0	41	8	45	

Fonte: Adaptado de SEAB/Deral (2017).

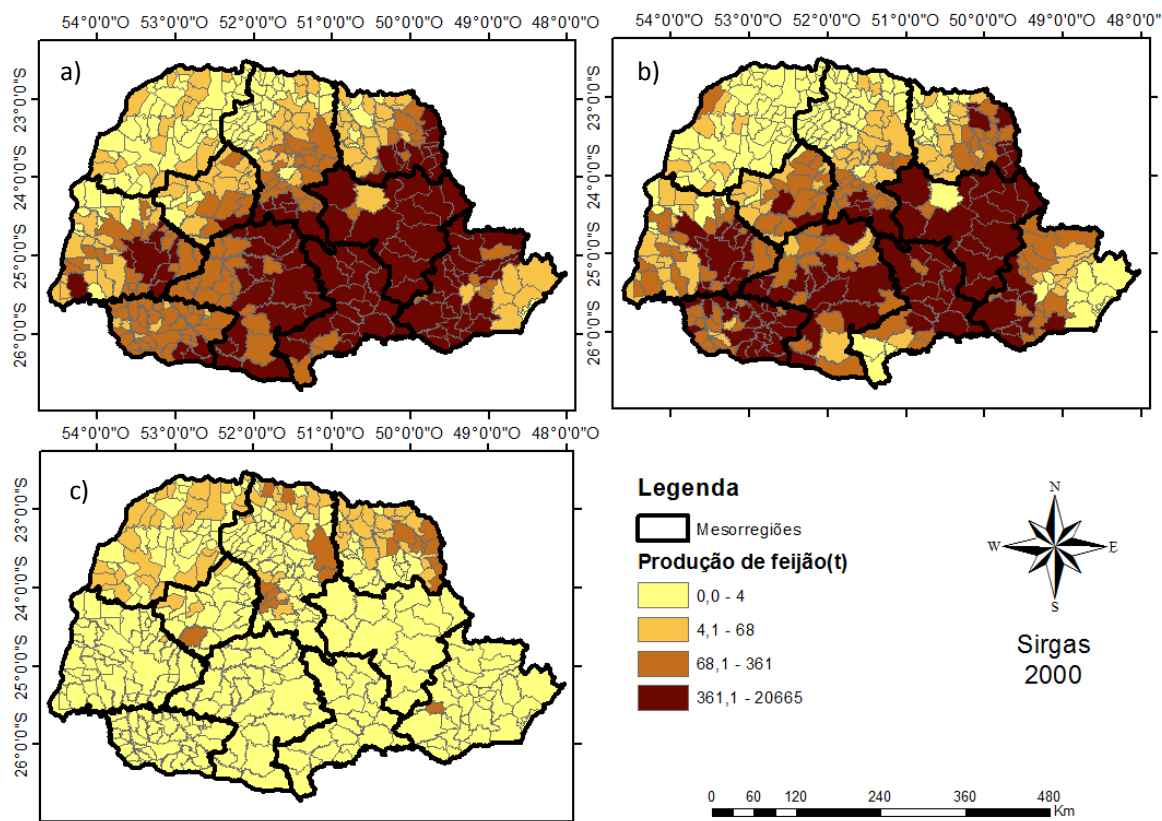


Figura 10 Produção média (2012 a 2016) de feijão [a) primeira safra; b) segunda safra, c) terceira safra] nos municípios paranaenses.
Fonte: IBGE (2017)

3.3 Sensoriamento remoto

O sensoriamento remoto é uma técnica de obtenção de informações dos objetos da superfície terrestre sem que haja um contato físico de qualquer espécie entre o sensor e o objeto (MENEZES; ALMEIDA, 2012).

O sensoriamento consiste da interação dos alvos terrestre (vegetação, solo, água, etc...) com a radiação eletromagnética, sendo essas interações divididas em três componentes: a refletida, a absorvida e a transmitida (JOSHI et al., 2016).

3.3.1 Sensoriamento remoto aplicado a agricultura

No caso do sensoriamento remoto aplicado à agricultura, o alvo de maior interesse para os estudos é a vegetação, possibilitando, com isso, o monitoramento do crescimento e o desenvolvimento das culturas, além de permitir detectar condições de estresse que limitam a produtividade. Desta forma, esta tecnologia permite obter informações espaciais e temporais detalhadas sobre o comportamento dos alvos nos diferentes ambientes (FERNANDES, 2009).

A radiação eletromagnética, ao interagir com o alvo terrestre, se divide em três componentes: a refletida, a absorvida e a transmitida. Cerca de 50% da energia que chega até a planta é absorvida; os outros 50% restantes são refletidos ou transmitidos (MOREIRA, 2005).

A resposta espectral da refletância da vegetação refere-se à capacidade da planta em refletir a radiação eletromagnética nos diferentes comprimentos de onda do espectro. (FERNANDES, 2009).

Os alvos terrestres apresentam diferentes valores de refletância nos diversos comprimentos de ondas, por exemplo: a vegetação aumenta a refletância na faixa do visível e diminui na faixa do infravermelho próximo (IVP) ao passar do período de crescimento para senescência, devido a alterações nos mecanismos biológicos (GUYOT, 1989) e, com isso, é possível diferenciar os alvos observados.

3.3.2 Índice de vegetação

Os índices de vegetação representam combinações matemáticas de diferentes bandas espectrais que têm como finalidade ressaltar o comportamento espectral da vegetação em relação aos demais alvos da superfície terrestre (VIÑA et al., 2016).

Em determinados comprimentos de onda, a vegetação pode ser confundida com demais alvos terrestres. Para facilitar a diferenciação de alvos os índices de vegetação utilizam as bandas que mais distinguem vegetação, como é o caso do infravermelho próximo, do vermelho, entre outras.

Os índices de vegetação mais utilizados são o Normalized Difference Vegetation Index (NDVI) (ROUSE et al., 1973) e o Enhanced Vegetation Index (EVI) (HUETE et al., 1997), sendo que o NDVI foi utilizado nos estudos de Couto Jr., Carvalho Jr. e Martins (2012), Andrade et al. (2017) e Zhang et al. (2017), entre outros; enquanto que o EVI foi utilizado em Grzegozewski et al. (2015), Mengue e Fontana (2015), Johann et al. (2016) e Zhong et al. (2016).

Sendo assim, o NDVI é calculado pela Equação 1.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad \text{Eq.(1)}$$

em que:

NIR - refletância no infravermelho próximo;

RED - refletância na região do vermelho.

O EVI (Equação 2), foi criado com o intuito de reduzir a influência do sinal do solo e da atmosfera sobre a resposta do dossel e para aprimorar a detecção da vegetação em regiões com maiores densidades de biomassa (KUPLICH; MOREIRA; FONTANA, 2013).

$$EVI = 2,5 \frac{NIR - RED}{1 + NIR + 6RED - 7,5BLUE} \quad \text{Eq. (2)}$$

em que:

NIR - Reflectância na região espectral do infravermelho próximo;

RED - Reflectância na região espectral do vermelho;

BLUE - Reflectância na região espectral do azul.

3.3.3 Sensores

Atualmente, a resolução espectral das imagens obtidas pelos sensores já ultrapassa centenas de bandas, e a resolução espacial de muitas imagens já é menor que 1 metro, possibilitando suas aplicações nas áreas de mapeamentos temáticos, monitoração ambiental, desmatamentos florestais, previsões de safras, entre outras (MENEZES; ALMEIDA, 2012).

Devido à gratuidade das imagens, o sensor Modis a bordo dos satélites Terra-Aqua e o sensor OLI, a bordo do satélite Landsat-8, e por serem muito utilizados em mapeamentos de uso e ocupação do solo, serão descritos com mais detalhes abaixo.

3.3.3.1 Modis

Modis (Moderate Resolution Imaging Spectroradiometer ou Espectroradiômetro de Imagem de Resolução Moderada) é um instrumento a bordo dos satélites Terra (originalmente conhecido como EOS AM-1) e Aqua (originalmente conhecido como EOS PM-1) que registra imagens de toda a superfície da Terra no máximo a cada 2 dias, adquirindo dados em 36 faixas espectrais (Tabela 6). O Modis está desempenhando um papel vital no desenvolvimento de modelos de sistemas terrestres globais e interativos, validados e capazes de prever as mudanças globais com precisão suficiente para ajudar em tomadas de decisões acertadas (NASA, 2017).

De acordo com Zhang et al. (2017), as imagens com resolução espacial elevada tendem a apresentar melhores resultados para classificação, pois aumenta a possibilidade de se obter pixels puros devido ao tamanho reduzido de cada pixel; entretanto, quanto maior a resolução espacial, maior será o custo computacional para trabalhar com as imagens.

Tabela 6 Características do sensor Modis a bordo dos satélites Terra e Aqua

Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal	Faixa Imageada
1	620 - 670 nm	250 m	1 a 2 dias	2330 x 5000 km
2	841 - 876 nm			
3	459 - 479 nm	500 m	1 a 2 dias	2330 x 5000 km
4	545 - 565 nm			
5	1230 - 1250 nm			
6	1628 - 1652 nm			
7	2105 - 2155 nm			
8	405 - 420 nm	1000 m	1 a 2 dias	2330 x 5000 km
9	438 - 448 nm			
10	483 - 493 nm			
11	526 - 536 nm			
12	546 - 556 nm			
13	662 - 672 nm			
14	673 - 683 nm			
15	743 - 753 nm			
16	862 - 877 nm			
17	890 - 920 nm			
18	931 - 941 nm			
19	915 - 965 nm			
20	3660 - 3840 nm			
21	3929 - 3989 nm			
22	3929 - 3989 nm			
23	4020 - 4080 nm			
24	4433 - 4498 nm			
25	4482 - 4549 nm			
26	1360 - 1390 nm			
27	6535 - 6895 nm			
28	7175 - 7475 nm			
29	8400 - 8700 nm			
30	9580 - 9880 nm			
31	10780 - 11280 nm			
32	11770 - 12270 nm			
33	13185 - 13485 nm			
34	13485 - 13785 nm			
35	13785 - 14085 nm			
36	14085 - 14385 nm			

Fonte: Embrapa (2017)

3.3.3.2 Landsat-8

Os sensores OLI (Operational Land Imager) e TIRS (Thermal Infrared Sensor), a bordo do satélite Landsat-8, trabalham em 11 bandas espectrais, sendo que as imagens obtidas por estes sensores apresentam três resoluções espaciais distintas, de 15, 30 e 120 m (BROCHADO; MACARIRINGUE, 2014).

A resolução temporal destes sensores é de 16 dias (Tabela 7), e os dados coletados pelos instrumentos a bordo do satélite estão disponíveis para download dentro de 24 horas da sua aquisição.

Tabela 7 Características do sensor OLI abordo dos satélites Landsat-8

Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal	Área Imageada
(B1) Azul	0,45 – 0,52 μm			
(B2) Verde	0,52 – 0,60 μm			
(B3) Vermelho	0,63 – 0,69 μm	30 m	16 dias	184 km
(B4) Infravermelho Próximo	0,76 – 0,90 μm			
(B5) Infravermelho Médio	1,55 – 1,75 μm			
(B6) Infravermelho Termal	10,4 – 12,5 μm	120 m	16 dias	184 km
(B7) Infravermelho Médio	2,08 – 2,35 μm	30 m	16 dias	184 km
(B8) Pancromático	0,52 – 0,90 μm	15 m	16 dias	184 km

Fonte: Embrapa (2017)

A detecção espacial de alvos agrícolas é facilitada, comparada com o sensor MODIS, por apresentar uma resolução espacial de 30 m para as bandas espectrais do visível e do infravermelho próximo. Em função disto, diversos trabalhos utilizam o sensor com um método de validação de mapeamentos agrícolas elaborados com imagens do sensor MODIS (GRZEGOZEWSKI et al., 2016; JOHANN et al., 2016; ZHANG et al., 2017).

3.3.4 Séries temporais

Estudos de séries temporais ou multi-temporais envolvem a variável tempo. No caso do sensoriamento remoto, correspondem à aquisição de imagens com mais de uma data.

O uso de séries multi-temporais em uma área cultivada permite acompanhar o comportamento dos alvos ao longo do seu desenvolvimento, além de possibilitar a comparação de diferentes comportamentos no ciclo da cultura (FENANDES, 2009). Isso ocorre devido aos diferentes estágios fenológicos que a cultura possui. De acordo com Zhong et al. (2014), o estágio fenológico pode ser diretamente utilizado para separar culturas de calendários distintos.

Além disso, a utilização de informações obtidas da fenologia das culturas permite acelerar o processo de classificação, pois gera regras mais rapidamente e as torna mais estáveis (ZHONG et al., 2016). Os estágios fenológicos mais utilizados referem-se a semeadura, pico vegetativo e colheita.

Na data da semeadura, o padrão espectro-temporal da superfície se assemelha à assinatura espectro-temporal do solo. Entretanto, a medida que a cultura cresce a influência do solo diminui e o comportamento da superfície tende a se ajustar ao comportamento de vegetação. Ao atingir o pico vegetativo, a cultura entra em senescência, quando o vigor vegetal e o porte das plantas diminuem, alterando o padrão espectral da superfície, que volta ao padrão do solo, de acordo com a Figura 11 (FORMAGGIO, 1989; ESQUERDO, 2007).

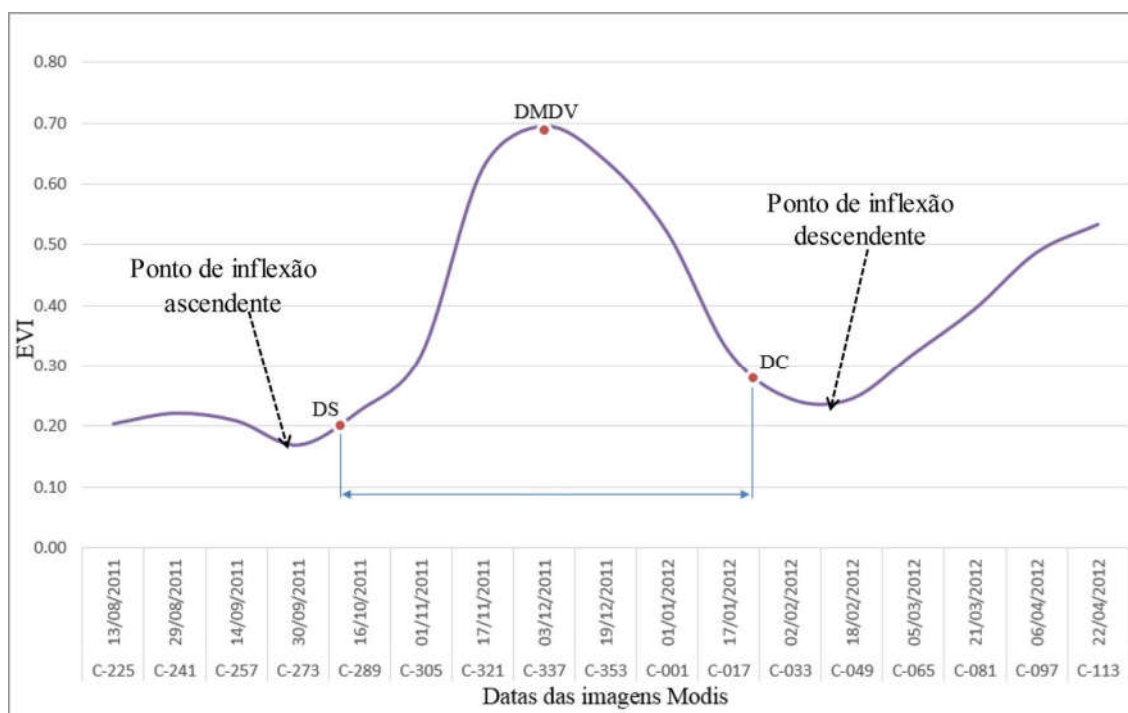


Figura 11 Padrão espectro-temporal do índice de vegetação EVI de uma cultura agrícola compreendendo o período do dia juliano (c) 225 ao 113, com as datas DS (data de semeadura), DMDV (data de máximo desenvolvimento vegetativo) e DC (data de colheita).
Fonte: Johann et al. (2016)

Por conta das diversas peculiaridades que o padrão espectral da vegetação apresenta ao decorrer do tempo, é recomendável trabalhar com séries temporais, pois, dessa forma, aumentará a quantidade de dados com grande potencial de realizar a separação dos alvos em estudo.

A utilização de séries temporais para realização de mapeamentos pode ser encontrada em vários estudos. Johann et al. (2012) estimaram áreas com culturas de soja e milho no Paraná utilizando imagens multi-temporais de EVI obtidas do sensor Modis para os anos-safra de 2004/05 a 2007/08.

Zhong et al. (2016) desenvolveram uma abordagem automatizada para mapear soja e milho no estado do Paraná para os anos de 2010 a 2015, utilizando uma série temporal de imagens do sensor Modis a fim de extrair informações referentes à fenologia das culturas e obtiveram resultados considerados bons, com Exatidão Global de 87,2% e Kappa de 0,80.

Couto Jr., Carvalho Jr. e Martins (2012) analisaram séries temporais do sensor Modis para descrever as sucessões de soja-milho e milho-soja em um sistema de plantio direto, utilizando um período de 12 anos. Neste trabalho, concluíram que o perfil espectro-temporal evidencia o comportamento da cultura ao longo do tempo.

O uso de séries temporais em procedimento de classificação de alvos é possível em função das culturas possuírem padrões fenológicos distintos (ZHANG et al., 2017). Além das

culturas, os demais alvos terrestres também possuem um determinado padrão espectral-temporal, como ilustra a Figura 12.

O comportamento dos perfis temporais de NDVI de cana-de-açúcar, soja, mata, cidade e pastagem (Figura 12) possuem diferenças entre si. Além disso, verifica-se que determinados alvos têm comportamentos mais estáveis, como é o caso da mata, diferentemente da soja, que tem uma significativa variação de valores de NDVI ao longo do seu ciclo (JOHANN; ARAÚJO; ROCHA, 2009).

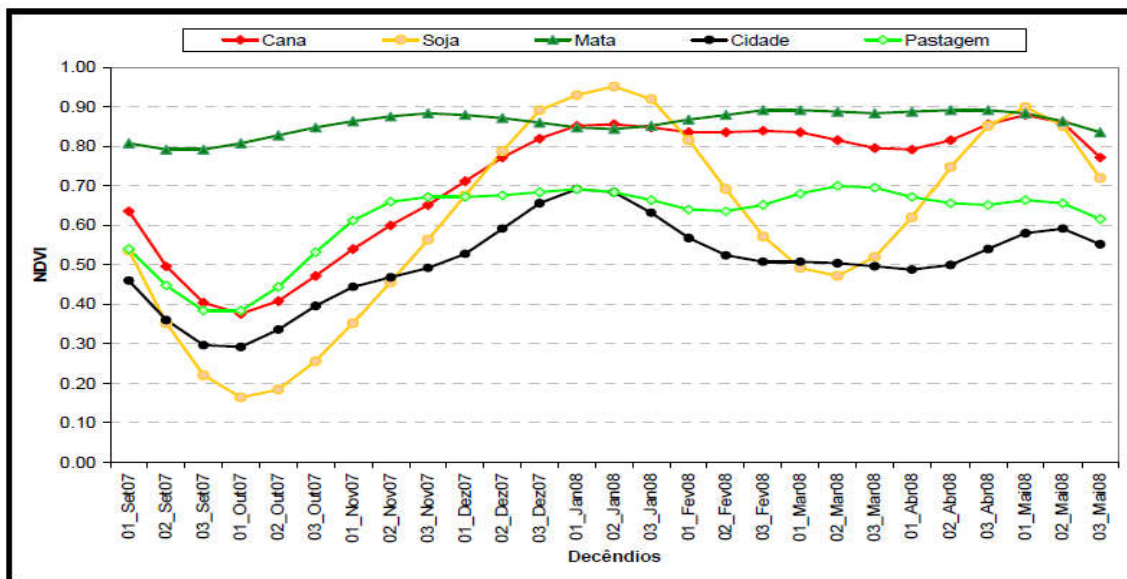


Figura 12 Perfil temporal do índice de vegetação NDVI de pixels de cana-de-açúcar, soja, mata, cidade e pastagem.

Fonte: Johann; Araújo; Rocha (2009).

Entretanto, o uso de séries temporais brutas de índice de vegetação apresenta limitações oriunda de coberturas de nuvens e outros tipos de ruídos (COUTO Jr.; CARVALHO Jr.; MARTINS, 2012). Por conta disso, ao se trabalhar com dados multi-temporais é indispensável a aplicação de técnicas de filtragem para eliminação destes ruídos. Estudos realizados por Adami (2010) mostraram que o uso dos filtros Hants, Savitzky–Golay e duplo-logística tiveram desempenho similar, sendo indicados para filtrar as séries temporais.

O filtro de Savitzky-Golay é um filtro de suavização de sinais analíticos (ADAMI, 2010), que se baseia na ideia de correção dos valores da amostra por uma ponderação com os valores vizinhos (COSTA, 2011).

De forma geral, o filtro Savitzky-Golay apresenta bons desempenhos para a maioria das situações, apresentando-se como uma ótima técnica para o estudo de pastagens e culturas agrícolas (ABADE, 2015).

3.3.5 Mapeamento

Mapeamento de vegetação é uma técnica importante tanto para a gestão dos recursos naturais, quanto para o agronegócio, pois permite compreender o uso do solo do território (XIE; SHA; YU, 2008).

A metodologia de mapeamento por imagem orbital, além de fornecer a localização das áreas das culturas, permite a realização de estimativas de área de forma mais objetiva que as utilizadas pelos órgãos oficiais de estatísticas agrícolas (JOHANN et al., 2012).

A utilização de índices de vegetação de imagens de satélite, aliada a uma moderada resolução espacial, como o sensor Modis, e analisada por meio de séries temporais possibilita o monitoramento regional de culturas agrícolas (JOHANN, 2012).

A aplicação dessa forma de mapeamento também foi utilizada por Vicente et al. (2012), que classificaram três classes de cobertura do solo (área de cana-de-açúcar, pasto e floresta) para o estado de São Paulo, com o uso de séries temporais de 6 anos de NDVI do sensor Vegetation, a bordo do satélite SPOT (Système Pour l'Observation de la Terre), e do algoritmo "Spectral Angle Mapper" (SAM).

Outro trabalho nessa mesma linha foi elaborado por Johann et al. (2012), que teve como objetivo estimar e mapear área com culturas de soja e milho para o estado do Paraná, utilizando imagens multi-temporais de EVI do sensor Modis. O mapeamento foi realizado com imagens de máximo e mínimo EVI, formando assim uma série temporal de imagem RGB onde R representava imagens de máximo EVI e GB representava imagem de mínimo EVI. Por fim, concluiu-se que a metodologia utilizada é altamente eficiente.

Na pesquisa realizada por Grzegozewski et al. (2015), o objetivo foi separar, mapear e estimar as áreas de soja e milho no estado do Paraná nos anos safras 2012/13 e 2013/14 utilizando imagens EVI/MODIS. Para realizar o mapeamento foram integradas duas metodologias distintas: a primeira consiste em elaborar imagens de mínimo e máximo EVI, criando uma composição RGB que identifica as culturas simultaneamente; e a segunda metodologia dotada foi uma classificação supervisionada SAM (Spectral Angle Mapper).

Son et al. (2016) também utilizaram séries temporais de EVI/MODIS; entretanto, o objetivo dos autores foi desenvolver uma abordagem de classificação da cultura de arroz baseada na fenologia para avaliar os sistemas de cultivo. As imagens utilizadas nesta pesquisa são da região de Mekong no Vietnã e referem-se ao período de 2000 a 2012.

Entretanto, existe a necessidade de realizar o mapeamento de maneira rápida e com baixo custo, o que requer a automatização do processo, de modo que este possa ser realizado várias vezes ao longo dos anos, tornando assim o processo de mapeamento mais econômico e dinâmico (ZHONG et al., 2016).

Segundo Zhong et al. (2016), o processo de automatização exige muito tempo e esforço para determinar as regras de classificação, além de ser essencial o conhecimento sobre práticas agrícolas locais e estágios fenológicos das culturas pesquisadas.

Metodologia semelhante foi utilizada por Zhang et al. (2017), que mapeou e estimou as áreas cultivadas para as principais culturas da Coreia do Norte, por meio de características fenológicas extraídas de séries temporais de NDVI do sensor Modis. Além disso, também propuseram uma metodologia automatizada a fim de facilitar futuros mapeamentos.

Para Mengue e Fontana (2015), o objetivo foi desenvolver uma metodologia de baixo custo para mapear culturas de verão no estado do Rio Grande do Sul no ano-safra 2011/12, com base no comportamento espectro-temporal de imagens EVI/MODIS e no modelo HAND.

Na pesquisa de Zhong et al. (2016), o objetivo foi desenvolver uma abordagem automatizada para mapear soja e milho no estado do Paraná utilizando o classificador de árvore de decisão. Foram utilizadas imagens de EVI do sensor Modis. A metodologia deste artigo consistiu em utilizar métricas extraídas dos estádios fenológicos para gerar regras de decisão do classificador, separando os alvos em soja e milho no Estado.

3.4 KDD

O processo de Descoberta de Conhecimento em Base de Dados (Knowledge Discovery in Databases – KDD) é um conjunto de atividades contínuas que tem como objetivo encontrar padrões válidos, novos, úteis e acessíveis em um banco de dados. Esse processo é composto por cinco etapas (Figura 13): seleção de dados, pré-processamento, formatação, mineração de dados e avaliação/interpretação (FAYYAD et al., 1996).

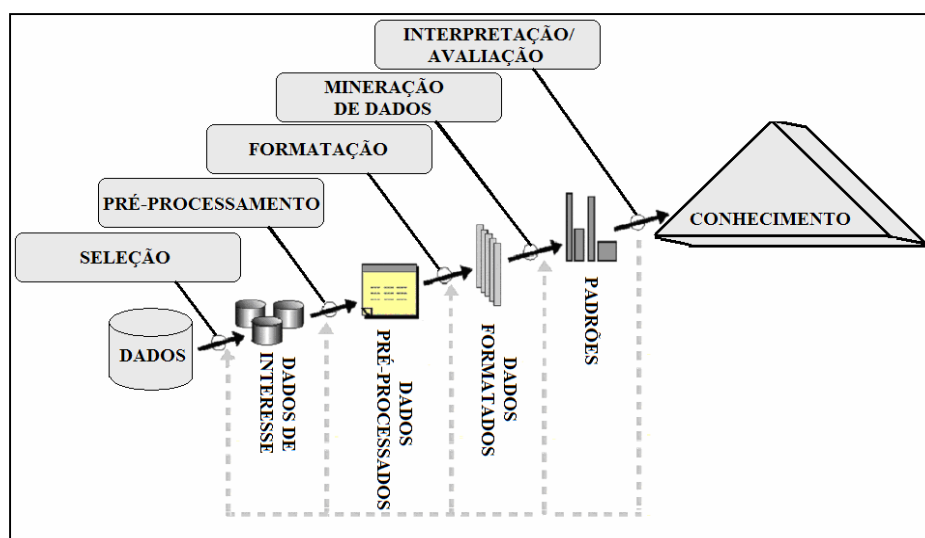


Figura 13 Processo Descoberta de Conhecimento em Base de Dados (KDD).
Fonte: Adaptado de Fayyad et al. (1996)

3.4.1 Seleção de dados

Na primeira fase é realizada a seleção dos dados de interesse. Nesse momento é necessário conhecer os objetos da pesquisa, de modo a possibilitar a obtenção de alguma informação nova sobre eles. O termo objetos se refere a pessoas, produtos, serviços, municípios ou qualquer outra entidade que possa ser analisada em uma quantia de atributos ou variáveis (HAIR Jr., 2005).

Ainda de acordo Tan, Steinbach e Kumar (2009), nessa fase devem ser priorizadas apenas as variáveis de interesse, sem verificar se seus valores são inconsistentes ou duplicados.

3.4.2 Pré-processamento

Para essa etapa o objetivo é excluir dados inconsistentes e ajustar variáveis de modo a melhorar a análise na etapa de mineração de dados.

Pode haver objeto que não tenha um ou mais valores atribuídos às variáveis. Existem casos em que esses valores não são coletados, e em outros casos as variáveis não são aplicáveis a todos objetos; entretanto, os valores em falta devem ser levados em conta durante a análise de dados (TAN; STEINBACH; KUMAR, 2009).

Para Boente, Goldschmidt e Estrela (2008), nessa fase ocorre a limpeza de dados, que consiste em corrigir informações errôneas de maneira a não comprometer o resultado do processo. Esses autores também ressaltam a importância de agregar mais informações aos dados selecionados, ou seja, realizar ajustes para que os dados possam fornecer mais conteúdo relevante ao processo.

3.4.3 Formatação

Nessa etapa ocorre a transformação das variáveis, de maneira a possibilitar a aplicação da etapa das demais etapas do KDD.

O processo de descoberta de conhecimento em base de dados pode ser bastante sensível à diferença de escalas e magnitude entre as variáveis, ou seja, as variáveis com maior dispersão apresentam um maior impacto sobre o resultado da mineração de dados (HAIR Jr. et al., 2005).

Por esse motivo é importante, para o resultado obtidos no KDD, diminuir esse impacto, e uma forma de fazer isso seria pelo processo de padronização das variáveis. Esse processo converte escores de dados iniciais em valores padronizados com média igual a 0 e desvio padrão igual a 1.

Segundo Tan, Steinbach e Kumar (2009), conjuntos de dados podem ter um grande número de variáveis e os algoritmos de mineração funcionam melhor com menos variáveis; portanto, reduzindo o número de variáveis, melhor será o desempenho do algoritmo.

Além do melhoramento do algoritmo a redução de dimensionalidade pode gerar um modelo mais compreensível e auxiliar na redução de características irrelevantes e de ruídos.

Ruídos são componentes aleatórios de um erro de medição, e estão relacionados à distorção dos valores e à adição de objetos ilegíveis (TAN; STEINBACH; KUMAR, 2009).

3.4.4 Mineração de dados

Essa etapa é considerada a principal etapa do processo KDD, sendo o núcleo do processo, pois é nesta etapa que se realiza a busca por conhecimentos úteis nos dados que foram selecionados pré-processados e formatados (MACEDO; MATOS, 2010).

A descoberta de conhecimento é realizada através de ferramentas computacionais, sendo que estas se dividem em duas categorias: tarefa de previsão cujo objetivo é prever o valor de um atributo baseando-se em valores dos outros atributos; e a tarefa descritiva, que tem o objetivo de derivar padrões que resumam os relacionamentos subjacentes nos dados (TAN; STEINBACH; KUMAR, 2009).

Dentro dessas categorias se destacam as tarefas relacionadas a classificação, associação e agrupamento de padrões (FAYYAD et al., 1996).

De acordo com Macedo e Matos (2010), a tarefa de classificação tem o objetivo de encontrar uma função ou regras de decisão para mapear um conjunto de dados em categorias pré-determinadas, denotadas por classes. Após descoberta desta função, pode ser aplicada a novos dados, buscando prever a classe em que estes se enquadram.

Uma forma de encontrar a regra de decisão é com a elaboração de uma árvore de decisão. Esta fornece uma abordagem de modelagem alternativa a estatística clássica, por conta da sua representação simbólica e interpretável do modelo que pode ser utilizada tanto para predição quanto para conhecer a estrutura predita (MEGETO et al., 2014).

O algoritmo Random Forest constrói várias árvores de decisão usando um subconjunto aleatório de atributos obtidos do conjunto de dados originais, em que cada subconjunto gera uma árvore de decisão. Para determinar a classe final de uma instância é verificado qual foi a classe mais votada dentre todas as árvores de decisão geradas, sendo este processo repetido para todas as instâncias (BREIMAN, 2011; LORENZETI, 2016). Segundo Lorenzetti (2016), esse algoritmo apresenta resultados melhores do que utilizando uma árvore de decisão apenas, porém demanda maior tempo computacional para a execução deste.

A tarefa de agrupamento tem como objetivo identificar grupos de dados onde estes tenham características semelhantes aos do mesmo grupo e os grupos tenham características diferentes entre si (WAGNER; ZABADAL; RIBEIRO, 2016).

Por fim, a tarefa de associação se utilizada para descobrir padrões que descrevam características associadas dentro dos dados. Esses padrões encontrados são representados na forma de regras ou subconjunto de características (TAN; STEINBACH; KUMAR, 2009).

3.4.5 Avaliação e Interpretação

A etapa de Interpretação dos resultados consiste em validar o conhecimento extraído (FAYYAD et al., 1996). Essa etapa tem como objetivo interpretar as informações obtidas por meio do processo de mineração de dados e organizar os resultados de maneira que estes possam ser compreendidos.

Para isso, é comum a utilização de técnicas de visualização na forma de tabelas ou gráficos. Através destas técnicas podem-se “observar rapidamente grandes quantidades de informações visuais e encontrar padrões nas mesmas” (TAN; STEINBACH; KUMAR, 2009).

4 METODOLOGIA

A presente dissertação visa mapear as áreas de floresta, demais alvos permanentes, culturas semi-perenes, soja/feijão, milho 1ª e 2ª safra e culturas de inverno, no Paraná, para cada um dos quatro anos-safra estudados (2013/14, 2014/15, 2015/16 e 2016/17), utilizando métricas extraídas de perfil espectro-temporal de índice de vegetação de imagens do sensor Modis (Figura 14).

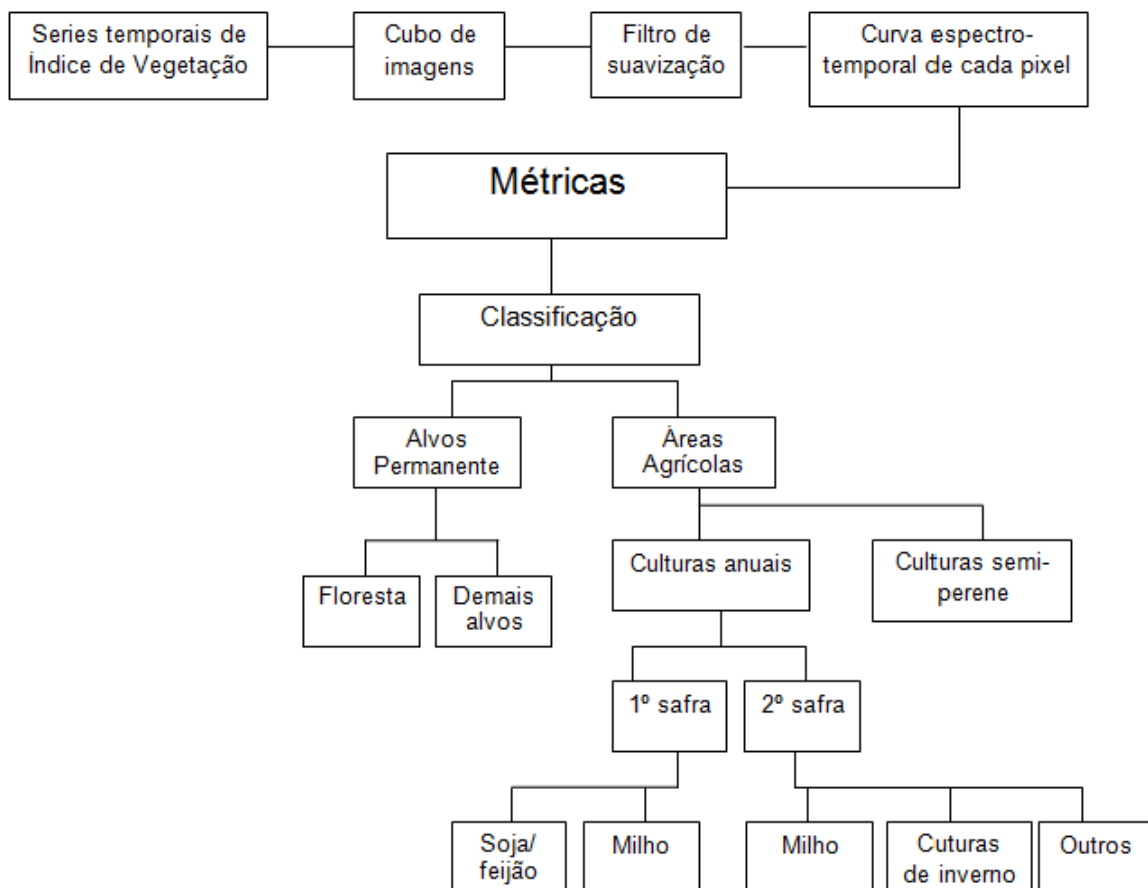


Figura 14 Fluxograma para a elaboração do mapeamento de um ano-safra seguindo o processo KDD

A metodologia para construção do mapeamento foi elaborada no software R versão 3.3.1 (R Development Core Team, 2017), por meio de rotinas desenvolvidas com a finalidade de automatizar o processo de mapeamento.

Os procedimentos metodológicos foram divididos em cinco etapas, de acordo com o processo KDD, sendo estas: seleção, pré-processamento, formatação/transformação, mineração de dados e avaliação/interpretação, de modo que essas etapas são utilizadas para elaborar o mapeamento de um ano-safra, sendo repetido o processo para a elaboração dos demais anos-safra.

Para realizar o mapeamento no software R, primeiramente selecionaram-se as imagens, sendo que estas foram adquiridas no site da Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA (EMBRAPA, 2017). As imagens utilizadas no processo foram de índice de vegetação EVI do sensor Modis (Terra e Aqua), para os anos-safra de 2013/14 a 2016/17.

Após selecionar as imagens, estas foram filtradas utilizando o filtro de suavização Savitzck-Goley aplicado com o auxílio do software Cymp v.1.07 (PALOSCHI, 2016).

As imagens filtradas serão utilizadas como dado de entrada no Software R, com a finalidade de extrair o perfil espectro temporal de cada pixel da imagem e, a partir deste, são elaboradas as métricas, representadas esquematicamente na Figura 15 e definidas na Tabela 8.

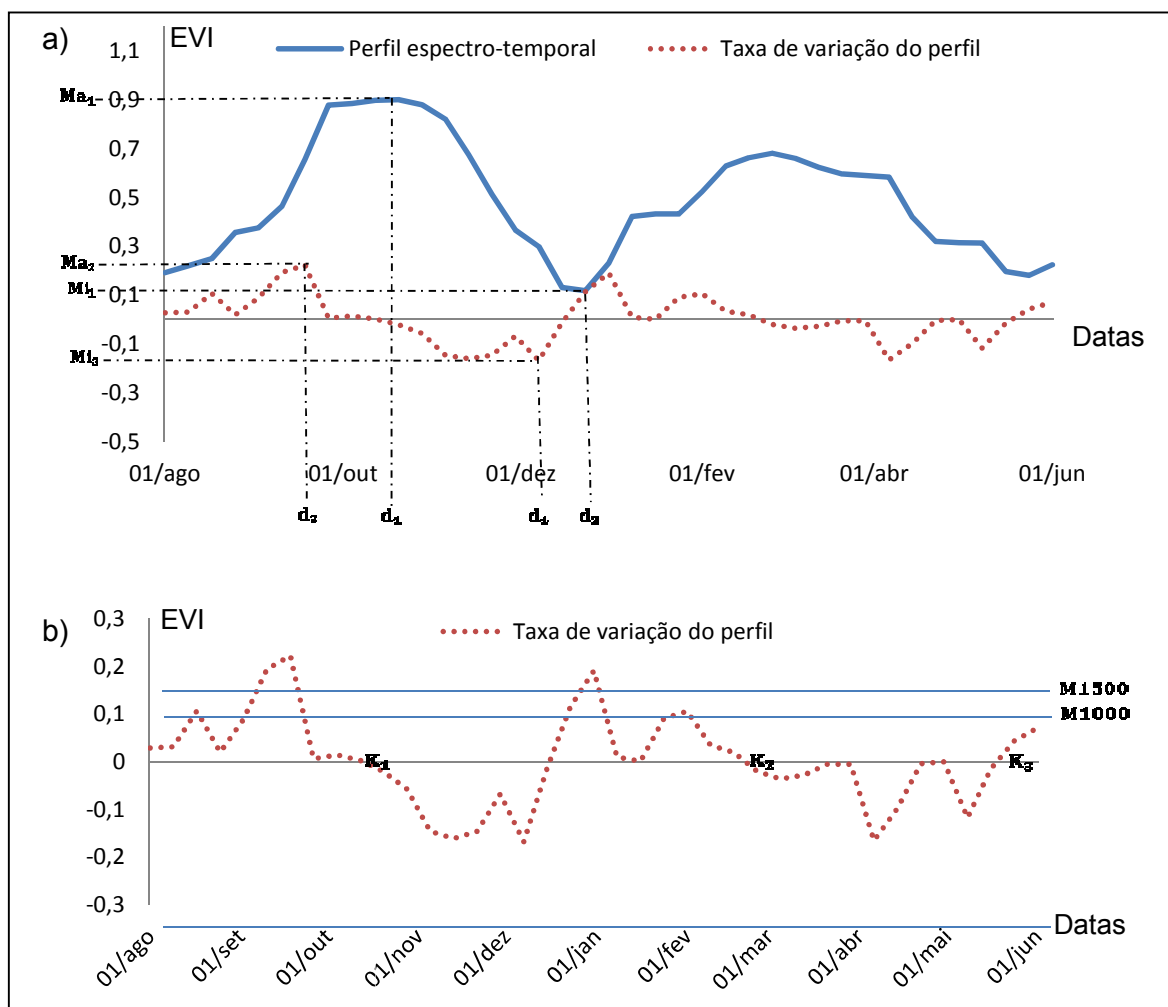


Figura 15 Representação esquemática das métricas obtidas de uma série espectro-temporal de índice e vegetação (a) com ênfase na curva referente a taxa de variação de uma série temporal (b).

Tabela 8 Métricas extraídas de uma série espectro-temporal de índice de vegetação (EVI)

Métricas	Descrição
Ma_1	Valor do máximo EVI registrado em cada pixel
Mi_1	Valor do mínimo EVI registrado em cada pixel
dif_1	Diferença entre o máximo e o mínimo EVI
Ma_2	Valor da maior taxa de variação do EVI registrado em cada pixel
Mi_2	Valor da menor taxa de variação do EVI registrado em cada pixel
K_i	Pontos de mudança da taxa de variação
d_1	Data que ocorre o máximo EVI para cada pixel
d_2	Data que ocorre o mínimo EVI para cada pixel
d_3	Data que ocorre a maior taxa de variação do EVI para cada pixel
d_4	Data que ocorre a menor taxa de variação do EVI para cada pixel
dif_2	Diferença entre as datas que ocorre o máximo e o mínimo EVI
dif_3	Diferença entre as datas que ocorre o Ma_2 e o Mi_2
M 0,15	Número de datas com taxa de variação maior que 1500
M 0,10	Número de datas com taxa de variação maior que 1000
Média	Média do EVI de todas as datas para cada pixel
Int	Soma dos valores de EVI de todas as datas para cada pixel
Dp	Desvio padrão de cada pixel
CV	Coefficiente de variação de cada pixel ao longo das datas

Para auxiliar no desenvolvimento de cada métrica foram elaborados os perfis espectro-temporais médios de cada alvo utilizando pixels puros de imagens do sensor MODIS, extraídos por meio de análise visual de imagem Landsat-8 (Figura 16).

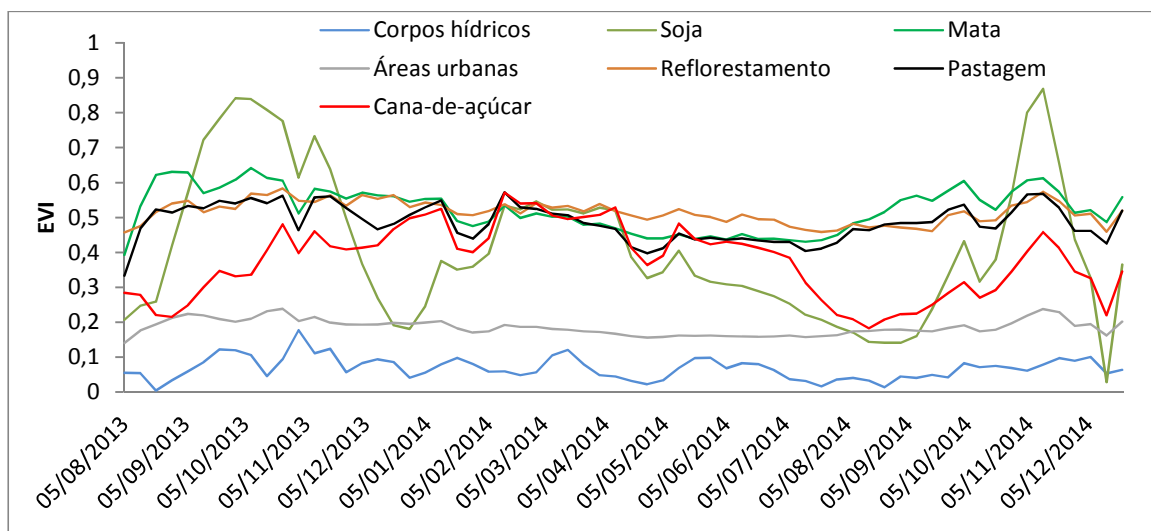


Figura 16 Perfil espectro-temporal médio de cada alvo.

Desse modo, cada uma das métricas foi construída para auxiliar na diferenciação dos alvos de interesse, sendo que o Ma_1 (Valor do máximo EVI registrado em cada pixel) e Mi_1 (Valor do mínimo EVI registrado em cada pixel) têm como finalidade separar pixels de culturas que apresentam valores elevados da refletância de EVI, além de possibilitar a diferenciação de floresta ou reflorestamento de pixels de cidade e água, pois enquanto áreas de mata e reflorestamento apresentam valores de Ma_1 mais elevados, ou seja, maior valores

de EVI do que a média, os alvos como cidade e água apresentam os menores valores de Mi_1 (Figura 16). Para esta mesma finalidade poderão ser usadas as métricas média e Int (Soma dos valores de EVI de todas as datas para cada pixel).

A métrica denotada por dif_1 (Diferença entre o máximo e o mínimo EVI), Dp (desvio padrão) e CV (Coeficiente de variação) podem diferenciar pixels que representam maiores variação temporal do EVI, como é o caso das culturas agrícolas (Figura 16).

Para as métricas Ma_2 (Valor da maior taxa de variação do EVI registrado em cada pixel), Mi_2 (Valor da menor taxa de variação do EVI registrado em cada pixel), $M_{0,10}$ (Número de datas com taxa de variação maior que 1000) e $M_{0,15}$ (Número de datas com taxa de variação maior que 1500), os maiores valores são características das culturas anuais, pois estes possuem rápido crescimento e ciclo curto de cultivo.

As métricas d_1, d_2, d_3 e d_4 referem-se às datas de ocorrência do Ma_1, Mi_1, Ma_2 e Mi_2 , respectivamente, sendo que estas variáveis se encontram no intervalo de 05 de agosto do primeiro ano até 27 de dezembro do segundo ano, sendo que cada imagem possui um intervalo de 8 dias, totalizando 64 imagens. Estas métricas possuem relevância para diferenciar alvos que possuem padrão espectro-temporal característicos, como é o caso das culturas de milho e soja. Para o mesmo propósito podem ser utilizadas as métricas dif_2 (Diferença entre as datas que ocorre o máximo e o mínimo EVI) e dif_3 (Diferença entre as datas que ocorre o Ma_2 e o Mi_2), que trazem informação referente ao ciclo das culturas.

Por fim, a métrica K_i (Pontos de mudança da taxa de variação) podem auxiliar na separação de alvos que não possuem vegetação dos demais, como é o caso dos alvos urbanos e corpos hídricos, pois estes tendem a não apresentar padrão no perfil espectro temporal, aumentando as quantidades de K_i (Figura 16).

4.1 Mineração de dados

Depois de calculadas as métricas do perfil temporal de EVI/Modis se iniciam os procedimentos para separação dos alvos de interesse. Esse procedimento foi realizado no software R, utilizando como dado de entrada uma matriz cujas linhas representam cada pixel da imagem Modis dentro do estado, e as colunas representam as métricas elaboradas.

Para a etapa de mineração de dados, utilizou-se a tarefa de classificação, por meio do algoritmo Random Forest.

Para a aplicação desse algoritmo elaborou-se um conjunto de amostras formado por pixels puros de cada alvo de interesse, obtidos por meio de uma análise visual das imagens Landsat-8 (Figura 17), e também por meio do perfil espectro-temporal do índice de vegetação no sistema de análise temporal da vegetação (SATVEG, 2017). Por decorrência da diferença da resolução espacial das imagens Modis e Landsat-8, utilizou-se uma grade de pixels Modis

(em que cada pixel possui 250 x 250 m) nas imagens Landsat-8, com o intuito de extrair apenas regiões que compreendessem um pixel inteiro do sensor Modis. Para realizar esta tarefa utilizou-se o software Arcgis 10.3.

Com os pixels puros extraídos construiu-se um conjunto de amostra, sendo que 25% deste conjunto foram utilizados para validação posterior do processo e os outros 75% formaram o “traindata”, sendo este utilizado como entrada para o processo de classificação, realizada no software R, gerando assim regras de decisão para a classificação dos pixels do “traindata”.

Na sequência, utilizam-se as regras geradas para determinar a classificação no restante dos dados.

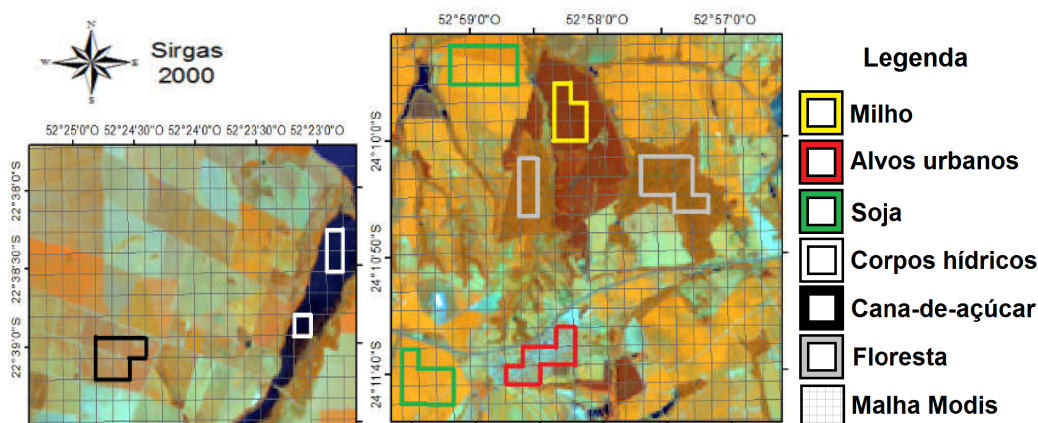


Figura 17 Método manual de seleção de alvos de culturas agrícolas e alvos permanentes na composição falsa cor RGB564 de Landsat-8 com o detalhamento da malha Modis.

Contudo, foram realizados cinco processos de classificação: separação de alvos agrícolas e alvos permanentes (1º artigo), separação dos alvos permanentes entre floresta e demais alvos permanentes (2º artigo), separação de áreas Agrícolas entre cultura anual e cultura semi-perene (2º artigo), separação de agricultura anual em soja/feijão (2º artigo), milho 1ª safra e a separação de milho 2ª safra e culturas de inverno (2º artigo) (Figura 14).

Como produto final dessa fase, obtiveram-se dois mapeamentos por ano-safra, sendo que nestes os pixels da imagem são classificados como floresta, demais alvos permanentes, culturas semi-perenes, soja/feijão, milho e cultura de inverno.

Os resultados gerados nos mapeamentos foram avaliados com base na taxa de exatidão global (EG) (Equação 3), índice Kappa (IK) (Equação 4), Erro de Comissão [EC] e Erro de Omissão [EO], tendo como referência o conjunto de validação.

$$EG = \frac{A}{m} * 100$$

Eq. (3)

$$IK = \frac{\left[m * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_{i.} * x_{.j}) \right]}{\left[m^2 - \sum_{i=1}^r (x_{i.} * x_{.j}) \right]} \quad (4) \quad \text{Eq. (4)}$$

em que:

A - acerto geral (pontos amostrais com acertos);

m - número de pontos amostrais;

n - número de observações;

x_{ij} - observação na linha i e coluna j ;

$x_{i.}$ - total marginal da linha i ;

$x_{.j}$ - total marginal da coluna j .

Utilizaram-se também os indicadores estatísticos: correlação (r), Erro Médio (ME) (Equação 5); Raiz do Erro Médio Quadrático (RMSE) (Equação 6) e o Coeficiente de Concordância Aprimorado de Willmott, Roberson e Matsuura (2012) (dr) (Equação 7), tendo como referência dados oficiais fornecido pelo IBGE.

$$ME = \frac{1}{n} * \sum_{i=1}^n (O - E) \quad \text{Eq. (5)}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^n (O - E)^2} \quad \text{Eq. (6)}$$

$$dr = 1 - \frac{\sum_{i=1}^n |E - O|}{2 \sum_{i=1}^n (|O - \bar{O}|)} \quad \text{Eq. (7)}$$

em que:

n - número de observações;

O - área da cultura usada como referência;

E - área da cultura obtida nos mapeamentos;

\bar{O} - área da cultura média usada como referência.

5 REFERÊNCIAS

- ABADE, N. A. **Uso de séries temporais NDVI do MODIS para mapeamento da cobertura vegetal natural e exótica em uma região de transição Cerrado-caatinga**. 2015. 71 f. Dissertação (Mestrado em Geografia). Universidade de Brasília – UnB, Brasília/DF, 2015.
- ADAMI, M. **Estimativa da data de plantio da soja por meio de séries temporais de imagens MODIS**. 2010. 163 f. Tese (Doutorado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos Campos/SP, 2010.
- ANDRADE, R. G.; BOLFE, É. L.; CASTRO V., D. de; NOGUEIRA, S. F. Avaliação das condições de pastagens no cerrado brasileiro por meio de geotecnologias. **Brazilian Journal of Sustainable Agriculture**, v. 7, n. 1, p. 34-41, 2017.
- APARECIDO, L. E. de O.; ROLIM, G. D. S.; RICHETTI, J.; SOUZA, P. S. de; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, Lavras/MG, v. 40, n. 4, p. 405-417, 2016.
- BOENTE, A. N. P.; GOLDSCHMIDT, R. R.; ESTRELA, V. V. Uma metodologia de suporte ao processo de descoberta de conhecimento em bases de dados. In: V Simpósio de Excelência em Gestão e Tecnologia – SEGeT. **Anais...**, v. 1, p. 4-5, Resende, Rio de Janeiro, 2008.
- BORÉM, A. **Escape Gênico – Os riscos do escape gênico da soja no Brasil**. Encarte Especial - Biotecnologia, Ciência e Desenvolvimento, p.101-107,1999
- BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 05 dez. 2017.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.
- BROCHADO, G. T.; MACARRINGUE, L. S. Uso de imagens do LISS-3/Resourcesat-1 e do OLI/Landsat-8 no estudo espacial da queda do nível de água no lago de Furnas no ano de 2012. **Revista Ambiente & Água**, Taubaté, v. 9, n. 4, p. 664-678, 2014.
- CANÉCHIO FILHO, V. **Principais Culturas**, vol. II. Campinas-SP: Instituto Campineiro de Ensino Agrícola, 1987.
- CECHIM JUNIOR, C.; JOHANN, J. A.; ANTUNES, J. F. G. Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, n. 6, p. 427-432, 2017.
- PORTAL BRASIL. **Agronegócio deve ter crescimento de 2% em 2017**. Disponível em: <<http://www.brasil.gov.br/economia-e-emprego/2016/12/agronegocio-deve-ter-crescimento-de-2-em-2017>>. Acesso em: 15 dez. 2017.
- COSTA, A. A. **Geração de trajetórias sub-ótimas para plataformas flutuantes do tipo Hovercraft**. 2011. 204 f. Tese (Doutorado em Engenharia Mecânica). Universidade Federal do Rio de Janeiro – UFRJ, Rio de Janeiro, 2010.
- COUTO JUNIOR, A.F.; CARVALHO JUNIOR, O.A.; MARTINS, E.S. Séries temporais MODIS aplicadas em sucessão de culturas de soja (*Glycine max* (L.) Merrill) e milho (*Zea mays* L.) em sistema de plantio direto. **Revista Brasileira de Cartografia**, v. 64, p. 405-418, 2012.

DEMARCHI, M. **Análise da conjuntura agropecuária-safra 2010/11**. Boletim informativo/outubro. Curitiba: SEAB – Secretaria da agricultura e do abastecimento do Estado do Paraná, 2010.

DILLON, S. L.; SHAPTER, F. M.; HENRY, R. J. Domestication to Crop Improvement: Genetic Resources for Sorghum and Saccharum. **Annals Of Botany**, London, v. 100, n. 5, p. 975-989, 2007.

EMBRAPA. **Imagens do sensor MODIS**. Disponível em: <<https://www.modis.cnptia.embrapa.br/geonetwork/srv/pt/main.home>>. Acesso em: 18 fev. 2017.

EMBRAPA. **Satélites em atividade**. Disponível em: <https://www.cnpm.embrapa.br/projetos/sat/conteudo/sensores_operantes.html>. Acesso em: 23 mar. 2017.

ESPÍRITO SANTO JÚNIOR, C.; SILVEIRA, M. A. T. **Atlas Geográfico do Paraná: O uso de novas tecnologias**. Curitiba: Secretaria de Estado da Educação do Paraná, 2007.

ESQUERDO, J. C. D. M. **Utilização de perfis multi-temporais do NDVI/AVHRR no acompanhamento da safra de soja no oeste do Paraná**. 2007. 186 f. Tese (Doutorado em Engenharia Agrícola). Universidade Estadual de Campinas, Campinas/SP, 2007.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHRUSAMY, R. **Advances in Knowledge Discovery & Data Mining**. California: AAAI/MIT, 1996.

FERNANDES, E. **Monitoramento da cultura de cana-de-açúcar no estado de São Paulo por meio de imagens SPOT Vegetation e dados meteorológicos**. 114 f. Dissertação (Mestrado em Engenharia Agrícola). Universidade Estadual de Campinas, Campinas/SP, 2009.

FORMAGGIO, A.R. **Características Agrônômicas e espectrais para sensoriamento remoto de trigo e feijão**. 161 f. Tese (Doutorado em Agronomia). Escola Superior de Agricultura "Luiz de Queiroz" – USP, Piracicaba/SP, 1989.

GUYOT, G. **Signatures spectrales des surfaces naturelles**. Collection "Télédétection Stellitaire". Caen: Édition Paradigme, 1989.

HAIR JUNIOR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise Multivariada de Dados**. Tradução de SANTANNA, A. S.; CHAVES NETO, A. Porto Alegre: Bookman, 2005.

HUETE, A.; LIU, H. Q.; BATCHILY, K.; LEWEEN, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, v.59, p.440-451, 1997.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Produção da Extração Vegetal e da Silvicultura**. Disponível em: <<https://sidra.ibge.gov.br/tabela/>>. Acesso em: 12 dez. 2017.

IPARDES – Instituto Paranaense de Desenvolvimento Social e Econômico. **Webpage**. Disponível em: <http://www.ipardes.gov.br/>. Acesso em: 02 fev. 2018.

JOHANN, J. A. **Calibração de dados agrometeorológicos e estimativa de área e produtividade de culturas agrícolas de verão no estado do Paraná**, 2012. Universidade Estadual de Campinas.

JOHANN, J. A.; BECKER, W. R.; OPAZO, M. A. U.; MERCANTE, E. Uso de imagens do sensor orbital Modis na estimação de datas do ciclo de desenvolvimento da cultura da soja. **Engenharia Agrícola**, Jaboticabal, v. 35, p. 1-15, 2016.

JOHANN, J. A.; ROCHA, J. V.; DUFT, D. G.; LAMPARELLI, R. A. C. Estimativa de áreas com culturas de verão no Paraná, por meio de imagens multitemporais EVI/Modis. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, 2012.

JOHANN, J.A.; ARAÚJO, G.K.D.; ROCHA, J.V. Avaliação do perfil temporal de NDVI decendial do sensor SPOT *Vegetation* em pixels “puros” e “não puros” derivados de mapa de uso da terra realizado a partir do sensor Landsat-5/TM. In: XIV SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO (SBSR). **Anais...** 2009. Natal/RN.

JOSHI, N.; BAUMANN, M.; EHAMMER, A.; FENSHOLT, R.; GROGAN, K.; HOSTERT, P.; JEPSEN, M. R.; KUEMMERLE, T.; MEYFROIDT, P.; MITCHARD, E. T. A.; REICHE, J.; RYAN, C. M.; WASKE, B. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. **Remote Sensing**, v. 8, n. 1, p. 70, 2016.

KUPLICH, T. M.; MOREIRA, A.; FONTANA, D. C. Série temporal de índice de vegetação sobre diferentes tipologias vegetais no Rio Grande do Sul. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.17, n.10, p.1116-1123, 2013.

LORENZETT, C. D. C.; TELÖCKEN, A. V. Estudo Comparativo entre os algoritmos de mineração de dados Random Forest e J48 na tomada de decisão. In: **II Simpósio de Pesquisa e Desenvolvimento em Computação**. **Anais...** Cruz Alta, RS, 2016.

MACEDO, D. C., MATOS, S. N. Extração de conhecimento através da mineração de dados. **Revista de Engenharia e Tecnologia**, v. 2, n. 2, p. 22-30, 2010.

MAPBIOMAS. **Algorithm Theoretical Basis Document & Results**. Disponível em: <<https://s3.amazonaws.com/mapbiomas-ecostage/Base+de+dados/Metodologia/ATBD-R+-+Mata+Atlantica+-+2017-03-17.pdf>>. Acesso em: 5 maio 2017.

MARCUSSI, A. B. **Caracterização do uso do solo e das áreas de preservação permanente, visando a adequação ambiental**. 2010. 66 f. Dissertação (Mestrado em Agronomia). Universidade Estadual Paulista, UNESP. Jaboticabal, 2010.

MEGETO, G. A. S.; OLIVEIRA, S. R. de M.; DEL PONTE, E. M.; MEIRA, C. A. A. Árvore de decisão para classificação de ocorrências de ferrugem asiática em lavouras comerciais com base em variáveis meteorológicas. **Engenharia Agrícola**, Jaboticabal, v. 34, n.3, p.590-599. 2014.

MENESES, P.R.; ALMEIDA, T. **Introdução ao processamento de imagens de sensoriamento remoto**. UnB, CNPq, Brasília, 266p, 2012. Disponível em: <<http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8>>. Acesso em: 18 fev. 2017.

MENGUE, V. P.; FONTANA, D. C. Assessment of spectral-temporal dynamics for mapping the main summer crops in the Rio Grande do Sul State. **Bragantia**, v. 74, p. 122, 2015.

MEZZADRI, F. P. **Cenário atual da pecuária de corte**: aspectos do Brasil com foco no estado do Paraná, ano 2007. Curitiba: SEAB/DERAL/DCA, 2007.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. Viçosa: UFV, 2005.

NASA – National Aeronautics and Space Administration. **About**. Disponível em: <<https://modis.gsfc.nasa.gov/about/>>. Acesso em: 21 fev. 201.

NONATO, R. T. **Aplicação de mineração de dados na identificação de áreas cultivadas com cana-de-açúcar em imagens de sensoriamento remoto no estado de São Paulo**, 2010. 128 f. Dissertação (Mestrado em Engenharia Agrícola). Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2010.

PALOSCHI, R. A. **Software aplicado a modelos de estimativa de produtividade agrícola**. 2016. 98 f. Dissertação (Mestrado em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná, Cascavel, 2016.

PALUDO, A.; JOHANN, J.A.; OPAZO, M.A.U.; RICHETTI, J.; BECKER, W. R. Uso de sensoriamento remoto no mapeamento da cobertura florestal, corpos hídricos e cidades para o estado do Paraná. In: XLIII Congresso Brasileiro de Engenharia Agrícola – CONBEA. **Anais...** Campo Grande, MS, 2014.

PARANÁ. **Governo do Paraná - Website**. Disponível em: <<http://www.cidadao.pr.gov.br>>. Acesso em: 08 jan. 2018.

PARANÁ. **Lei estadual 9.491 de 21 de dezembro de 1990**. Disponível em: <<http://www.legislacao.pr.gov.br/legislacao/listarAtosAno.do?action=exibir&codAto=4924&indice=1&anoSpan=1990&anoSelecionado=1990&isPaginado=true>>. Acesso em: 15 dez. 2017.

PARANÁ. Secretaria da Fazenda do Paraná – SEAB. **Sefanet**. Disponível em: <<http://www.sefanet.pr.gov.br>>. Acesso em: 08 dez. 2017.

PARANÁ. **Transferência de 25% do ICMS aos municípios (cota-parte ICMS)**. Disponível em: <<http://www.liderespublicos.pr.gov.br/2016/11/36/Transferencia-de-25-do-ICMS-aos-municipios-cota-parte-ICMS.html>>. Acesso em: 15 dez. 2017.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acesso em: 15 fev. 2018.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with ERTS. In: Earth Resources Technology Satellite-1 Symposium, 3, Washington, v. 1, p. 309-317, 1973. **Proceedings...** Washington: NASA, 1973.

SATVEG – Sistema de Análise Temporal Da Vegetação. **SatVeg**. Disponível em <<https://www.satveg.cnptia.embrapa.br/satveg>>. Acesso em: 25 out. 2017.

SEAB – Secretaria de Estado da Agricultura e do Abastecimento; DERAL – Departamento de Economia Rural. **Comparativo de área, Produção e Rendimento para a Cultura do trigo Safra 2014/15 e 2015/16**. Disponível em: <<http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=74>>. Acesso em: 17 fev. 2017.

SON, N. T.; CHEN, C. F.; CHEN, C. R.; DUC, H. N.; CHANG, L. Y. A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. **Remote Sensing**, v. 6, n. 1, p. 135-156, 2013.

TAN, P.N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining: Mineração de dados**. Rio de Janeiro: Ciência Moderna, 2009. 932 p.

VICENTE, L. E.; GOMES, D.; VICTORIA, D. C.; GARCON, E. A. M.; BOLFE, E. L.; ANDRADE, R. G.; SILVA, G. B. S. Séries temporais de NDVI do sensor SPOT Vegetation e algoritmo SAM aplicados ao mapeamento de cana-de-açúcar. **Pesquisa Agropecuária Brasileira**, v. 1, p. 1337-1345, 2012.

VIÑA, A.; LIU, W.; ZHOU, S.; HUANG, J.; LIU, J. Land surface phenology as an indicator of biodiversity patterns. **Ecological Indicators**, v. 64, p. 281-288, 2016.

WAGNER, M. M.; RIBEIRO, V.; ZABADAL, J. Mineração de Dados Meteorológicos Empregando Dados de Temperatura: o caso de uma cidade gaúcha. **Interfaces Científicas-Exatas e Tecnológicas**, v. 2, n. 2, p. 57-70, 2016.

WILLMOTT, C. J.; ROBERSON, S. M.; MATSUURA, K. A Refined Index of Model Performance. **International Journal of Climatology**, v. 32, p. 2088-2094, 2012.

WOITKE, R.; MOURÃO, G. N. A Qualidade da Gestão Fiscal nos Municípios Paranaenses. In: 3º Simpósio de Pesquisa e 9º Seminário de Iniciação Científica, 2015, Curitiba. **Caderno PAIC**, v. 16. p. 11-28. Curitiba: FAE, 2015.

XIE, Y.; SHA, Z.; YU, M. Remote sensing imagery in vegetation mapping: a review. **Journal of plant ecology**, v. 1, n. 1, p. 9-23, 2008.

ZHANG, H.; LI, Q.; LIU, J.; JSHANG, J.; DU, X.; ZHAO, L.; WANG, N.; DONG, T. Crop Classification and Acreage Estimation in North Korea Using Phenology Features. **Giscience & Remote Sensing**, v.54, n. 3, p. 1-26, 2017.

ZHONG, L.; GONG, P.; BIGING, G. S. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. **Remote Sensing Environment**, Berlín, v. 140, p. 1-13, 2014.

ZHONG, L.; HU, L.; YU, L.; GONG, P.; BIGING, G. S. Automated mapping of soybean and corn using phenology. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 119, p. 151-164, 2016.

6 ARTIGOS

6.1 Mapeamento semiautomático de áreas agrícolas e permanentes com métricas do perfil espectro-temporal de EVI/MODIS

Resumo: Este trabalho teve como objetivo desenvolver uma metodologia para extrair áreas agrícolas e permanentes no estado do Paraná entre os anos-safra 2013/14 e 2016/17, utilizando séries temporais MOD13Q1 e MYD13Q1. A metodologia proposta neste trabalho seguiu os procedimentos do Processo de descoberta de conhecimento em base de dados – KDD, em que a tarefa de classificação foi executada pelo algoritmo Random Forest. A validação da acurácia espacial dos mapeamentos fez-se com amostras de áreas agrícolas e permanentes (corpos hídricos, alvos urbanos e demais áreas não utilizadas para agricultura), obtendo-se exatidão global entre 94,75 e 98,38% e Kappa superior a 0,92. Além disso, os dados municipais da área agrícola mapeada foram comparados com as estatísticas oficiais brasileira, obtendo-se forte correlação linear, com 0,88 para os anos-safra 2013/14 e 2014/15 e 0,87 para o ano-safra 2015/16. Por fim, a partir dos mapeamentos gerados para cada ano-safra gerou-se um único mapeamento com as áreas agrícolas e permanentes no Paraná. Contudo, a metodologia proposta obteve êxito na automatização do processo de mapeamento de alvos agrícolas e de alvos permanentes, por meio de rotinas escritas em linguagem R. Desse modo, a metodologia de classificação utilizando as métricas elaboradas neste trabalho mostra potencial no que se refere a qualidade dos mapeamentos obtido.

Palavras-chave: Série temporal, Índice de vegetação, Mineração de dados, KDD.

6.1.1 Introdução

O Brasil é o principal produtor e exportador de commodities agrícolas do mundo (FAO, 2015; SCHULTZ et al., 2015). Dentro do cenário nacional, o estado do Paraná possui destaque no setor agropecuário, desempenhando um papel importante na economia. Diante desta realidade, possuir informações precisas referentes ao uso da agropecuária é de suma importância para auxiliar em tomadas de decisões sobre escoamento de safra, armazenamento de grãos, oferta dos produtos agrícolas e preços de mercado (SON et al., 2013; LOBELL et al., 2015; SILVA JUNIOR et al., 2016).

Essas informações podem ser obtidas por meio de mapeamentos de áreas agrícolas (DOS SANTOS et al., 2014; GRZEGOZEWSKI et al., 2016; ZHONG et al., 2016), que fornecem dados sobre estimativas de área semeada e participação estadual por cultura agrícola. No estado do Paraná, os mapeamentos são anualmente realizados pela Secretaria da Agricultura e do Abastecimento (SEAB) e pela Companhia Nacional de Abastecimento (CONAB). Entretanto, com o desenvolvimento de técnicas de sensoriamento remoto, a elaboração de mapeamentos vem se aprimorando, tanto na qualidade, quanto no tempo de elaboração destes (ZHANG et al., 2017).

Embora tais técnicas tenham facilitado a execução de mapeamentos, também surgiram desafios relacionados à análise das imagens orbitais. Um destes desafios é a distinção de alvos terrestres similares, ou seja, alvos que apresentam refletâncias

semelhantes, por exemplo, alvos de pastagem. Cechim Junior et al. (2017) apresentam que normalmente esses alvos ocorrem certa confusão espectral com áreas de cana-de-açúcar e alvos de soja se confundem com o milho (GRZEGOZEWSKI et al., 2016; BECKER et al., 2017).

Para amenizar esse problema, podem ser utilizados índices de vegetação, como o Normalized Difference Vegetation Index (NDVI) (ROUSE et al., 1973) e o Enhanced Vegetation Index (EVI) (HUETE et al., 1997). Esses índices facilitam a distinção de alvos de vegetação dos demais, pois utilizam em sua composição bandas do espectro eletromagnético que apresentam maiores índices para alvos de vegetação. Outra estratégia para distinguir alvos terrestres é a utilização de séries temporais EVI/MODIS, conforme adotado por Couto Jr., Carvalho Jr. e Martins (2012), Johann et al. (2012), Zhong et al. (2016), Silva Junior et al. (2017) e Zhang et al. (2017), pois permitem a identificação e a diferenciação de culturas agrícolas em função de cada uma delas ter um padrão espectro-temporal distinto (JOHANN et al., 2016).

Além da distinção de alvos similares, outro desafio relacionado à análise de imagens orbitais refere-se ao tempo necessário para o processamento dos dados. Esse desafio é agravado ao se trabalhar com séries temporais, pois aumenta consideravelmente o número de imagens a processar. Uma solução para diminuir o tempo de processamento é por meio da automatização dos processos, de modo que este possa ser realizado várias vezes ao longo dos anos, tornando-o mais econômico e dinâmico (ZHONG et al., 2016).

Todavia, mesmo com a automatização dos processos, o uso de séries temporais para grandes áreas (escalas regional, estadual, nacional e global) gera uma enorme quantidade de dados. Por conta disso, é imprescindível a utilização de técnicas e ferramentas computacionais para obter informações relevantes e em menor tempo. Logo, este trabalho aplicou o processo de Descoberta de Conhecimento em Base de Dados ou “Knowledge Discovery in Databases – KDD”, que, segundo Fayyad et al. (1996), é um processo não trivial de descoberta de padrões válidos, úteis, novos e acessíveis em grandes bases de dados, ou seja, a finalidade é extrair conhecimento a partir de um banco de dados.

Diante disto, o objetivo deste estudo foi criar uma metodologia semiautomática para mapear as áreas com culturas agrícolas e alvos permanentes para o estado do Paraná para os anos-safra 2013/14, 2014/15, 2015/16 e 2016/17, utilizando métricas extraídas, de maneira semiautomática, do perfil espectro-temporal de cada pixel da imagem de índice de vegetação EVI/MODIS.

6.1.2 Material e métodos

6.1.2.1 Área de estudo

O local de estudo compreende o estado do Paraná, localizado na região sul do Brasil, entre os paralelos 22° 29' S e 26° 43' S e entre os meridianos 48° 2' W e 54° 38' W, e subdividido em 399 municípios distribuídos em 10 mesorregiões (Figura 18). O estado possui quatro tipos de clima, sendo estes o Aw, Cfa, Cfb e Cwa (APARECIDO et al., 2016).

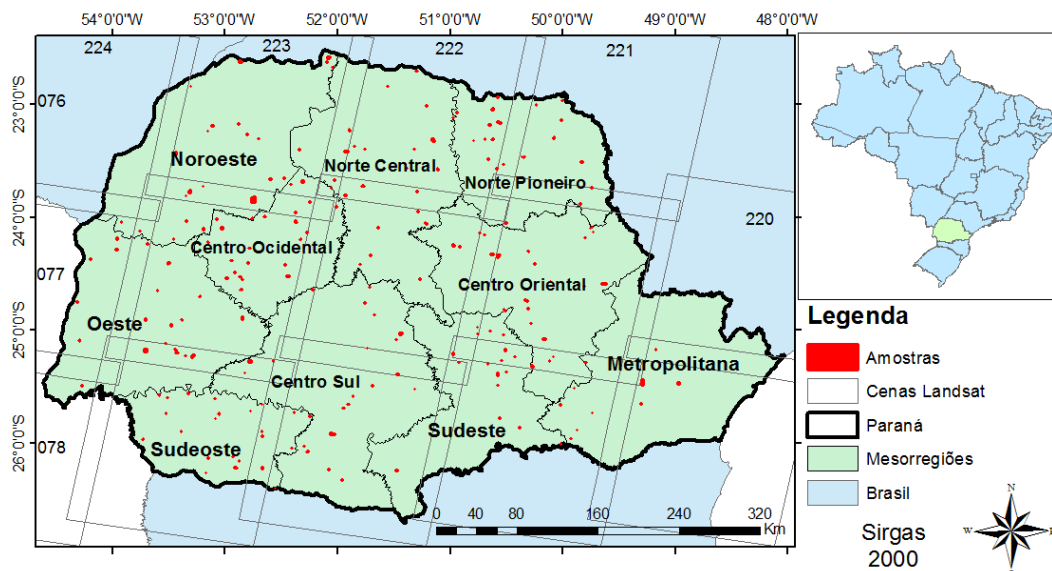


Figura 18 Localização do Estado do Paraná com as 10 mesorregiões e path-row do satélite Landsat-8.

6.1.2.2 Processo de Descoberta de conhecimento em base de dados

Os procedimentos adotados para mapeamento de áreas agrícolas e áreas com alvos permanentes para cada ano-safra (2013/14, 2014/15, 2015/16 e 2016/17) são ilustrados no fluxograma (Figura 19). A metodologia para construção de cada mapeamento foi dividida em cinco etapas, de acordo com o processo KDD, sendo estas: seleção, pré-processamento, formatação/transformação, mineração de dados e avaliação/interpretação.

Seleção

Os dados utilizados no processo são oriundos de séries temporais e da extração de pixels puros, ou seja, pixels sem mistura espectral de demais alvos (GRZEGOZEWSKI et al., 2016). Para a construção da série temporal utilizaram-se imagens sensor MODIS (MOD13Q1 e MYD13Q1). Especificamente, utilizou-se o índice de vegetação EVI (Equação 1), com resolução espacial de 250 metros e resolução temporal de 8 dias, adquiridas gratuitamente no

site da Empresa Brasileira de Pesquisa Agropecuária (Embrapa Informática Agropecuária, 2017). Os pixels puros foram selecionados em mosaicos estaduais de cenas do Landsat-8/OLI na composição RGB-564, com resolução espacial de 30 metros, as quais foram adquiridas gratuitamente no site United States Geological Survey (USGS, 2017). Os pixels puros foram empregados na construção do conjunto treinamento do processo KDD e posteriormente na validação dos mapeamentos.

$$EVI = 2,5 \frac{NIR - RED}{1 + NIR + 6RED - 7,5BLUE} \quad \text{Eq. (1)}$$

em que:

NIR - Reflectância na região espectral do infravermelho próximo;

RED - Reflectância na região espectral do vermelho;

BLUE - Reflectância na região espectral do azul.

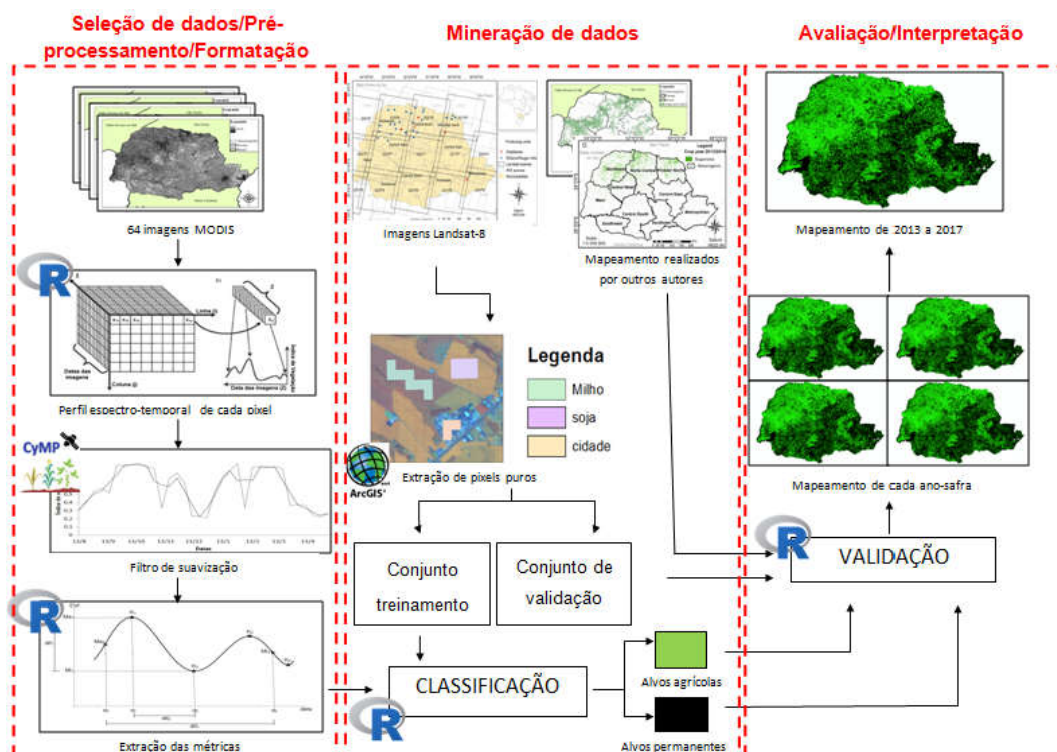


Figura 19 Fluxograma para a separação de alvos agrícolas e alvos permanentes, com as divisões do processo KDD.

Cada ano-safra (2013/14 a 2015/16) contabilizou 64 imagens EVI/MODIS, entre 05 de agosto do ano1 a 27 de dezembro do ano2, de forma que abrangesse todo o ciclo das culturas utilizadas no estudo, totalizando 202 imagens em todo o estudo.

Pré-processamento

O pré-processamento dos dados contemplou as etapas de filtragem da série temporal e formação de um cubo espectro-temporal do índice de vegetação EVI/MODIS.

A aplicação da filtragem é necessária, pois as imagens que compõem a série temporal podem apresentar algumas interferências de ruídos ocasionadas por nuvens, falhas do detector, dentre outros (MENGUE; FONTANA, 2015), podendo gerar informações falsas. Em função disto, aplicou-se o filtro de suavização Savitzky-Golay, com auxílio do software CyMP v.1.07 (PALOSCHI, 2016) (Figura 19).

Posteriormente, estas imagens foram empilhadas (cubo temporal) de forma crescente, em que a primeira imagem corresponde à data mais antiga (Figura 20a). Desse modo, cada pixel da imagem possui 64 respostas de EVI, possibilitando a construção do perfil espectro-temporal para cada pixel da imagem (Figura 20b).

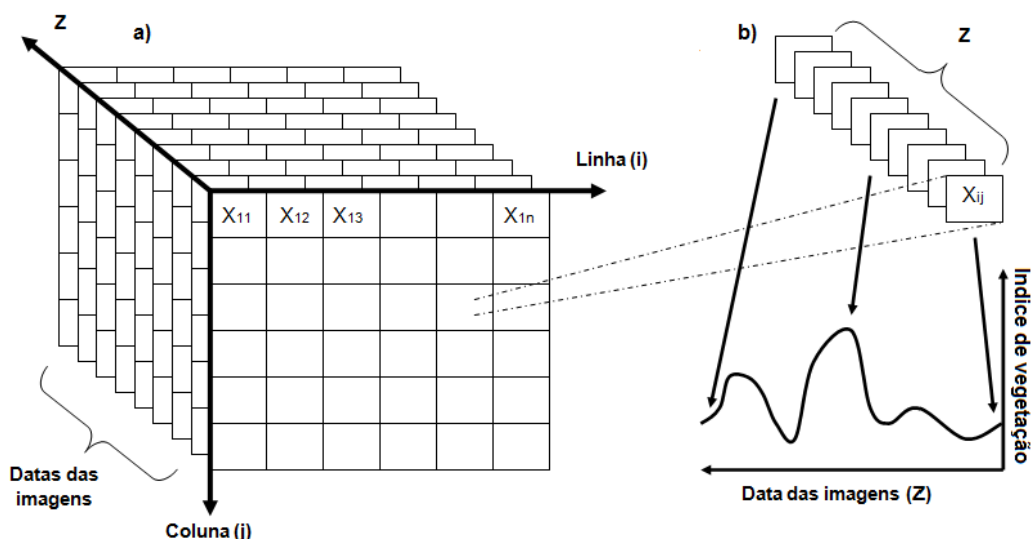


Figura 20 Representação do cubo de imagem (a) utilizada para geração do padrão espectro-temporal do índice de vegetação (b).

Formatação/transformação

Utilizando o perfil espectro-temporal filtrado de EVI/MODIS como dado de entrada, elaboraram-se métricas para auxiliar na separação dos alvos, representados esquematicamente na Figura 21 e definidos na Tabela 9. Esse processo foi desenvolvido por meio de rotinas elaboradas no software R versão 3.3.1 (R DEVELOPMENT CORE TEAM, 2017).

Desse modo, para cada um dos pixels da imagem MODIS no estado do Paraná, extraíram-se as 15 métricas da Tabela 9, dando origem a uma matriz com 5.501.808 linhas

(representando cada pixel) e 15 colunas (representando as métricas). Esta matriz de dados foi utilizada como dado de entrada para a próxima etapa do KDD.

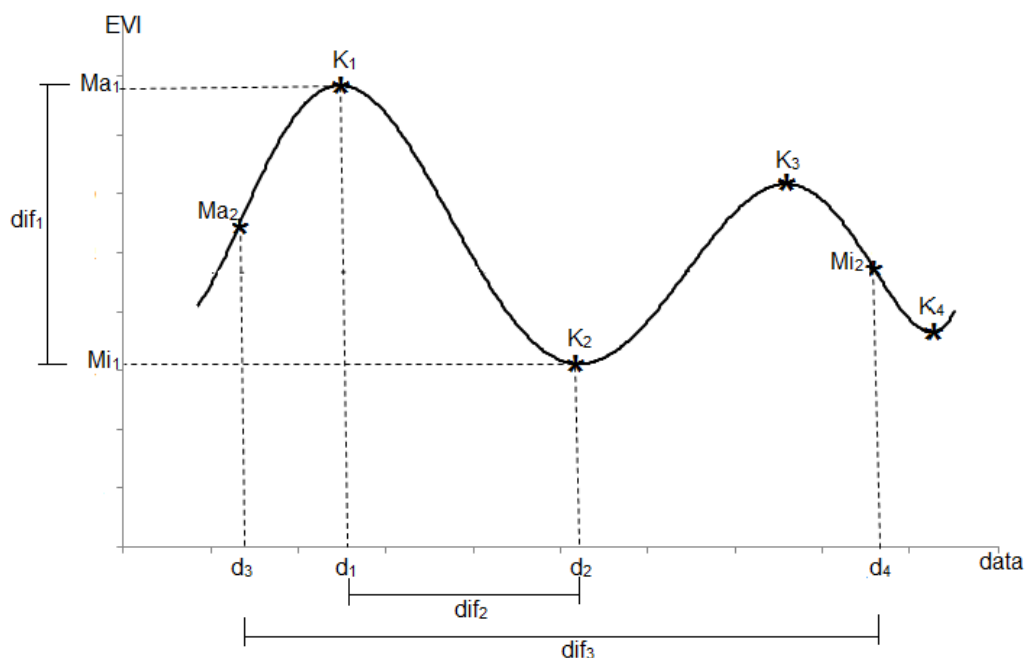


Figura 21 Forma de extração das métricas de uma série temporal.

Tabela 9 Métricas extraídas de uma série espectro-temporal de índice de vegetação (EVI)

Métricas	Descrição
Ma_1	Valor do máximo EVI registrado em cada pixel
Mi_1	Valor do mínimo EVI registrado em cada pixel
dif_1	Diferença entre o máximo e o mínimo EVI
Ma_2	Valor da maior taxa de variação do EVI registrado em cada pixel
Mi_2	Valor da menor taxa de variação do EVI registrado em cada pixel
K_i	Pontos de mudança da taxa de variação
d_1	Data que ocorre o máximo EVI para cada pixel
d_2	Data que ocorre o mínimo EVI para cada pixel
d_3	Data que ocorre a maior taxa de variação do EVI para cada pixel
d_4	Data que ocorre a menor taxa de variação do EVI para cada pixel
dif_2	Diferença entre as datas que ocorre o máximo e o mínimo EVI
dif_3	Diferença entre as datas que ocorre o Ma_2 e o Mi_2
Média	Média do EVI de todas as datas para cada pixel
Dp	Desvio padrão de cada pixel na série temporal
CV	Coefficiente de variação de cada pixel ao na série temporal

Mineração de dados

Nesta fase do trabalho, utilizou-se a tarefa de classificação com o objetivo de encontrar uma função ou regras de decisão para mapear um conjunto de dados em categorias pré-determinadas, denotadas por classes (por exemplo, áreas com culturas agrícolas ou alvos permanentes). Uma forma de encontrar estas regras é por meio de uma

árvore de decisão. Esta fornece uma modelagem alternativa à estatística clássica, por conta da sua representação simbólica e interpretável do modelo que pode ser utilizada tanto para predição quanto para conhecer a estrutura predita (MEGETO et al., 2014). Assim, conhecidas as regras de decisão, elas foram aplicadas a novos dados (pixels), buscando prever a qual classe elas pertenciam (MACEDO; MATOS, 2010).

O algoritmo Random Forest foi empregado para realizar a tarefa de classificação. Sua principal característica é a construção de várias árvores de decisão, usando um subconjunto aleatório de variáveis obtidas do conjunto de dados originais, em que cada subconjunto gera uma árvore de decisão. Para determinar a classe final de uma instância (pixel) é verificado qual foi a classe mais votada dentre todas as árvores de decisão geradas, sendo este processo repetido para todas as instâncias (BREIMAN, 2011; LORENZETI, 2016). Segundo Lorenzetti (2016), esse algoritmo apresenta resultados superiores do que utilizando uma única árvore de decisão, porém demanda maior tempo computacional para a execução da rotina.

Para a aplicação desse algoritmo, elaborou-se um conjunto de amostras (Figura 18) formado por pixels puros de cada alvo de interesse (Figura 22), obtidos por meio de uma análise visual das imagens Landsat-8. Em virtude da diferença da resolução espacial das imagens MODIS e Landsat-8, utilizou-se uma grade de pixels MODIS (em que cada pixel possui 250 x 250m) sobre as imagens Landsat-8 com o intuito de extrair apenas regiões que compreendiam um pixel inteiro do sensor MODIS (Figura 22). Estes procedimentos foram operacionalizados no software Arcgis 10.3.

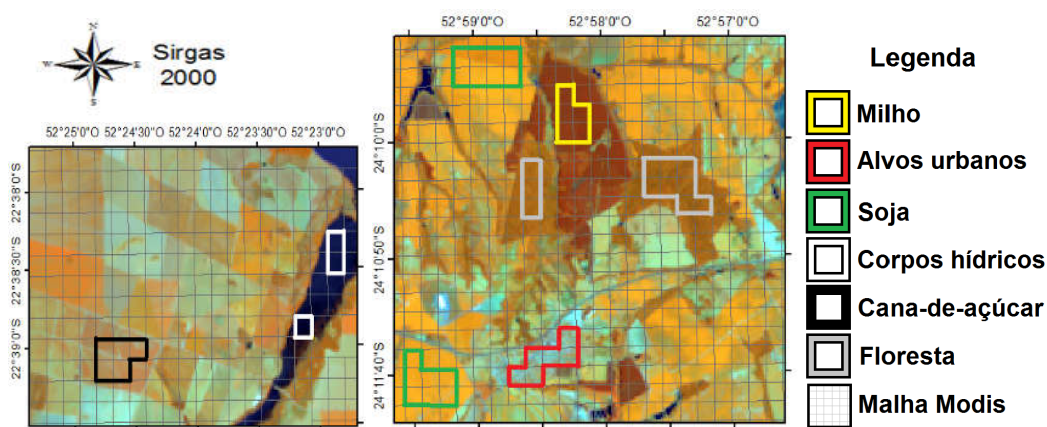


Figura 22 Método de seleção de alvos de culturas agrícolas e alvos permanentes na composição RGB564 com o detalhamento da malha MODIS.

Com os pixels puros extraídos construiu-se um conjunto de amostra (Figura 19), em que 75% desse conjunto foi utilizado para elaborar o conjunto de treinamento (traindata), sendo este utilizado como entrada para o processo de classificação, realizada no software R. Geraram-se assim regras de decisão para classificação dos pixels do traindata e posteriormente essas regras foram utilizadas para classificar o restante dos pixels do Paraná. Os demais 25% dos dados (validation data) foram utilizados para validação da classificação,

gerando, como produto final, a classificação (ou mapeamento) dos pixels de áreas de agricultura e os pixels de alvos permanentes (Figura 19).

Avaliação e interpretação

O processo de avaliação dos mapeamentos segmentou-se em quatro etapas. Na etapa I determinou-se o melhor conjunto de amostras; na etapa II validaram-se os mapeamentos de cada ano-safra utilizando o validation data. Para a etapa III elaborou-se um mapeamento único, utilizando os quatro mapeamentos de cada ano-safra, e se validou este mapeamento por meio do Erro de omissão (EO), reclassificando-o caso o EO apresente valor superior a 40%. Por fim, na etapa IV validou-se o mapeamento único gerado na etapa III após as reclassificações.

A etapa I foi utilizada para testar a melhor configuração de dados para os mapeamentos, sendo que para isso utilizaram-se três conjuntos de dados (CA1, CA2 e CA3) para o ano-safra 2013/14. Os testes foram necessários em função das amostras de cana-de-açúcar apresentarem grande variação do perfil espectro-temporal de EVI entre os pixels puros selecionados. No primeiro conjunto de dados (CA1) utilizaram-se pixels de pastagem, soja, floresta, reflorestamento, alvos urbanos, corpos hídricos e cana-de-açúcar (Figura 18). Para o CA2 utilizaram-se as mesmas amostras de CA1, com exceção das amostras de cana-de-açúcar, pois agora consideraram-se apenas os pixels de cana-de-açúcar com padrão espectro-temporais semelhantes (cana-de-açúcar modificada). Por fim, o CA3 utilizou o conjunto de dados de CA1, sem as amostras da cana-de-açúcar.

A tomada de decisão baseou-se na Exatidão Global (EG) (Equação 2) e índice Kappa (IK) (Equação 3), calculados com base no conjunto de validação de pixels puros para o ano-safra 2013/14. Além disso, foram utilizados também mapeamentos de soja e milho (GRZEGOZEWSKI et al., 2015), de cana-de-açúcar (CECHIM JUNIOR et al., 2017) e de alvos permanentes (PALUDO et al., 2014) para comparação espacial, ou seja, verificou-se a proporção de áreas destes mapas que possuíam alvos diferentes dos encontrados no mapeamento do ano-safra 2013/14 para cada conjunto de dados (CA1, CA2, CA3).

$$EG = \frac{A}{m} * 100 \quad \text{Eq. (2)}$$

$$IK = \frac{\left[m * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_i * x_j) \right]}{\left[m^2 - \sum_{i=1}^r (x_i * x_j) \right]} \quad \text{Eq. (3)}$$

em que:

A - acerto geral (pontos amostrais com acertos);

m - número de pontos amostrais;

x_{ij} - observação na linha i e coluna j ;

x_i - total marginal da linha i ;

x_j - total marginal da coluna j .

Após determinar o melhor conjunto de dados (CA1, CA2, CA3), este foi utilizado para elaborar os mapeamentos de cada ano-safra, iniciando a etapa II, de modo que a acurácia espacial foi avaliada com base no IK, EG, Erro de omissão (EO) e Erro de comissão (EC).

Também se fez a comparação da área municipal dos mapeamentos deste estudo com os dados oficiais do Instituto Brasileiro de Geografia e Estatística (IBGE). Esses dados se referem à área agrícola (soja, milho e feijão de primeira safra, além das lavouras de cana-de-açúcar) de cada município do Paraná para os anos-safra 2013/14, 2014/15 e 2015/16, pois para o ano-safra 2016/17 os dados oficiais ainda não foram divulgados. A análise foi realizada pelos indicadores estatísticos: erro médio (ME) (Equação 4), erro médio quadrático (RMSE) (Equação 5), índice de concordância aprimorado de Willmott (dr) (WILLMOTT; ROBERSON; MATSUURA, 2012) (Equação 6) e coeficiente de correlação de Pearson (r).

$$ME = \frac{1}{n} * \sum_{i=1}^n (O - E) \quad \text{Eq. (4)}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^n (O - E)^2} \quad \text{Eq. (5)}$$

$$dr = 1 - \frac{\sum_{i=1}^n |E - O|}{2 \sum_{i=1}^n (O - \bar{O})} \quad \text{Eq. (6)}$$

em que:

n - número de observações;

O - área da cultura usada como referência;

E - área da cultura obtida nos mapeamentos;

\bar{O} - área da cultura média usada como referência.

O coeficiente de correlação foi avaliado de acordo com Callegari-Jaques (2003), que classifica $|r|$ maior que 0,9 como muito forte correlação, $|r|$ entre 0,7 e 0,9 como forte correlação, $|r|$ entre 0,4 e 0,7 com moderada correlação, $|r|$ entre 0,2 e 0,4 fraca correlação e $|r|$ menor que 0,2 como muito fraca correlação.

Após a elaboração dos mapeamentos de cada ano-safra (2013/14 a 2016/17), iniciou-se a etapa III, na qual construiu-se um único mapeamento (Figura 23), sendo este

segmentado em cinco classes: a classe 1 representava os pixels caracterizados como alvo permanente nos quatro anos-safra; a classe 2, os pixels caracterizados como alvo permanente em três dos quatro anos-safra; a classe 3 abrangia os pixels caracterizados como alvo permanente em dois anos-safra; a classe 4 possuía os pixels como alvo permanente em apenas um ano-safra; e a classe 5 continha os pixels caracterizados apenas como área agrícola nos quatro anos-safra (Figura 23a).

Após esta segmentação, cada uma das cinco classes, foram reclassificadas entre alvos permanentes e alvos agrícolas, por meio de análise visual utilizando um mosaico de imagens Landsat-8 na composição colorida falsa-cor RGB-564, gerando por fim, um único mapeamento de alvos permanentes e área agrícola, divididos em cinco classes para todo o Paraná (Figura 23b).

Este mapeamento único foi avaliado por meio do Erro de omissão (EO), tendo como referência um conjunto de amostras de 400 pixels, obtidos por meio de amostragem aleatória estratificada, sendo 50% sorteada sobre a área agrícola e 50% sobre os alvos permanentes, e analisados sobre o mosaico de imagens Landsat-8.

Para melhorar a qualidade e acurácia espacial deste mapeamento, utilizou-se o EO de cada classe, de modo que, se a classe apresentasse um EO superior a 40%, ela era reclassificada (Figura 23c). O processo de reclassificação utilizou o intervalo interquartil dos perfis espectro-temporal de EVI/MODIS da classe com menor EO (Figura 23c). Assim, se até 75% do perfil espectro-temporal do EVI de cada pixel estava contido neste intervalo, este pixel era reclassificado com a classe de menor EO (Figura 23d). Este processo foi repetido até que não houvessem mais classes com EO superior a 40%.

Em seguida, as classes com mesmo alvo foram agrupadas, obtendo com isso um mapeamento único final segmentado em 2 classes (alvo permanente e alvo agrícola). Por fim, na etapa IV o mapeamento único foi avaliado por meio do EG e IK, EO e EC, utilizando o mesmo procedimento de sorteio de 400 amostras, descrito anteriormente.

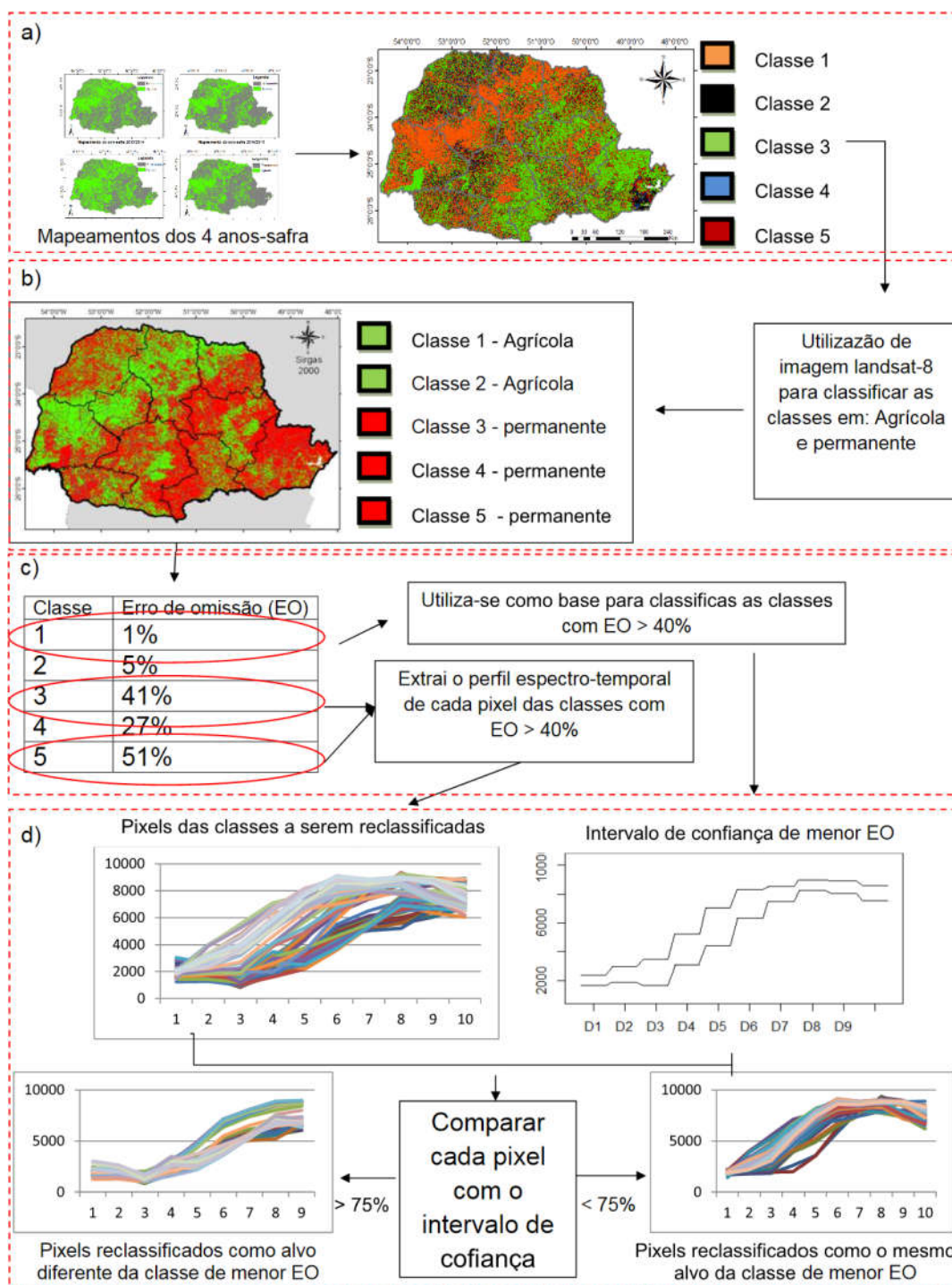


Figura 23 Fluxograma para reclassificação das classes com erro de omissão superior a 40%. a) processo de segmentação em cinco classes; b) Verificação dos alvos de cada classe; c) Processo de decisão de quais classes seriam reclassificadas e qual classe seria utilizada como base para elaborar o intervalo interquartil; d) comparação dos pixels das classes reclassificadas com o intervalo interquartil e definição da nova classe.

6.1.3 Resultados

6.1.3.1 Método de verificação do conjunto de dados para os mapeamentos

O conjunto de dados CA2 (Tabela 10) apresentou os maiores valores de IK e EG, como também o menor percentual de diferença (14,38%) quando comparado ao mapeamento de soja e milho realizado por Grzegozewski et al. (2015). Desse modo, este conjunto de dados foi utilizado para os mapeamentos de todos os anos-safra (2013/14 a 2016/17).

Tabela 10 Valores de exatidão global (EG), índice Kappa (IK), e taxa de erro para os três conjuntos de amostras testados

Testes	Conjunto validação		Diferença % (mapa de soja e milho) ¹	Diferença % (mapa de cana- de-açúcar) ²	Diferença % (mapa de floresta, alvos urbanos e corpos hídricos) ³
	Exatidão Global	Índice Kappa			
CA1	96,81	0,95	15,48	7,23	19,57
CA2	98,38	0,98	14,38	14,84	16,11
CA3	97,38	0,96	26,67	48,10	12,73

¹ Grzegozewski et al. (2016); ² Cechim Junior et al. (2017); ³ Paludo et al. (2014)

6.1.3.2 Mapeamentos dos alvos para os anos-safra 2013/14 a 2016/17

Utilizando as amostras de CA2 efetuou-se a metodologia proposta, classificando os alvos permanentes e as áreas agrícolas para os anos-safra 2013/14 a 2016/17 (Figura 24). Há maior concentração de áreas agrícolas na região que se estende do oeste ao nordeste do estado, abrangendo as mesorregiões Oeste, Cento ocidental, Norte central e Norte pioneiro (Figura 24), corroborando com resultados apresentados por Johann et al. (2012), França Junior; Villa; Villa (2015), Johann et al. (2016) e Zhong et al. (2016), que descrevem essa região como a que possui a maior concentração das culturas soja e milho. Identificam-se também grandes áreas destinadas a alvos permanentes, como, por exemplo, a região do parque nacional do Iguaçu, localizado na região sudoeste do estado e na mesorregião metropolitana de Curitiba.

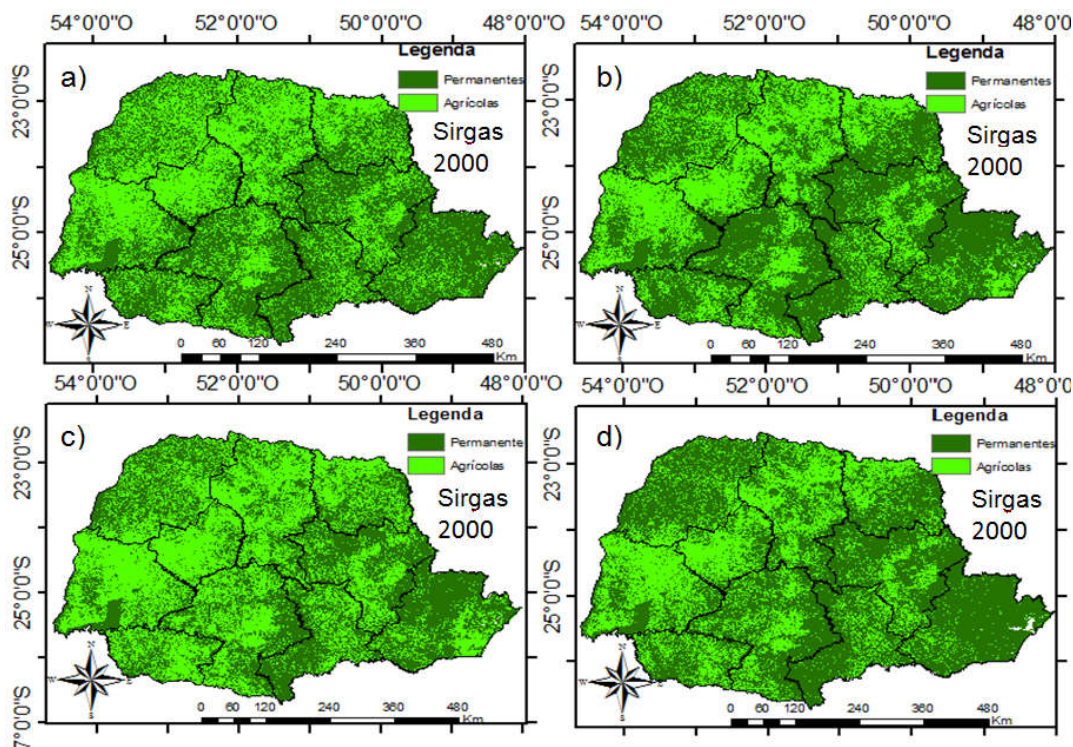


Figura 24 Mapeamento de áreas agrícolas e alvos permanentes para os anos-safra 2013/14 (a), 2014/15 (b), 2015/16 (c) e 2016/17 (d).

A acurácia espacial dos mapeamentos, avaliadas pelo EG e IK, revelou EG superior a 98,34 para todos os mapeamentos (Figura 25), o que é considerado aceitável por Foody (2002). Landis e Koch (1977) consideram um índice Kappa excelente quando superior a 0,80, sendo esta condição cumprida em todos os mapeamentos, variando entre 0,95 e 0,99 (Figura 25).

Os resultados obtidos são semelhantes aos encontrados na literatura, visto que Zhong et al. (2016) obtiveram 87,2% de Exatidão Global e Kappa de 0,80, utilizando uma metodologia automatizada para mapear soja e milho com imagens do sensor MODIS. Formaggio et al. (2010), utilizando Mineração de dados e Object-Based Image Analysis (OBIA), obtiveram uma exatidão global de 94% e Kappa de 0,87 para o mapeamento de cana-de-açúcar no Brasil com imagens Landsat. Grzegozewski et al. (2016) calcularam o IK e a EG como base nos dados obtidos de imagens Landsat, ou seja, com maior resolução espacial, resultando em uma exatidão global de 92,25% para milho e 87,75% para a soja, e IK de 0,79 e 0,75 para milho e soja, respectivamente.

Para o erro de omissão (EO) todos os valores foram inferiores a 5,5%, e os erros de comissão (EC) apresentaram valores variando de 0,2 a 7,5% (Figura 25). Para Grzegozewski et al. (2015), os valores de EO e EC variaram de 4,81 a 10,33% para o milho e de 10,47 a 12,87 para soja. Sendo assim, no presente estudo foram obtidos resultados tão bons ou melhores que os descritos na literatura.

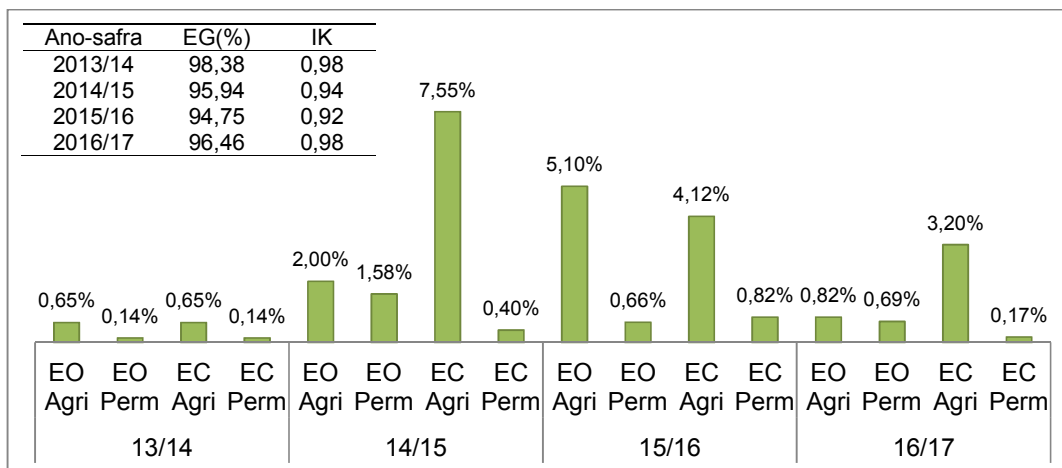


Figura 25 Erro de omissão (EO) e Erro de comissão (EC) para alvo agrícola (Agri) e alvos permanentes (Perm) nos anos-safra 2013/14 a 2016/17.

O mesmo ocorreu com os indicadores estatísticos; no entanto, os mapeamentos subestimaram os dados oficiais, sendo que o erro médio (ME) indicou que a área obtida pelo mapeamento foi superestimada entre 3.256 ha (2014/15) e 10.873 ha (2015/16) quando comparado aos dados oficiais. A raiz do erro médio (RMSE) evidenciou que a dimensão do erro produzido entre as comparações de estimativas de área teve uma variação entre 10.195 ha (ano-safra 2014/15) e 16.886 ha (ano-safra 2015/16) (Figura 26).

Esta superestimação também é evidenciada na Figura 27, pois no ano-safra 2013/14 a área mapeada foi 39,4% maior que os dados oficiais e nos anos-safra 2014/15 e 2015/16 esse aumento foi de 19 e 64%, respectivamente. Entretanto, a superestimação dos alvos agrícolas pode ser justificada, visto que os dados oficiais levam em conta apenas as culturas de soja, milho 1ª safra, feijão 1ª safra e cana-de-açúcar, e o mapeamento estimou as áreas destas culturas e também a área de pastagem do Paraná, que não possui informação de dado oficial.

Entretanto, os mapeamentos obtiveram alta correlação (r) com os dados oficiais, sendo este de 0,88 para os anos-safra 2013/14 e 2014/15 e 0,87 para o ano-safra 2015/16 (Figura 26). Cechim Junior et al. (2017), que propuseram uma metodologia para mapear áreas de cultivo de cana-de-açúcar entre 2010/11 e 2013/14, no Paraná, utilizando imagens dos satélites Landsat/TM/OLI e IRS/LISS-3, encontraram correlações com os dados oficiais do IBGE que variou de moderada a forte ($0,64 \leq r_s \leq 0,80$). Melo (2017) realizou um mapeamento de áreas de pastagem no estado de São Paulo e obteve um (r) de 0,83. Grzegozewski et al. (2016) obtiveram valores de r de 0,72 para o mapeamento de milho no estado do Paraná e 0,95 para mapeamento de soja, sendo ambos referentes ao ano-safra 2012/13.

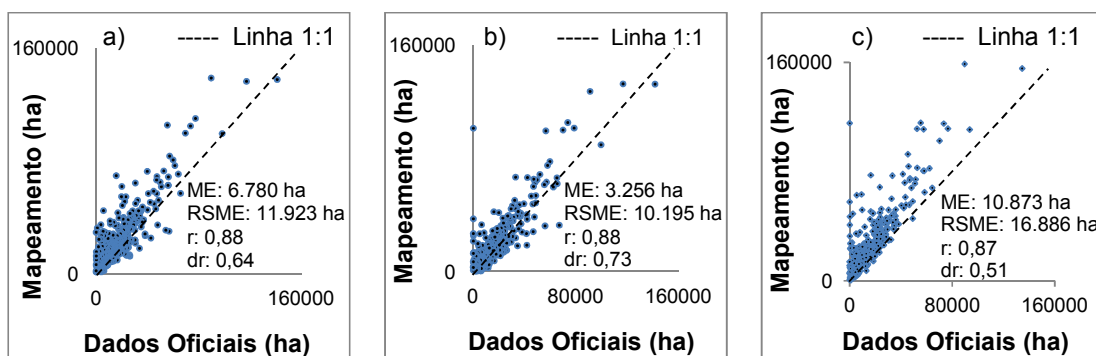


Figura 26 Gráficos de dispersão entre dados oficiais de área agrícola de cada município e os mapeamentos realizados para os anos-safra 2013/2014 (a), 2014/2015 (b) e 2015/2016 (c).

O índice de concordância aprimorado de Willmott (dr) indicou valores de 0,64, 0,73 e 0,51 para os anos-safra 2013/14, 2014/15 e 2015/16, respectivamente; portanto, este possui em média 0,63.

Ao analisar a correlação a nível municipal (Figura 26) tem-se que, na maioria dos municípios, existe correlação entre os dados oficiais e os dados estimados, embora alguns se destaquem por apresentar valores que destoam dos dados oficiais. Um exemplo disto é o município de Irati (Figura 27), que foi o município mais subestimado no ano safra 2013/14 e o terceiro no ano safra 2015/16. Outro exemplo é o município de Guaratuba, que apresentou as maiores diferenças das áreas agrícolas informadas pelos órgãos oficiais com as áreas agrícolas obtidas no mapeamento (Figura 27).

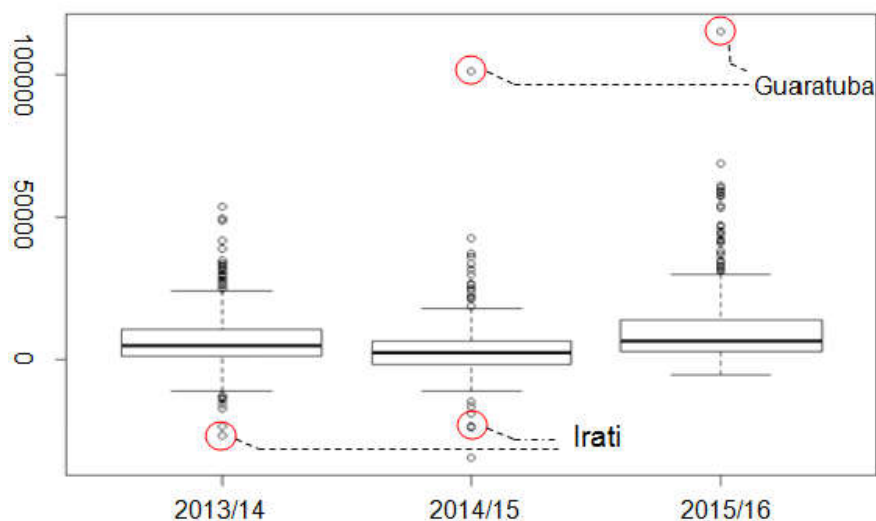


Figura 27 Boxplot das diferenças dos dados oficiais e o mapeamento para os anos-safra 2013/14, 2014/15 e 2015/16.

De maneira geral, o mapeamento superestimou os dados oficiais (Figura 28). Este valor elevado da área agrícola dos mapeamentos ocorreu por conta das áreas de pastagem, que para o mapeamento elaborado foram consideradas como áreas agrícolas; no entanto, os dados oficiais excluem essas áreas do cálculo de áreas agrícolas.

Além disso, os mapeamentos dos anos-safras 2014/15 e 2015/16 classificaram, erroneamente, regiões do litoral paranaense e do parque de Ilha grande como alvo agrícola. Para amenizar esse erro, elaborou-se um único mapeamento para os quatro anos-safra analisados.

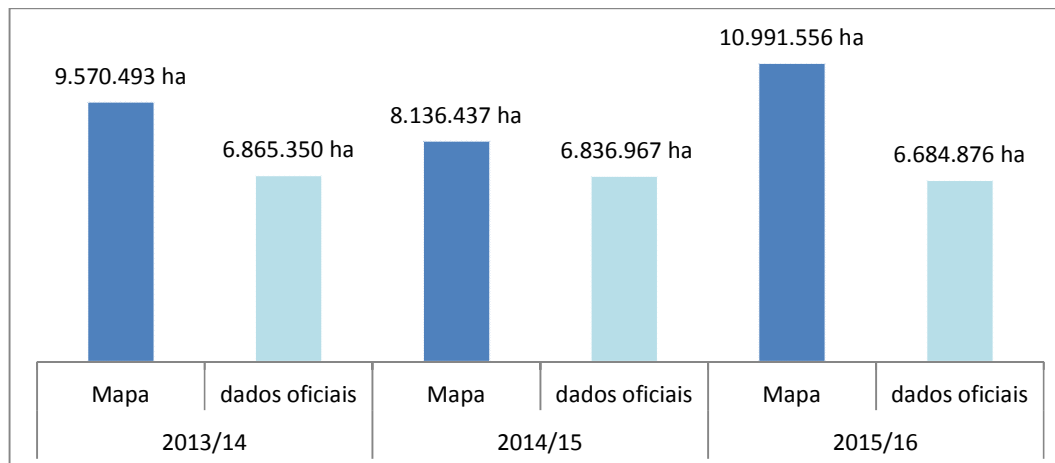


Figura 28 Área total agrícola (ha) do mapeamento elaborado e dos dados oficiais para o Paraná nos anos-safra 2013/14 a 2015/16.

6.1.3.3 Mapeamento único dos anos-safra

A união dos mapeamentos do período estudado resultou em um único mapa de área agrícola e alvos permanentes entre 2013 e 2017, segmentado em cinco classes (Figura 29).

Desse modo a classe 1 foi reclassificada como alvo permanente (Figura 30b), e a classe 5 foi reclassificada como área agrícola, pois nesses casos todos os anos-safra apresentavam essa classificação (Figura 30f).

Para determinar as classes 2, 3 e 4 foi realizada uma análise visual, verificando-se que estas classes eram formadas, na maioria dos casos, por pixels mistos, ou seja, áreas que possuíam dois ou mais alvos em um só pixel. Tal condição se dá pela média resolução espacial do MODIS e pelo fato do grande número de pequenas propriedades no estado, pois, de acordo com Zhong et al. (2016) e Silva Junior et al. (2017), as pequenas propriedades possuem misturas espectrais, o que dificulta a realização do mapeamento. Contudo, verificou-se que a classe 2 era formada por pixels que possuíam em sua composição uma maior porcentagem de alvos permanentes. Por decorrência disto, esta classe foi reclassificada como alvo permanente (Figura 30c).

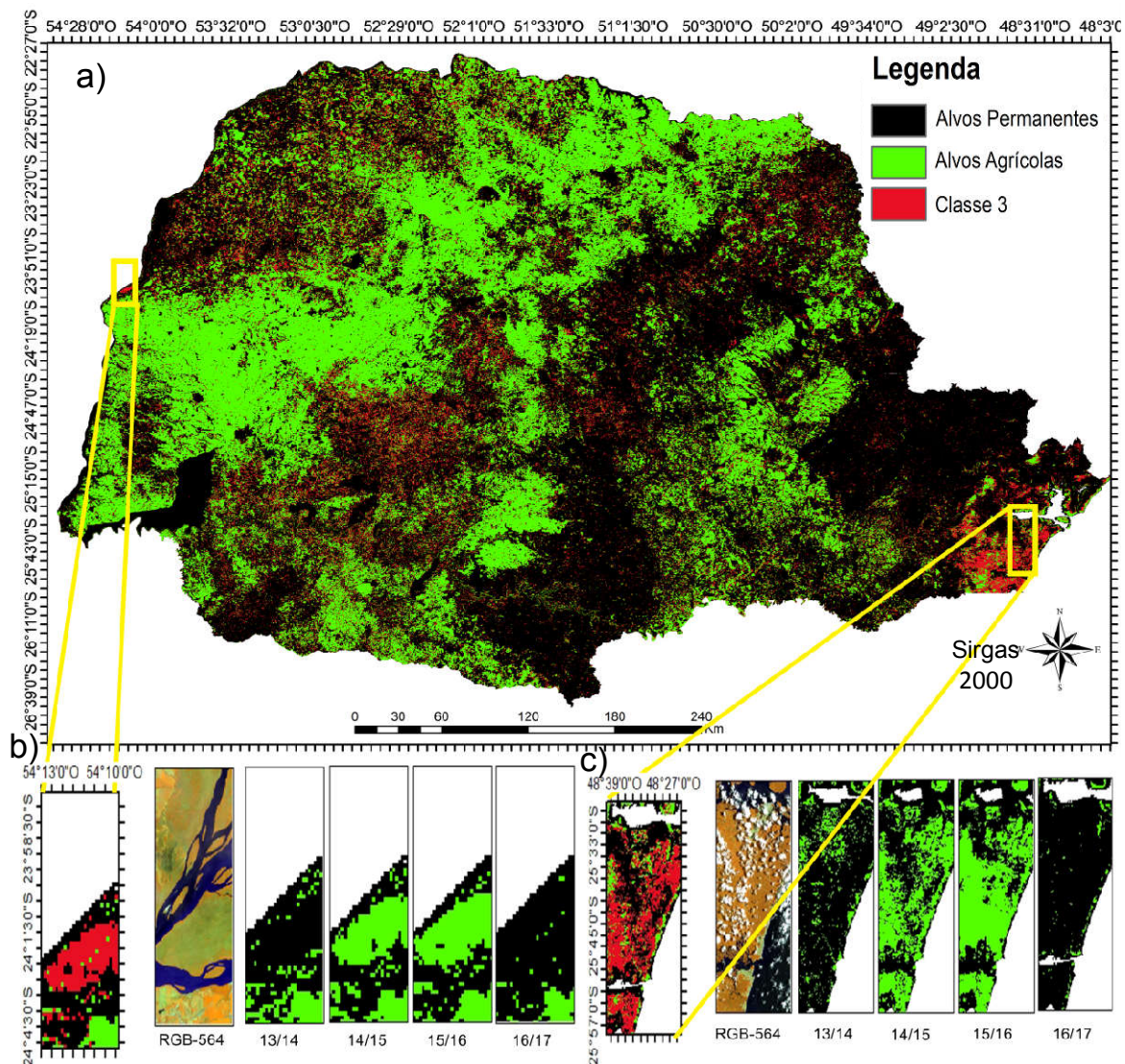


Figura 29 Mapeamento de áreas agrícolas e alvos permanentes do estado do Paraná entre os anos-safra 2013/14 e 2016/17 (a) com ênfase nas áreas com maior concentração da classe 3: Parque nacional de Ilha Grande (b); e litoral paranaense (c).

A classe 3 era formada por pixels mistos, além de pixels localizados em duas áreas específicas do estado (Figura 29), sendo uma delas o litoral do estado (Figura 29c), considerando que de acordo com o Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES, 2017), em 2006 os municípios do litoral possuíam apenas 12,2% de seu território voltado a agropecuária. Desse modo, tem-se que a agricultura nessa região é pouco desenvolvida. A outra área localizada da classe 3 encontra-se ao sul da Ilha Grande (Figura 29b), entre o rio Paraná, onde localiza-se o Parque Nacional de Ilha Grande, uma Unidade de Conservação de Proteção Integral. Portanto, trata-se de uma região que possui apenas alvos permanentes. Por conta disso, esta classe foi reclassificada como alvo permanente (Figura 30d).

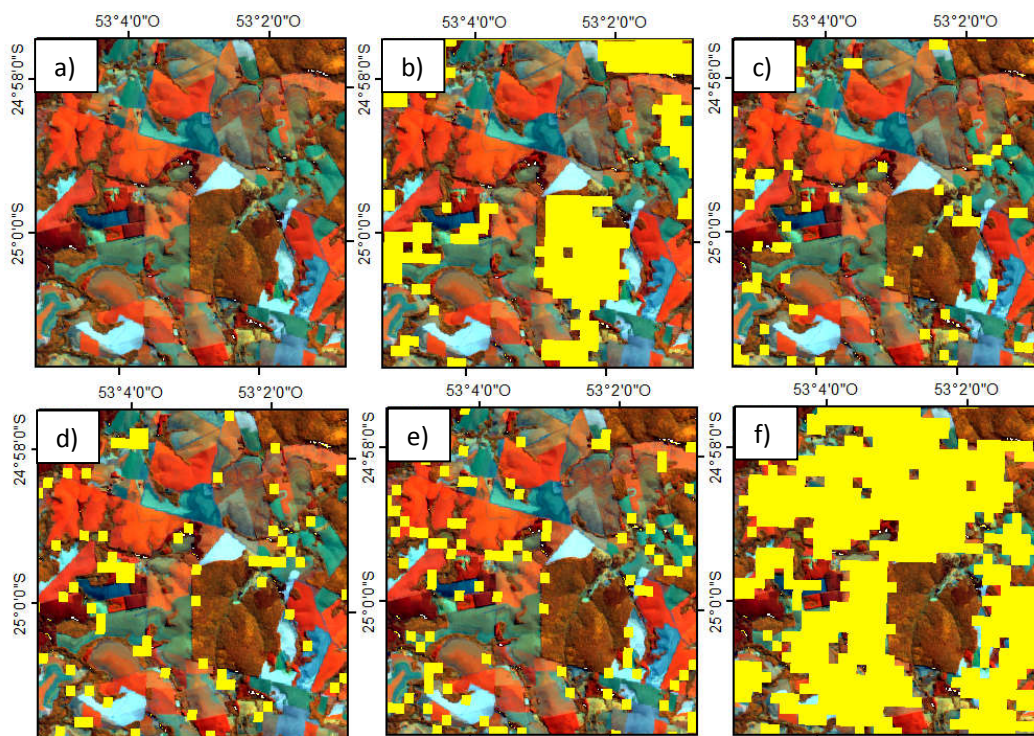


Figura 30 Exemplo da classificação de cada uma das classes conforme composição colorida RGB-564 para cenas Landsta-8/OLI. a) Imagem RGB-564 original; b) Imagem RGB-564 com destaque para os pixels classificados na classe 1; c) Imagem RGB-564 com destaque para os pixels classificados na classe 2; d) Imagem RGB-564 com destaque para os pixels classificados na classe 3; e) Imagem RGB-564 com destaque para os pixels classificados na classe 4; f) Imagem RGB-564 com destaque para os pixels classificados na classe 5.

Por fim, a classe 4 é formada por pixels mistos que possuem pequena porcentagem de alvos permanentes. De acordo com Zhong et al. (2016), em áreas irregulares uma pequena porcentagem de vegetação pode alterar o perfil espectro-temporal de uma cultura. Por conta disso, esta classe foi reclassificada como área agrícola (Figura 30e).

Para a avaliação deste mapeamento utilizou-se um conjunto de 400 amostras extraídas de maneira aleatória, para verificar o EO de cada classe. Desse modo, tem-se que a classe 1 apresentou o menor EO e as classes 3 e 4 apresentaram EO maior que 40%; portanto, essas duas foram reclassificadas utilizando o intervalo interquartil dos pixels da classe 1 (Figura 23).

Após a reclassificação obteve-se o mapeamento final das áreas agrícolas e dos alvos permanentes (Figura 31).

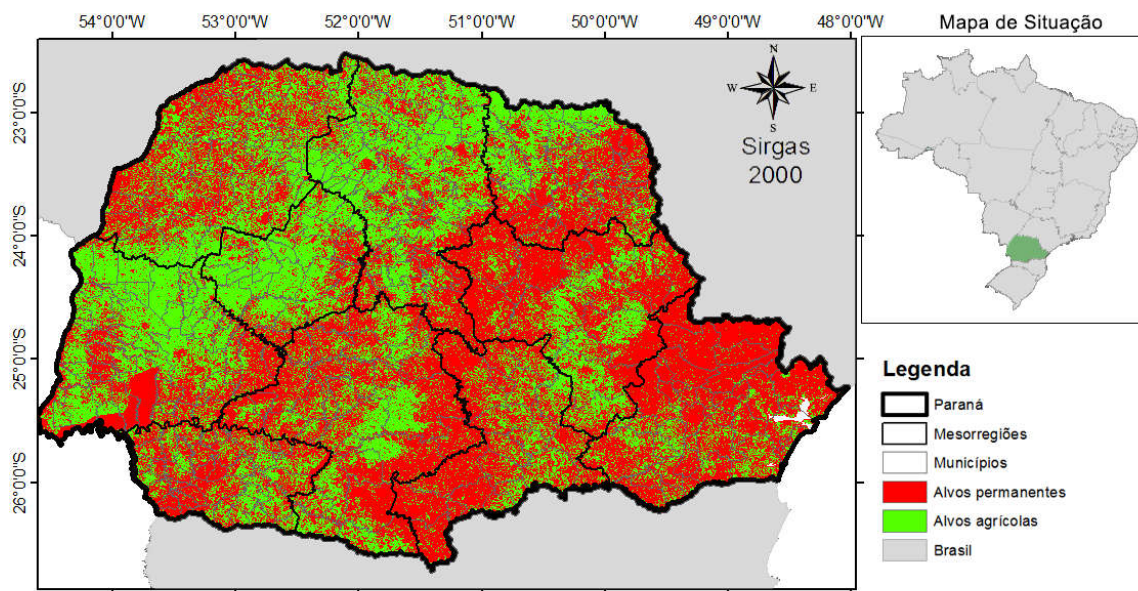


Figura 31 Mapeamento final de áreas agrícolas e alvos permanentes no Paraná para o período de 2013 a 2017.

A avaliação da acurácia espacial deste mapeamento por meio da matriz de confusão (Tabela 11) resultou em uma EG de 77,75% e um IK de 0,555, sendo considerado “bom”, conforme Landis e Koch (1977). Além disso, este valor é próximo aos encontrados por Silva Junior et al. (2017), que obtiveram um índice de 0,59 para mapeamento de soja no estado do Paraná utilizando imagens MODIS.

Tabela 10 Matriz de confusão das áreas agrícolas e dos alvos permanentes

Classes	Referência		Erro de inclusão (EI)	
	Alvo agrícola	Alvo permanente		
Preditos	Alvo agrícola	176	24	12%
	Alvo permanente	65	135	32,5%
Erro de comissão (EC)				
		26,97%	15,09%	EG = 77,75% IK = 0,555

Os EI variaram de 12 a 32,5% e os EC foram de 26,97 e 15,09%, o que representa a proporção de amostras de referência que foram excluídas da classe a que pertenciam.

Os valores se assemelham aos obtidos por Silva Junior, Frank e Rodrigues (2013), que ao analisar seis metodologias de mapeamentos, tiveram EI entre 5 e 38% para a classe soja e o EC entre 13 e 82%. Já Zhang et al. (2017), em três anos de mapeamentos na Coreia do Norte, encontraram EC maiores que 21,1% para o arroz, e superiores a 31,2% para a soja. Já os EI foram superiores a 14,2% para arroz e 22,3% para soja.

6.1.4 Conclusões

A metodologia automática proposta para realização dos mapeamentos de culturas agrícolas e alvos permanentes, para todos os anos-safra (2013/14 a 2016/17) produziu mapeamentos com acurácia espacial excelente, tomando-se como referência as imagens falsa-cor RGB564 do satélite Landsat-8.

A estimativa de área agrícola em escala municipal apresentou correlação de 0,88 para os anos-safra 2013/14 e 2014/15, e 0,87 para o ano-safra 2015/16, sendo, portanto, considerada forte correlação com os dados oficiais. De acordo com os métodos de validação, a metodologia proposta obteve êxito na separação de alvos agrícolas e de alvos permanentes.

A metodologia de mapeamento proposta obteve êxito na automatização do processo de mapeamento, tendo em vista ser possível a automatização da etapa de extração de métricas do perfil espectro-temporal de EVI/MODIS, com rotinas escritas em linguagem R. Isto irá facilitar a realização de futuros mapeamentos com baixo tempo de processamento.

Por fim, a metodologia de classificação utilizando as métricas elaboradas neste trabalho, por meio de um software livre, mostra potencial no que se refere à qualidade dos mapeamentos obtido.

6.1.5 Referências

APARECIDO, L. E. de O.; ROLIM, G. D. S.; RICHETTI, J.; SOUZA, P. S. de; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, v. 40, n. 4, p. 405-417, 2016.

BECKER, W. R.; JOHANN, J. A.; RICHETTI, J.; SILVA, L. C. de A. Data mining techniques for separation of summer crop based on satellite images. **Engenharia Agrícola**, v. 37, n. 4, p. 750-759, 2017. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69162017000400750&lng=en&tlng=en>. Acesso em: 15 dez. 2017.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.

CALLEGARI-JACQUES, S. M. Testes Não-Paramétricos. In: CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2003.

CECHIM JUNIOR, C.; JOHANN, J. A.; ANTUNES, J. F. G. Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, n. 6, p. 427-432, 2017.

COUTO JUNIOR, A.F.; CARVALHO JUNIOR, O.A.; MARTINS, E.S. Séries temporais MODIS aplicadas em sucessão de culturas de soja (*Glycine max* (L.) Merrill) e milho (*Zea mays* L.) em sistema de plantio direto. **Revista Brasileira de Cartografia**, v. 64, p. 405-418, 2012.

EARTHEXPLORER (USGS). **USGS Home**. Disponível em: <<http://earthexplorer.usgs.gov>>. Acesso em: 14 maio 2017.

EMBRAPA INFORMÁTICA E AGROPECUÁRIA. **Banco de produtos Modis**. Disponível em: <<http://www.modis.cnptia.embrapa.br/geonetwork/srv/pt/main.home>>. Acesso em: 12 set. 2017.

FAO – Food and Agriculture Organization. **Statistics Division**. Disponível em: <<http://faostat3.fao.org/browse/Q/QC/E>>. Acesso em: 17 mar. 2018.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHRUSAMY, R. **Advances in Knowledge Discovery & Data Mining**. California: AAAI/MIT, 1996.

FOODY, G. M. Status of land cover classification accuracy assesment. **Remote Sensing of Environment**, v.80, p.185-201, 2002.

FORMAGGIO, A.R.; VIEIRA, M.A.; RENNÓ, C.D.; AGUIAR, D.A.; MELLO, M.P. Object-Based Image Analysis and Data Mining for mapping sugarcane with Landsat imagery in Brazil. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 38, p. 553-562, 2010.

FRANÇA JUNIOR, P.; VILLA, M. E. C. D.; VILLA, J. F. D. Características geomorfológicas e a produção de culturas temporárias no estado do Paraná-Brasil. **Revista Formação**, v. 1, n. 22, p. 71-86, 2015.

GRZEGOZEWSKI, D. M; JOHANN, J. A.; URIBE-OPAZO, M. A.; MERCANTE, E.; COUTINHO, A. C. Mapping soya bean and corn crops in the State of Paraná, Brazil, using EVI images from the MODIS sensor. **International Journal of Remote Sensing**, v. 37, n. 6, p. 1257-1275, 2016.

HUETE, A.; LIU, H. Q.; BATCHILY, K.; LEWEEN, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, v. 59, p. 440-451, 1997.

IPARDES – Instituto Paranaense de Desenvolvimento Social e Econômico. **Webpage**. Disponível em: <http://www.ipardes.gov.br/>. Acesso em: 31 out. 2017.

JOHANN, J. A.; BECKER, W. R.; OPAZO, M. A. U.; MERCANTE, E. Uso de imagens do sensor orbital Modis na estimação de datas do ciclo de desenvolvimento da cultura da soja. **Engenharia Agrícola**, Jaboticabal, v. 35, p. 1-15, 2016.

JOHANN, J. A.; ROCHA, J. V.; DUFT, D. G.; LAMPARELLI, R. A. C. Estimativa de áreas com culturas de verão no Paraná, por meio de imagens multitemporais EVI/Modis. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, 2012.

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

LOBELL, D.B.; THAU, D.; SEIFERT, C.; ENGLE, E.; LITTLE, B. A scalable satellite-based crop yield mapper. **Remote Sensing of Environment**, v.164, p.324-333, 2015.

LORENZETT, C. D. C.; TELÖCKEN, A. V. Estudo Comparativo entre os algoritmos de mineração de dados Random Forest e J48 na tomada de decisão. In: **II Simpósio de Pesquisa e Desenvolvimento em Computação. Anais...** Cruz Alta, RS, 2016.

MACEDO, D. C.; MATOS, S. N. Extração de conhecimento através da mineração de dados. **Revista de Engenharia e Tecnologia**, v. 2, n. 2, p. 22-30, 2010.

MEGETO, G. A. S.; OLIVEIRA, S. R. de M.; DEL PONTE, E. M.; MEIRA, C. A. A. Árvore de decisão para classificação de ocorrências de ferrugem asiática em lavouras comerciais com base em variáveis meteorológicas. **Engenharia Agrícola**, v. 34, n. 3, p. 590-599, 2014.

MELO, M. R. S. **Detecção e mapeamento das pastagens por meio de séries temporais do sensor Modis**. 2017. 106 f. Tese (Doutorado em Engenharia Agrícola). Universidade Estadual de Campinas. Campinas, 2017.

MENGUE, V. P.; FONTANA, D. C. Assessment of spectral-temporal dynamics for mapping the main summer crops in the Rio Grande do Sul State. **Bragantia**, v. 74, p. 122, 2015.

PALOSCHI, R. A. **Software aplicado a modelos de estimativa de produtividade agrícola**. 2016. 98 f. Dissertação (Mestrado em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná, Cascavel, 2016.

PALUDO, A.; JOHANN, J.A.; OPAZO, M.A.U.; RICHETTI, J.; BECKER, W. R. Uso de sensoriamento remoto no mapeamento da cobertura florestal, corpos hídricos e cidades para o estado do Paraná. In: XLIII Congresso Brasileiro de Engenharia Agrícola – CONBEA. **Anais...** Campo Grande, MS, 2014.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acesso em: 15 fev. 2018.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with ERTS. In: Earth Resources Technology Satellite-1 Symposium, 3, Washington, v. 1, p. 309-317, 1973. **Proceedings...** Washington: NASA, 1973.

SANTOS, J.S. dos; FONTANA, D.C.; SILVA, T.S.F.; RUDORFF, B.F.T. Identification of the spatial and temporal dynamics for estimating soybean crop area from MODIS images in the Rio Grande do Sul, Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.18, p.54-63, 2014.

SCHULTZ, B.; IMMITZER, M.; FORMAGGIO, A. R.; SANCHES, I. D. A.; BARRETO LUIZ, A. J.; ATZBERGER, C. Self-guided segmentation and classification of multi-temporal Landsat-8 images for crop type mapping in Southeastern Brazil. **Remote Sensing**, v. 7, p. 14482-14508, 2015.

SILVA JUNIOR, C. A.; NANI, M. R.; TEODORO, P. E.; SILVA, G. F. C. Vegetation Indices for Discrimination of Soybean Areas: A New Approach. **Agronomy Journal**, v. 109, n. 4, p. 1-13, 2017.

SILVA JUNIOR, C. A.; NANNI, M. R.; TEODORO, P. E.; SILVA, G. F. C.; LIMA, M. G.; ERI, M. Comparison of mapping soybean areas in Brazil through perceptron neural networks and vegetation indices. **African Journal of Agricultural Research**, v. 11, p. 114413-114424, 2016.

SILVA JUNIOR, C. A.; FRANK, T.; RODRIGUES, T. C. S. Discriminação de áreas de soja por meio de imagens EVI/MODIS e análise baseada em geo-objeto. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.18, n.1, p.44-53, 2013.

SON, N. T.; CHEN, C. F.; CHEN, C. R.; DUC, H. N.; CHANG, L. Y. A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. **Remote Sensing**, v. 6, n. 1, p. 135-156, 2013.

WILLMOTT, C. J.; ROBERSON, S. M.; MATSUURA, K. A Refined Index of Model Performance. **International Journal of Climatology**, v. 32, p. 2088-2094, 2012.

ZHANG, H.; LI, Q.; LIU, J.; JSHANG, J.; DU, X.; ZHAO, L.; WANG, N.; DONG, T. Crop Classification and Acreage Estimation in North Korea Using Phenology Features. **Giscience & Remote Sensing**, v. 54, n. 3, p. 1-26, 2017.

ZHONG, L.; HU, L.; YU, L.; GONG, P.; BIGING, G. S. Automated mapping of soybean and corn using phenology. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 119, p. 151-164, 2016.

6.2 Uso da dinâmica espectro-temporal de EVI/MODIS para mapeamento semiautomático de florestas e culturas semi-perenes e anuais no Paraná

Resumo: O conhecimento da localização e da quantidade de áreas destinadas a agricultura, florestas nativas ou plantadas é relevante para que os gestores públicos tomem suas decisões pautadas em dados fidedignos com a realidade. Além disso, parte das receitas de ICMS advindas do Fundo de Participação aos Municípios (FPM) depende de dados de produção agropecuária, número de propriedades rurais e fator ambiental. Diante disso, este trabalho teve como objetivo elaborar uma metodologia semiautomática para mapear as áreas de floresta e área agrícola (soja/feijão, milho 1^a e 2^a safras, culturas de inverno e culturas semi-perene) no estado do Paraná para os anos-safra 2013/14 a 2016/17, utilizando séries temporais de índices de vegetação EVI/Modis. A metodologia seguiu as etapas do Processo de descoberta de conhecimento em base de dados – KDD, sendo elaboradas métricas extraídas do perfil espectro-temporal de cada pixel e posteriormente realizada a classificação de todos os pixels do estado com base nessas métricas, utilizando o algoritmo Random Forest. Para a validação dos mapeamentos utilizaram-se amostras extraídas de imagens Landsat-8, obtendo-se índices de exatidão global maior que 84,37% e um índice kappa variando entre 0,63 a 1 – portanto, considerados mapeamentos com boa ou excelente acurácia espacial. Os dados municipais das áreas de soja/feijão, milho 1^a safra, milho 2^a safra e culturas de inverno mapeadas foram confrontados com as estatísticas oficiais, obtendo-se coeficientes de correlação linear entre 0,61 a 0,90, indicando moderada ou forte correlação com os dados oficiais.

Palavras-chave: Série temporal, EVI, Mineração de dados, mapeamento agrícola.

6.2.1 Introdução

O Brasil é o principal exportador de commodities agrícolas do mundo (SCHULTZ et al., 2015), sendo o segundo país que mais produz soja e o terceiro na produção de milho (CONAB, 2017).

No cenário nacional, o estado do Paraná se destaca no setor agropecuário por ser o maior produtor de trigo, com 50,2% da produção total do país, e o segundo estado com a maior produção de soja (17,17%) e de milho (18,25%) (CONAB 2017). Na pecuária, o estado

foi responsável por 14% da produção de leite do país em 2016, e possuía 4,3% dos rebanhos de bovinos destinado ao abate (SEAB/DERAL, 2018).

Diante disto, é fundamental conhecer o uso e a ocupação do território paranaense, incluindo as áreas destinadas a agricultura (culturas semi-perenes e anuais) e áreas permanentes (floresta/reflorestamento, corpos hídricos e alvos urbanos). Estes dados, fidedignos com a realidade, passam a ter maior importância aos municípios quando se avalia o cenário de repasse de ICMS, que é realizado por meio do Fundo de Participação aos Municípios (FPM). Parte da receita anual, obtida pelo FPM, de cada município deriva de dados de produção agropecuária (8%), número de propriedades rurais (6%) e fator ambiental (5%) – conhecido como ICMS Ecológico (PARANÁ, 2017). Além disto, o mercado de commodities agrícolas e os governos se utilizam de dados da produção agrícola para tomadas de decisão referentes a época de semeadura e colheita (JOHANN et al., 2016), armazenamento de grãos, escoamento das safras e oferta dos produtos agrícolas.

No Paraná, a divulgação dos dados de estatísticas oficiais de produção agrícola é realizada pela Secretaria da Agricultura e do Abastecimento (SEAB) e pela Companhia Nacional de Abastecimento (CONAB), que, em geral, utilizam estimativas obtidas de questionários aplicados a técnicos do setor agrícola (BECKER et al., 2017). Entretanto, a elaboração de mapeamentos vem se aprimorando, tanto na qualidade quanto no custo dos mesmos. Isto se justifica pelo desenvolvimento das técnicas de geoprocessamento, uso de imagem de satélite com melhores resolução espacial e temporal, além de softwares que permitem processar essas imagens de maneira rápida e eficaz.

O sensor Modis possui grande aplicabilidade para mapeamento de grandes regiões, tendo em vista que possui resolução temporal de 1 a 2 dias, resolução espacial de 250 metros e disponibilização gratuita das imagens (NASA, 2018).

Entretanto, a utilização de imagens de satélite para realizar mapeamentos apresentam alguns desafios, sendo um deles a dificuldade em distinguir alvos terrestres similares.

Para amenizar esse problema recomenda-se a utilização de índices de vegetação, como o Enhanced Vegetation Index (EVI) (HUETE et al., 1997). A utilização deste índice potencializa a distinção de alvos de vegetação dos demais, pois utilizam em sua composição bandas do espectro eletromagnético que apresentam maiores refletância para alvos de vegetação.

Além disso, a utilização de séries temporais também auxilia na separação de alvos. Contudo, esta análise aumenta o tempo necessário e o custo do processamento dos dados. Uma solução para contornar isto é a automatização dos processos, de modo que este possa ser realizado várias vezes ao longo dos anos (ZHONG et al., 2016).

Entretanto, mesmo com a automatização dos processos, o uso de séries temporais gera uma enorme quantidade de dados, demandando a utilização de técnicas e ferramentas computacionais para obter informações relevantes dos dados analisados. Contudo, neste

artigo utilizou-se o processo de Descoberta de Conhecimento em Base de Dados ou “Knowledge Discovery in Databases – KDD”, sendo este constituído por cinco etapas: seleção, pré-processamento, transformação/formatação, mineração de dados e avaliação/interpretação dos resultados (FAYYAD et al., 1996).

Diante disso, este trabalho teve como objetivo elaborar uma metodologia objetiva e semiautomática, baseada no processo KDD, para mapear as áreas de floresta e área agrícola (soja/feijão, milho 1ª e 2ª safras, culturas de inverno e culturas semi-perene) no estado do Paraná para os anos-safra de 2013/14 a 2016/17, utilizando séries temporais de índices de vegetação EVI/ Modis.

6.2.2 Material e métodos

6.2.2.1 Área de estudo

A área de estudo compreende o estado do Paraná, localizado na região sul do Brasil, entre os paralelos 22° 29' S e 26° 43' S e entre os meridianos 48° 2' W e 54° 38' W, e subdividido em 399 municípios distribuídos em 10 mesorregiões (Figura 32). O estado enquadra-se em três tipos de clima, sendo estes o Cfa, o Cfb e o Cwa (APARECIDO et al., 2016).

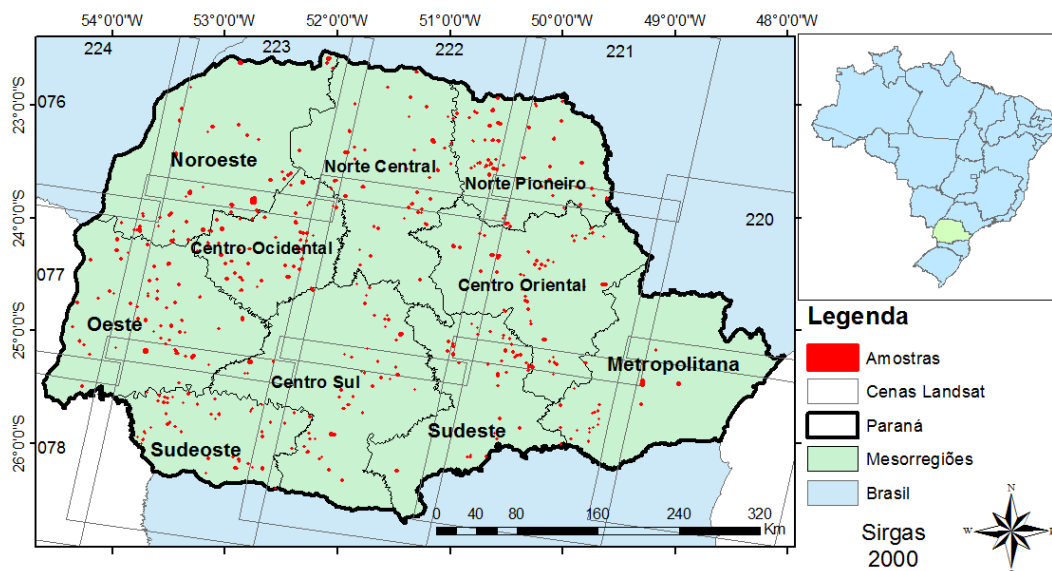


Figura 32 Localização do Estado do Paraná com as 10 mesorregiões, 399 municípios e *tiles* do satélite Landsat-8.

6.2.2.2 Processo de descoberto de conhecimento em base de dados – KDD

Os procedimentos metodológicos seguem as etapas do KDD, resultando em dois mapeamentos por ano-safra (2013/14, 2014/15, 2015/16 e 2016/17), um referente à primeira safra e outro referente à segunda safra (Figura 33) de culturas anuais no estado do Paraná.

A metodologia para construção de cada mapeamento foi dividida em cinco etapas: seleção, pré-processamento, formatação/transformação, mineração de dados e avaliação/interpretação.

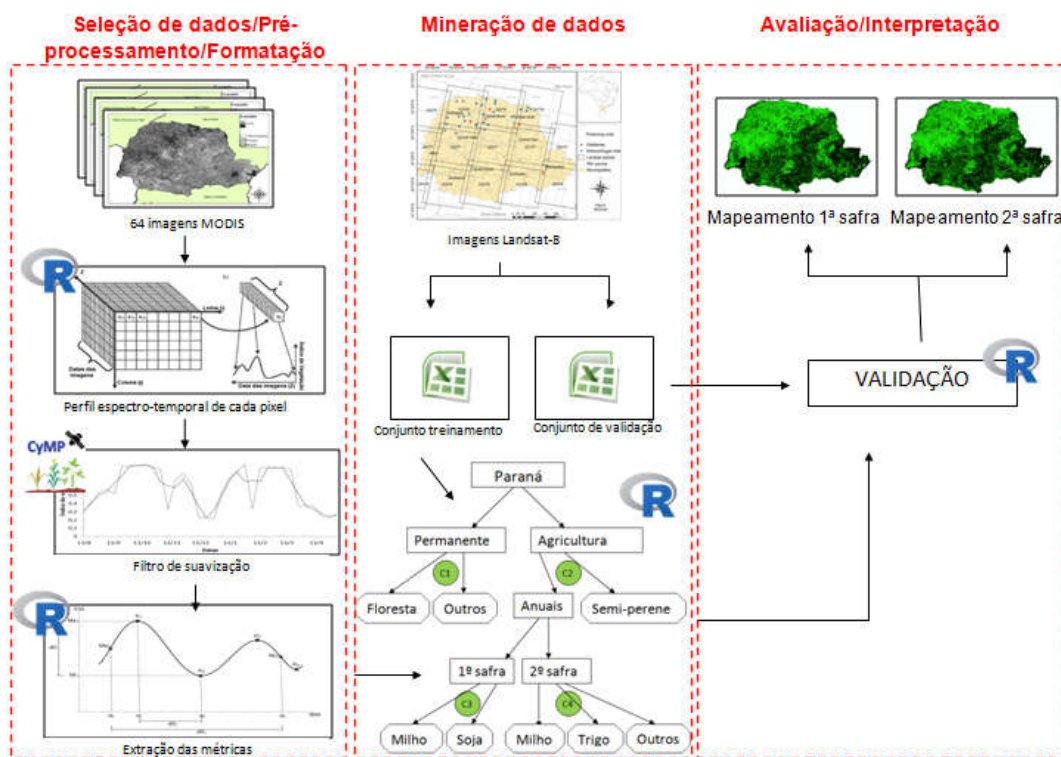


Figura 33 Fluxograma para a separação floresta, demais alvos permanentes, cultura semi-perene, soja/feijão, milho 1ª e 2ª safra, culturas de inverno e outros, com as divisões do processo KDD.

Seleção

Nesta etapa selecionaram-se os dados a serem utilizados durante todo o processo KDD. Primeiramente, elaborou-se a série temporal de imagens do sensor Modis, satélites TERRA (produto MOD13Q1) e AQUA (produto MYD13Q1) do índice de vegetação Enhanced Vegetation Index – EVI (Equação 1), com resolução espacial de 250 m e resolução temporal de 8 dias.

$$EVI = 2,5 \frac{NIR - RED}{1 + NIR + 6RED - 7,5BLUE} \quad \text{Eq. (1)}$$

em que:

NIR - Reflectância na região espectral do infravermelho próximo;

RED - Reflectância na região espectral do vermelho;

BLUE - Reflectância na região espectral do azul.

A série temporal abrangeu imagens de 08/2013 a 08/2017, totalizando 187 imagens. Para cada ano-safra utilizaram-se 64 imagens entre o período de Ago/Ano1 a Dez/Ano2. O ano-safra 2016/17 contemplou 49 imagens do período de Ago/2016 a Ago/2017 em função da disponibilidade de imagens para a realização do trabalho.

Pré-processamento

Em função das imagens apresentarem ruídos ocasionados por nuvens, falhas do detector, dentre outros (MENGUE; FONTANA, 2015), resultando em informações falsas, aplicou-se o filtro de suavização Savitzky–Golay, com auxílio do software CYMP v.1.07 (PALOSCHI, 2016), obtendo-se séries temporais de imagens de EVI/Modis filtradas.

Na sequência, estas imagens foram empilhadas de forma crescente, ou seja, a primeira imagem foi a data mais antiga até a imagem mais recente, resultando um cubo de imagens. Logo, cada pixel da imagem possui 64 respostas de EVI, possibilitando, assim, a construção de um perfil espectro-temporal para cada pixel (Figura 34).

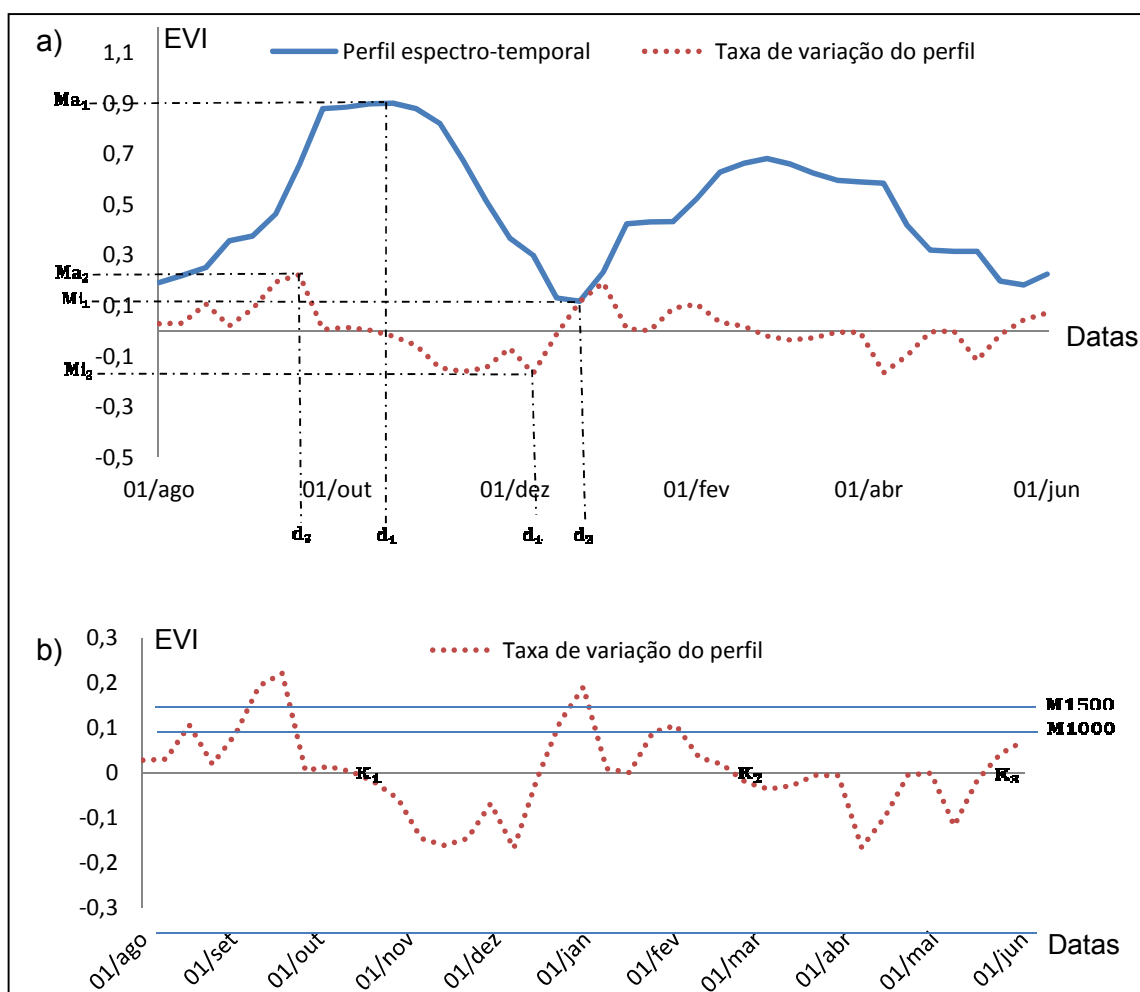


Figura 34 Representação esquemática das métricas obtidas de uma série espectro-temporal de índice e vegetação (a) com ênfase na curva referente a taxa de variação de uma série temporal (b).

Formatação/transformação

Com o intuito de realizar o mapeamento para cada alvo analisado, criaram-se métricas com base no perfil espectro-temporal de EVI filtrado (Figuras 34 e definidas na Tabela 12). Estas métricas foram implementadas por meio de rotinas no software R versão 3.3.1 (R DEVELOPMENT CORE TEAM, 2017), de maneira a automatizar o seu processo de extração.

Desse modo, para cada pixel da imagem EVI/Modis sobre o estado do Paraná, extraíram-se as 18 métricas (Tabela 12), originando uma matriz com as linhas representando cada um dos pixels e 18 colunas representando as métricas elaboradas. Esta matriz de dados foi utilizada como dado de entrada para a próxima etapa do KDD.

Tabela 11 Métricas extraídas de uma série espectro-temporal de índice de vegetação (EVI)

Métricas	Descrição
Ma_1	Valor do máximo EVI registrado em cada pixel
Mi_1	Valor do mínimo EVI registrado em cada pixel
dif_1	Diferença entre o máximo e o mínimo EVI
Ma_2	Valor da maior taxa de variação do EVI registrado em cada pixel
Mi_2	Valor da menor taxa de variação do EVI registrado em cada pixel
K_i	Pontos de mudança da taxa de variação
d_1	Data que ocorre o máximo EVI para cada pixel
d_2	Data que ocorre o mínimo EVI para cada pixel
d_3	Data que ocorre a maior taxa de variação do EVI para cada pixel
d_4	Data que ocorre a menor taxa de variação do EVI para cada pixel
dif_2	Diferença entre as datas que ocorre o máximo e o mínimo EVI
dif_3	Diferença entre as datas que ocorre o Ma_2 e o Mi_2
M1500	Número de datas com taxa de variação maior que 1500
M1000	Número de datas com taxa de variação maior que 1000
Média	Média do EVI de todas as datas para cada pixel
Int	Soma dos valores de EVI de todas as datas para cada pixel
Dp	Desvio padrão de cada pixel
CV	Coefficiente de variação de cada pixel ao longo das datas

Mineração de dados

Nesta etapa foi efetuada a tarefa de classificação por meio do algoritmo Random Forest. Para a aplicação desse algoritmo elaborou-se um conjunto de amostras (Figura 32) formado por pixels puros (GRZEGOZEWSKI et al., 2016) de cada alvo de interesse, selecionados através de uma grade de pixels EVI/Modis (em que cada pixel possui 250 x 250m) sobreposta às imagens do Landsat-8/OLI (Figura 35), com o intuito de extrair apenas pixels do alvo de interesse, sem mistura espectral. A seleção dos pixels foi executada no software Arcgis 10.3. Para confirmar o alvo, analisou-se o perfil espectro-temporal deste por meio da ferramenta sistema de análise temporal da vegetação (SATVEG), disponibilizada pela Embrapa (SATVEG, 2017).

Com os pixels puros extraídos construiu-se um conjunto de amostras, sendo que 75% foi utilizado para elaborar o processo de treinamento do algoritmo (“traindata”) e os outros 25% foram utilizados para validar o processo (validation data). O “traindata” são os dados utilizados pelo algoritmo para gerar regras de classificação, e com base nessas regras, os restantes dos dados são classificados, sendo que, primeiramente todos os pixels da área de estudo foram separados em pixels de alvos permanentes e pixels de área agrícola, conforme resultados obtidos no primeiro artigo desta dissertação.

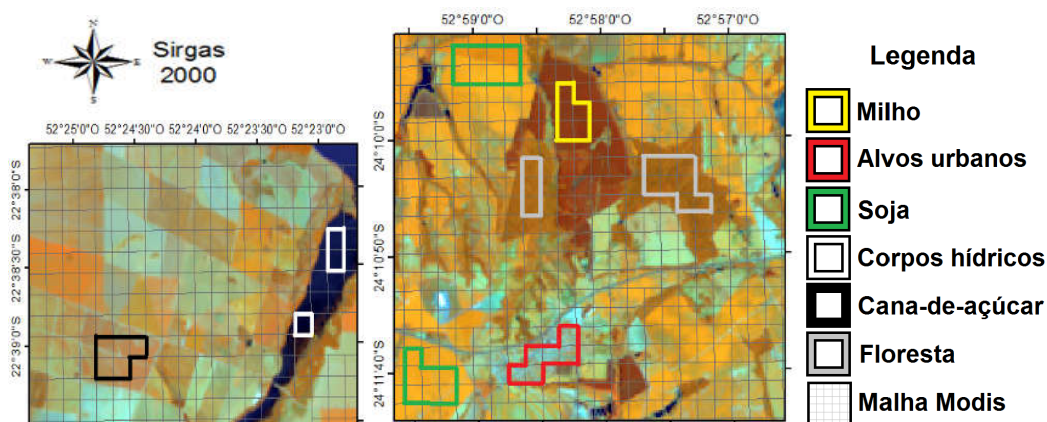


Figura 35 Método de Seleção de alvos de culturas agrícolas e alvos permanentes na composição falsa cor RGB564 com o detalhamento da malha Modis.

Em seguida, foram realizadas quatro etapas de classificação (Figura 36). Primeiramente, dentro do mapa de alvos permanentes, fez-se a separação entre a área de floresta/reflorestamento (flo) dos demais alvos permanentes (dap). Em seguida, o mapa de agricultura foi separado em culturas anuais ou semi-perenes (asp). Os pixels classificados como culturas anuais foram separados em culturas de 1ª e 2ª safra, sendo que esse processo de separação se baseou apenas nas datas das imagens. Por fim, o mapa de cultura de 1ª safra foi reclassificado como soja/feijão (s/f) e milho 1ª safra (mil1), e as culturas de 2ª safra foram segmentadas em milho 2ª safra (mil2), culturas de inverno (ci) e outros alvos (out) (Figura 36). Todos estes mapeamentos foram elaborados no software R, utilizando o algoritmo Random Forest.

a) Mapeamento de floresta e reflorestamento

Para os alvos permanentes realizou-se uma nova classificação entre os pixels de floresta/reflorestamento e demais alvos permanentes. Os dados de entrada do algoritmo Random Forest foram as métricas: \mathbf{Ma}_1 , \mathbf{Mi}_1 , $\mathbf{Média}$, \mathbf{CV} , \mathbf{K}_i e $\mathbf{M1500}$ (Tabela 12).

A escolha desta métricas se justifica pela facilidade que estas evidenciam os alvos de floresta/reflorestamento, pois as métricas \mathbf{Ma}_1 , $\mathbf{Média}$ e \mathbf{Mi}_1 distinguem o alvo floresta dos demais alvos que possuem menor refletância ao longo de uma safra (como, por exemplo, corpos hídricos e alvos urbanos). Já as métricas \mathbf{CV} , \mathbf{K}_i e $\mathbf{M1500}$ auxiliam na distinção de floresta e reflorestamento dos alvos agrícolas classificados erroneamente.

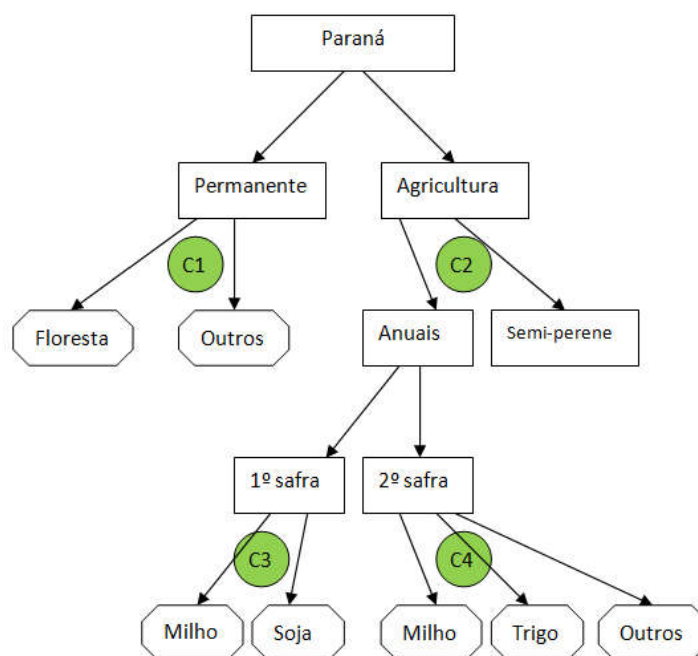


Figura 36 Procedimento de classificação com algoritmo Random Forest dos pixels EVI/Modis no Paraná em alvos permanentes (C1), alvos com culturas agrícolas anuais ou semi-perenes (C2) e culturas anuais de primeira safra (C3) e segunda safra (C4).

b) Mapeamento de culturas semi-perenes

Para os pixels inicialmente classificados como agricultura realizou-se a separação entre agricultura anual e semi-perene, utilizando as métricas: \mathbf{Ma}_1 , \mathbf{Mi}_1 , \mathbf{Ma}_2 e $\mathbf{M1000}$, como dados de entrada (Tabela 12). Estas métricas apresentam características diferentes para os alvos analisados, em que \mathbf{Ma}_1 , \mathbf{Ma}_2 e $\mathbf{M1000}$ possuem os maiores valores em pixels de agricultura anual, e a \mathbf{Mi}_1 apresentam valores maiores para agricultura semi-perene.

c) Mapeamento de culturas anuais de 1ª safra

Utilizando os pixels classificados como agricultura anual e as 30 primeiras imagens da série temporal (de Ago/Ano1 a Fev/Ano2), elaborou-se uma nova série temporal referente à primeira safra. Posteriormente, as 18 métricas foram recalculadas para a nova série temporal e, em seguida, foram utilizadas para reclassificar os pixels agrícola entre soja/feijão e milho 1ª safra.

Neste processo foram utilizadas as métricas: d_1 , d_2 , d_3 , d_4 , dif_2 e dif_3 (Tabela 12), as quais diferenciam as culturas analisadas em função das datas de máximo, mínimo, taxas de variação do EVI, como também dos ciclos das culturas.

d) Mapeamento de culturas anuais de 2^o safra

Para a segunda safra foram utilizados os pixels de agricultura anual, porém empregando uma nova série temporal com as imagens de janeiro a agosto do mesmo ano. Com isto, foram recalculadas as 18 métricas, reclassificando os pixels entre milho 2^a safra, culturas de inverno (aveia, centeio, cevada, trigo e triticale) e outros. Para a classificação destas classes foram utilizadas as 18 métricas elaboradas.

Avaliação e interpretação

Em cada processo de classificação foram utilizados a Exatidão Global (EG) (Equação 2), o índice Kappa (Equação 3), o Erro de omissão (EO) e o Erro de comissão (EC), utilizando como referência os dados do conjunto de validação.

$$EG = \frac{A}{m} * 100 \quad \text{Eq. (2)}$$

$$IK = \frac{\left[m * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_i * x_j) \right]}{\left[m^2 - \sum_{i=1}^r (x_i * x_j) \right]} \quad \text{Eq.(3)}$$

em que:

A - acerto geral (pontos amostrais com acertos);

m - número de pontos amostrais;

n - número de observações;

x_{ij} - observação na linha i e coluna j;

x_i - total marginal da linha i;

x_j - total marginal da coluna j.

Também se realizou a comparação da área municipal mapeada com os dados oficiais de áreas de soja/feijão 1^a safra, milho 1^a e 2^a safra e culturas de inverno (aveia, centeio, cevada, trigo e triticale) obtidos do IBGE, para os anos-safra 2013/14, 2014/15 e 2015/16. Tal procedimento não foi executado para o ano-safra 2016/17 pela não divulgação dos dados oficiais. A análise foi realizada pelos indicadores estatísticos: erro médio (ME) (Equação 4), erro médio quadrático (RMSE) (Equação 5), índice de concordância aprimorado de Willmott (dr) (WILLMOTT; ROBERSON; MATSUURA, 2012) (Equação 6) e coeficiente de correlação de Pearson (r).

$$ME = \frac{1}{n} * \sum_{i=1}^n (O - E) \quad \text{Eq. (4)}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^n (O - E)^2} \quad \text{Eq. (5)}$$

$$dr = 1 - \frac{\sum_{i=1}^n |E - O|}{2 \sum_{i=1}^n (|O - \bar{O}|)} \quad \text{Eq. (6)}$$

em que:

n - número de observações;

O - área da cultura usada como referência;

E - área da cultura obtida nos mapeamentos;

\bar{O} - área da cultura média usada como referência.

6.2.3 Resultados

Em todos os anos-safra, os mapeamentos apresentam semelhanças na espacialização dos alvos (Figura 37), em que os alvos de soja/feijão, milho e trigo se concentram do Oeste ao Norte Pioneiro, bem como no Centro Oriental Paranaense. A agricultura semi-perene está presente em maior concentração nas regiões Norte e Noroeste do estado, em função das condições climáticas favoráveis ao plantio de cana-de-açúcar nesta área (CECHIM JUNIOR et al., 2017).

De maneira geral os mapeamentos possuem padrões semelhantes. A maior diferença visual nos mapeamentos ocorreu do ano-safra 2015/16 (Figuras 37e e 37f), que possui maior ocorrência de alvos permanentes, em especial na região noroeste do estado.

6.2.3.1 Validação dos mapeamentos com os dados dos conjuntos de amostras de pixels puros

Nos quatro processos de classificação do ano-safra 2013/14 (alvos permanentes, alvos agrícolas de 1ª e 2ª safras), o índice Kappa variou entre 0,53 e 1 (Tabela 13), com EG superior a 80% (Tabela 13). O mapeamento das culturas da 1ª safra teve os menores IK (0,53). Os Erros de Omissão (EO) foram maiores nos alvos de milho 1ª safra (45,6%) e

culturas de inverno (46,2%), e o Erro de Comissão (EC) foi maior para o milho 1ª safra (19,6%) e para as culturas de inverno (30%).

Para o mapeamento referente ao ano-safra 2014/15 (Figura 37c e 37d), o IK dos quatro processos de classificação variou entre 0,55 a 1, em que o menor IK corresponde às culturas da 1ª safra e o maior corresponde ao mapeamento de florestas/reflorestamento e demais alvos permanentes. Os maiores valores de erros de omissão e comissão se encontram nos alvos de culturas de primeira safra, ou seja, soja/feijão e milho (Figura 38b). Além disso, o EG dos mapeamentos ficou entre 79,47 a 100% (Tabela 13).

Para o ano-safra 2015/16 (Figura 37e e 37f), o IK dos quatro processos de classificação variou entre 0,63 e 1, sendo que o menor valor corresponde ao mapeamento de 1ª safra e o maior ao mapeamento de alvos permanentes (floresta/reflorestamento e demais alvos), bem como o mapeamento de culturas anuais de 2ª safra. O EG dos quatro mapeamentos foi superior a 80% (Tabela 13). Os erros de omissão (EO) e comissão (EC) foram maiores para os alvos de culturas de inverno e milho 1ª safra, conforme Figura 38c.

O ano-safra 2016/17 (Figura 37g), o IK dos mapeamentos de alvo permanente, alvo agrícola e 1ª safra foram superiores a 0,92 com EG superior a 90% (Tabela 13). O EO foi superior para o alvo de milho 1ª safra (17,46%). O EC foi superior para o milho 1ª safra e para a soja/feijão, com 8,77 e 7,01%, respectivamente (Figura 38d).

Em média, os mapeamentos de alvo permanente, alvo agrícola (culturas anuais e semi-perene) e 2ª safra obtiveram valores de EG superiores aos 85%, considerados aceitáveis por Foody (2002). Além disso, estes mapeamentos apresentaram um IK classificado como excelente (superior a 0,8), enquanto o mapeamento de 1ª safra foi considerado bom (superior a 0,6 e inferior a 0,8), de acordo com a avaliação de Landis e Koch (1997).

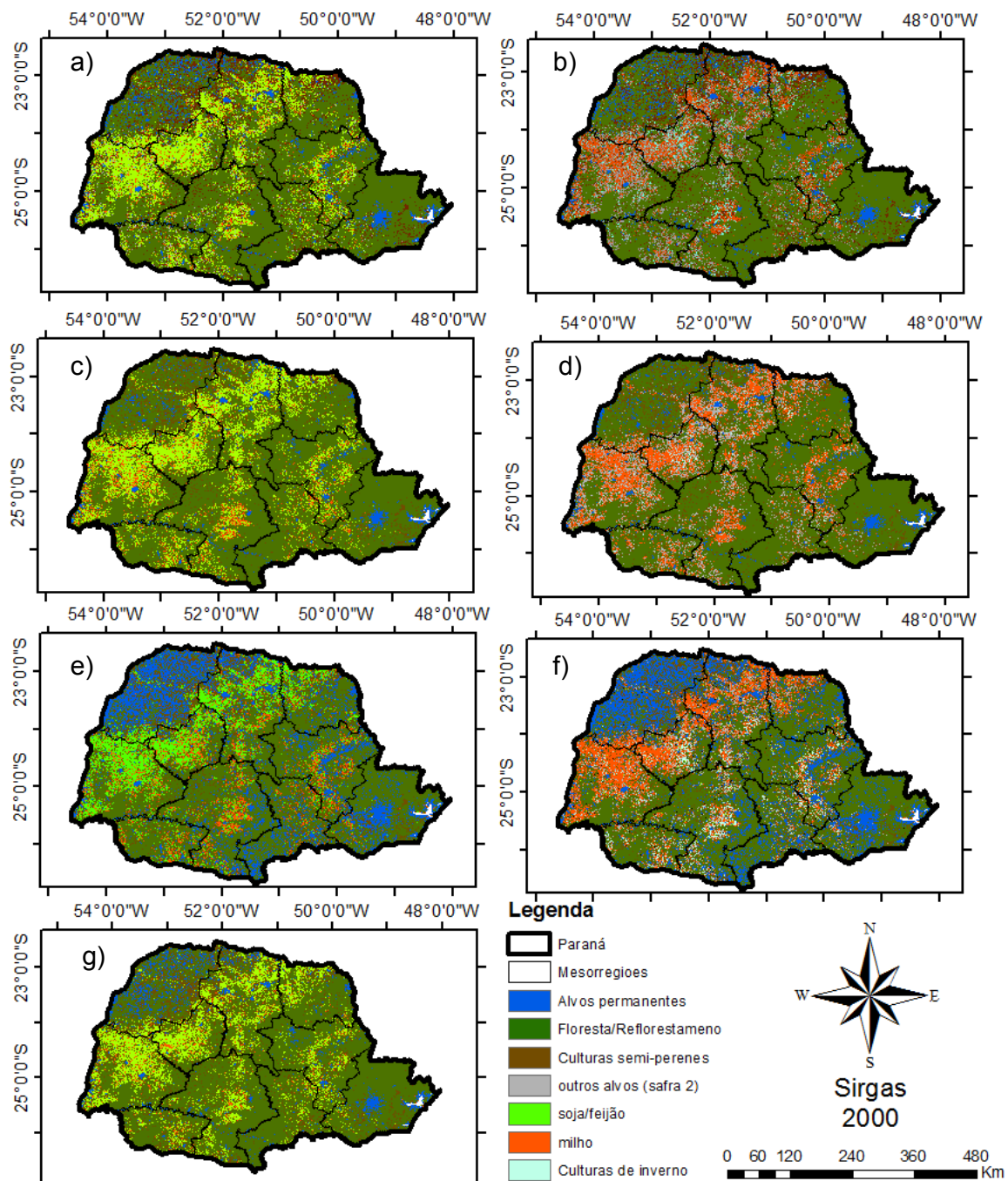


Figura 37 Mapeamento de alvos permanentes, floresta, culturas semi-perenes soja/feijão, milho 1ª e 2ª safra, culturas de inverno e outros, para os anos-safra 2013/14 a 2016/17, sendo que para 2013/14 tem-se o mapeamento 1ª safra (a) e 2ª safra (b), para o ano-safra 2014/15 tem-se 1ª safra (c) 2ª safra (d) no ano-safra 2015/16 tem-se o mapeamento da 1ª safra (e) e 2ª safra (f) e por fim o ano-safra 2016/17 possui apenas mapeamento da 1ª safra (g).

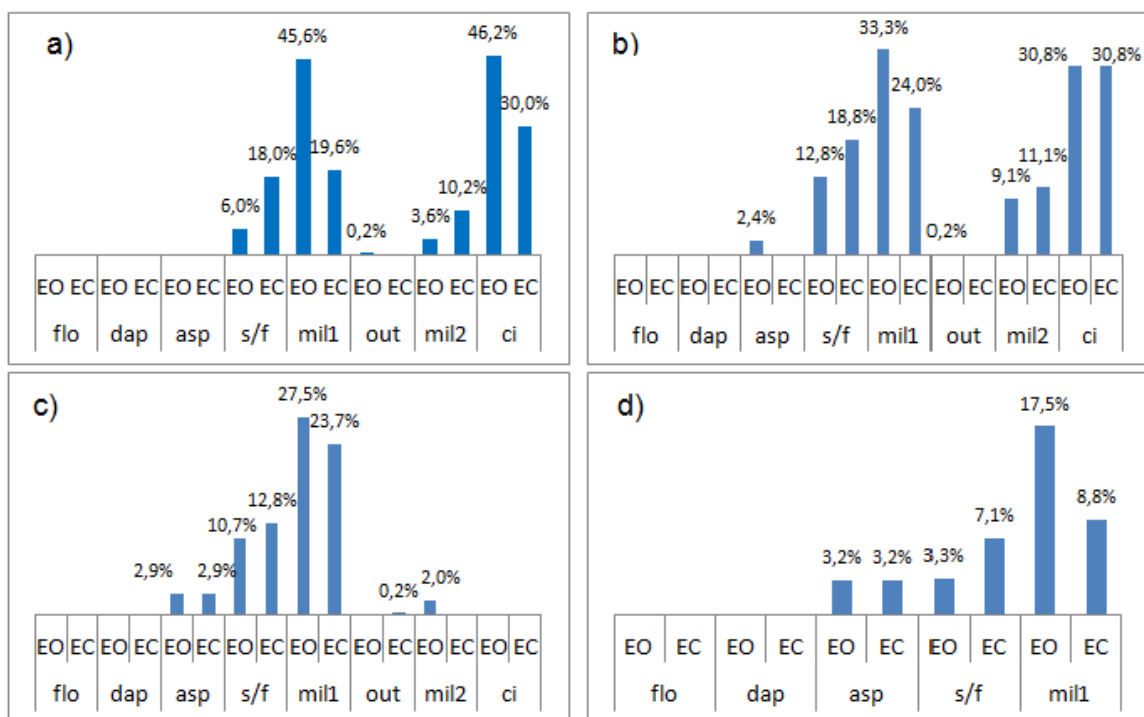


Figura 38 Erro de Omissão (EO) e Erro de comissão (EC) para os alvos floresta/reflorestamento (flo), demais alvos permanente (dap), agricultura semi-perene (asp), soja/feijão (s/f), milho 1ª safra (mil1), outros (out), milho 2ª safra (mil2) e culturas de inverno (ci) para os anos-safra 2013/14 (a); 2014/15 (b); 2015/16 (c) e 2016/17(d).

Tabela 12 Valores de Exatidão global (EG) e índice Kappa (IK) para os quatro anos-safra

	2013/2014		2014/2015		2015/2016		2016/2017		Média	
	EG	IK	EG	IK	EG	IK	EG	IK	EG	IK
Alvo permanente	100	1	100	1	100	1	100	1	100	1
Alvo agrícola	100	1	99,46	0,98	98,79	0,96	98,93	0,96	99,29	0,98
1ª safra	81,65	0,53	79,47	0,55	83,87	0,63	92,49	0,82	84,37	0,63
2ª safra	94,3	0,92	93,5	0,90	99,03	0,99	-	-	95,61	0,94

Os resultados obtidos são semelhantes aos encontrados na literatura, visto que Cechim Junior et al. (2017) obtiveram IK de 0,85 para o mapeamento de culturas de verão (soja e milho) na Bacia do Alto Paraguai (BAP) para o ano-safra 2000/01; já no ano-safra 2015/16 o esse valor foi de 0,86. Zhong et al. (2016) obtiveram um EG de 87,2% e IK de 0,804, empregando uma metodologia automatizada para mapear soja e milho no Paraná, utilizando imagens Modis. Souza et al. (2015) obtiveram EG de 86% e Kappa de 0,78 utilizando o índice Wide Dynamic Range Vegetation Index (WDRVI) para mapear soja e milho

no Paraná para o ano-safra 2010/11. Diante disso, os resultados obtidos no presente artigo foram tão bons quanto os descritos na literatura.

Para o EO, os resultados obtidos variaram de 0,13 a 30,97% e o EC variou de 0,06 a 20,26%. Estes valores se assemelham com os obtidos por Zhang et al. (2017), que encontrou um EO para a classificação de arroz na Coreia do Norte maior que 21,1% para os três anos estudados, e para a classificação de soja os erros ultrapassaram 31,2%; já o EC apresentou valores maiores que 14,2 para arroz e 22,3% para soja. Antunes et al. (2012) obteve valores EO de 3,4 e 67,6%, e para o EC os valores foram de 8,3 a 40% para o mapeamento de soja e não soja em oito municípios paranaenses para o ano-safra 2003/04.

6.2.3.2 Comparação com dados oficiais

Comparando os resultados obtidos com os dados oficiais, disponibilizados pelo IBGE (IBGE, 2018), o mapeamento que apresentou a maior diferença, em média, foi de soja/feijão, que subestimou os dados oficiais em 30,5% (Figura 39). De acordo com Zhong et al. (2016), nos pixels de áreas agrícolas com mistura espectral, quando este possui uma pequena porcentagem de vegetação densa, esta porcentagem pode alterar o perfil espectro-temporal do pixel, dificultando a classificação deste como agricultura.

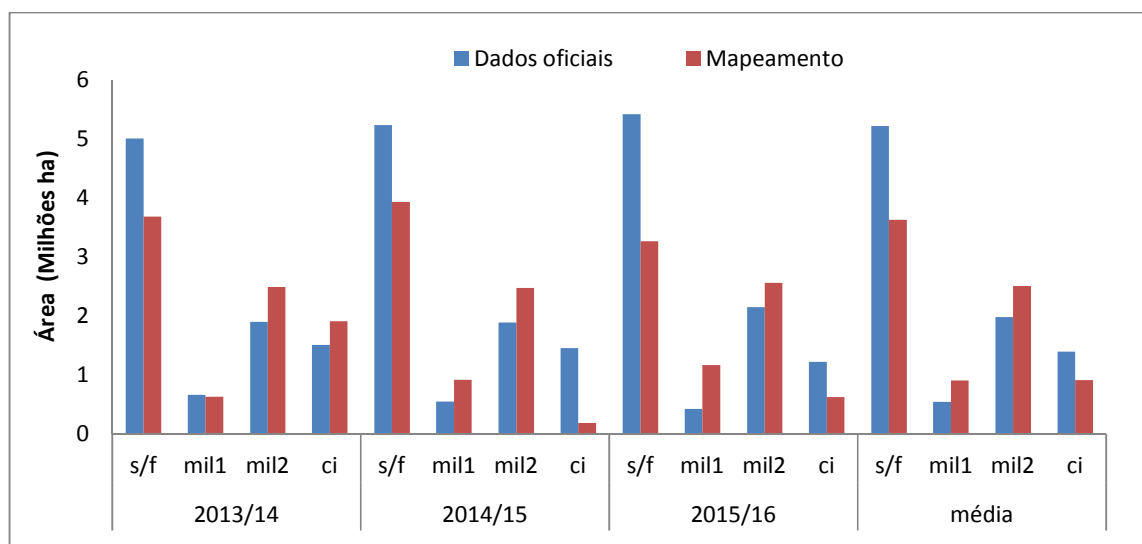


Figura 39 Comparação das áreas (ha) do mapeamento com os dados oficiais para os alvos soja/feijão (s/f), milho 1ª safra (mil1), milho 2ª safra (mil2) e culturas de inverno (ci).

De modo geral, os indicadores estatísticos mostram que na média dos três anos o ME variou de -3.996 ha (mapeamento de soja/feijão) a 1.326 ha (mapeamento de milho 2ª safra). Os maiores erros médios também foram atribuídos aos dois mapeamentos citados. Na média dos três anos-safra, o RMSE variou, em média, de 3.534 ha a 7.688 ha (Tabela 14). Este erro pode estar atrelado à média resolução espacial do sensor empregado, como também ao grande número de pequenas propriedades agrícolas no Paraná, pois, de acordo com Zhong

et al. (2016) e Silva Junior et al. (2017), esta característica dificulta o processo de mapeamento em função da maior mistura espectral de alvos num único pixel.

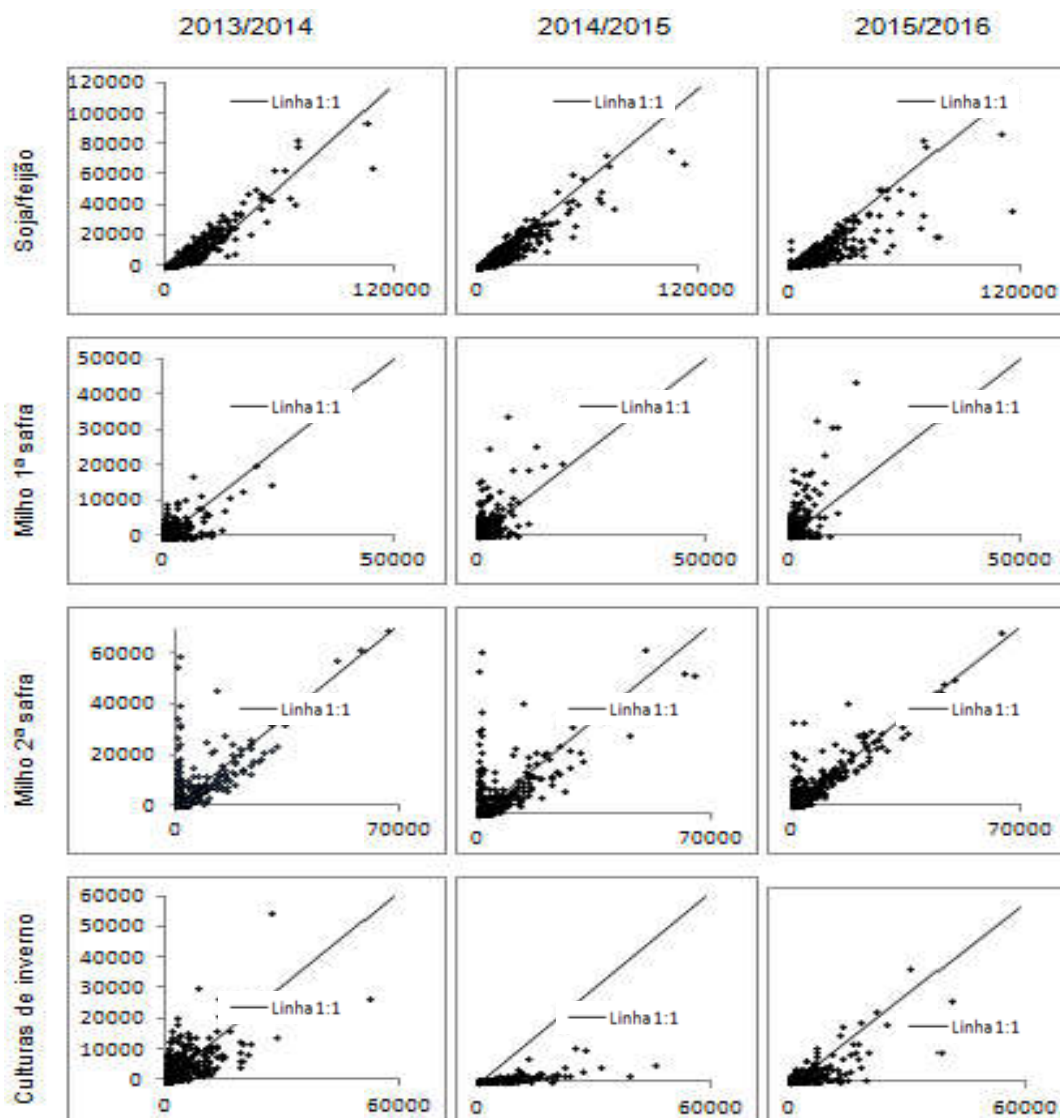


Figura 40 Gráficos de dispersão para os alvos agrícolas (Soja/feijão, milho 1ª e 2ª safra e culturas de inverno) para os anos-safra 2013/14 a 2015/16.

Utilizando o coeficiente linear (r), em média, o mapeamento de milho 1ª safra possui moderada correlação com os dados oficiais, enquanto os demais mapeamentos apresentam forte correlação (Figura 40). Para o r a variação foi de 0,3 a 0,78 (Tabela 14).

A correlação dos mapeamentos executados por Cechim Junior et al. (2017), que propôs uma metodologia para mapear áreas de cultivo de cana-de-açúcar no Paraná utilizando imagens dos satélites Landsat/TM/OLI e IRS/LISS-3 para os anos-safra 2010/2011 a 2013/2014, comparada aos dados oficiais do IBGE variou de moderada a forte ($0,64 \leq r_s \leq 0,80$). Melo (2017) realizou um mapeamento de áreas de pastagem no estado de São Paulo,

resultando num (r) 0,83 e dr de 0,81. Além disso, o ME foi de aproximadamente 1.400 ha e o RMSE de 6.711 ha.

Os resultados do ME e RMSE também foram semelhantes ao trabalho realizado por Mengue e Fontana (2015), cujo objetivo foi mapear as culturas de verão no estado do Rio Grande do Sul utilizando o sensor MODIS/EVI. Neste, o ME foi de 436 ha e -963 ha para arroz e soja, respectivamente, e para o RMSE os valores foram de 2.133 ha para arroz e 6.665 ha para soja. Logo, isto indica que os resultados obtidos no presente trabalho se assemelham com os demais encontrados na literatura.

Tabela 13 Média dos Indicadores estatísticos: Erro médio (ME), Erro médio quadrático (RMSE) Índice de concordância aprimorado de Willmott (dr) e coeficiente de correlação de Pearson (r) para soja/feijão, milho 1ª safra, milho 2ª safra e culturas de inverno

Ano-Safra	Estatísticas	Soja/Feijão	Milho safra1	Milho safra2	Culturas de inverno
2013/14	ME	-3.318	-76	1.483	-1.004
	RMSE	6.185	2.446	7.282	4.916
	dr	0,81	0,56	0,69	0,62
	r	0,93	0,57	0,72	0,68
2014/15	ME	-3.261	-925	1.475	3.169
	RMSE	6.665	3.591	7.321	5.930
	dr	0,80	0,35	0,67	0,59
	r	0,93	0,58	0,7	0,74
2015/16	ME	-5.408	-1.861	1.021	1.496
	RMSE	10.213	4567	4.611	3.609
	dr	0,73	-0,02	0,79	0,71
	r	0,83	0,69	0,90	0,77
Média	ME	-3.996	-953	1.326	1.220
	RMSE	7.687	3.535	6.405	4.819
	dr	0,78	0,30	0,72	0,64
	r	0,90	0,61	0,77	0,73

6.2.4 Conclusões

Com base nas amostras extraídas de imagens com alta resolução espacial, os resultados dos mapeamentos de cada ano-safra são considerados bons e excelentes de acordo com o índice Kappa.

A estimativa de área agrícola em escala municipal apresentou de moderada a forte correlação quando comparada com dados oficiais.

As métricas elaboradas de maneira automática permitiram a obtenção de êxito na classificação, gerando uma metodologia semiautomática para auxiliar em mapeamentos futuros com baixo tempo de processamento.

Com o mapeamento das culturas agrícolas é possível que o município conteste o órgão oficial quanto à efetiva área de produção agrícola, caso haja divergência com os dados oficiais, tendo em vista que esses dados são utilizados para o cálculo do FPM.

Além disso, o mapeamento de florestas e reflorestamento possibilita aos municípios obter dados da localização dessas áreas, permitindo acessar o FPM referente ao fator ambiental ou ICMS ecológico, visto que 38,35% dos municípios paranaense não recebem esse recurso.

6.2.5 Referências

ANTUNES, J. F. G.; MERCANTE, E.; ESQUERDO, J. C. D. M.; LAMPARELLI, R. A.; ROCHA, J. V. Estimativa de área de soja por classificação de imagens normalizada pela matriz de erros. **Pesquisa Agropecuária Brasileira**, v. 47, p. 1288-1294, 2012.

APARECIDO, L. E. de O.; ROLIM, G. D. S.; RICHETTI, J.; SOUZA, P. S. de; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, v. 40, n. 4, p. 405-417, 2016.

BECKER, W. R.; JOHANN, J. A.; RICHETTI, J.; SILVA, L. C. de A. Data mining techniques for separation of summer crop based on satellite images. **Engenharia Agrícola**, v. 37, n. 4, p. 750-759, 2017. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69162017000400750&lng=en&tng=en>. Acesso em: 10 dez. 2017.

CECHIM JUNIOR, C.; JOHANN, J. A.; ANTUNES, J. F. G. Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, n. 6, p. 427-432, 2017.

CECHIM JÚNIOR, C; JOÃO FRANCISCO GONÇALVES ANTUNES, J. F. G; JOHANN, J. A; ESQUERDO, J. C. D. M. Mapping of summer agricultural crops in the Alto Paraguai basin through EVI/Modis time series. **Geografia**, v. 42, n. 3, p. 129-143, 2017

CONAB – Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de grãos. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/17_07_12_11_17_01_boletim_graos_julho_2017.pdf>. Acesso em 08 dez. 2017.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHRUSAMY, R. **Advances in Knowledge Discovery & Data Mining**. California: AAAI/MIT, 1996.

FOODY, G. M. Status of land cover classification accuracy assesment. **Remote Sensing of Environment**, v. 80, p. 185-201, 2002.

GRZEGOZEWSKI, D. M.; JOHANN, J. A.; URIBE-OPAZO, M. A.; MERCANTE, E.; COUTINHO, A. C. Mapping soya bean and corn crops in the State of Paraná, Brazil, using EVI images from the MODIS sensor. **International Journal of Remote Sensing**, v. 37, n. 6, p. 1257-1275, 2016.

HUETE, A.; LIU, H. Q.; BATCHILY, K.; LEWEEN, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, v. 59, p. 440-451, 1997.

IBGE. **Produção Agrícola Municipal – PAM 2016**. Disponível em: <<https://sidra.ibge.gov.br/pesquisa/pam/tabelas>>. Acesso em: 10 jan. 2018.

JOHANN, J. A.; BECKER, W. R.; OPAZO, M. A. U.; MERCANTE, E. Uso de imagens do sensor orbital Modis na estimação de datas do ciclo de desenvolvimento da cultura da soja. **Engenharia Agrícola**, v. 35, p. 1-15, 2016.

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

MELO, M. R. S. **Deteção e mapeamento das pastagens por meio de séries temporais do sensor Modis**. 2017. 106 f. Tese (Doutorado em Engenharia Agrícola). Universidade Estadual de Campinas. Campinas, 2017.

MENGUE, V. P.; FONTANA, D. C. Assessment of spectral-temporal dynamics for mapping the main summer crops in the Rio Grande do Sul State. **Bragantia**, v. 74, p. 122, 2015.

NASA – National Aeronautics and Space Administration. **MODIS: Moderate Resolution Imaging Spectroradiometer**. Disponível em: <<https://modis.gsfc.nasa.gov/about/>>. Acesso em: 10 dez. 2017.

PALOSCHI, R. A. **Software aplicado a modelos de estimativa de produtividade agrícola**. 2016. 98 f. Dissertação (Mestrado em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná. Cascavel, 2016.

PARANÁ. **Transferência de 25% do ICMS aos municípios (cota-parte ICMS)**. Disponível em: <<http://www.liderespublicos.pr.gov.br/2016/11/36/Transferencia-de-25-do-ICMS-aos-municipios-cota-parte-ICMS.html>>. Acesso em: 04 dez. 2017.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acesso em: 15 fev. 2018.

SATVEG **Sistema de Análise Temporal da Vegetação (SATVEG)**. Disponível em: <<https://www.satveg.cnptia.embrapa.br/satveg>>. Acesso em: 25 out. 2017.

SCHULTZ, B.; IMMITZER, M.; FORMAGGIO, A. R.; SANCHES, I. D. A.; BARRETO LUIZ, A. J.; ATZBERGER, C. Self-guided segmentation and classification of multi-temporal Landsat-8 images for crop type mapping in Southeastern Brazil. **Remote Sensing**, v. 7, p. 14482-14508, 2015.

SEAB – Secretaria da Agricultura e do Abastecimento do Paraná. DERAL – Departamento de Economia Rural. **Estimativa de safra**. Disponível em: <<http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=74>>. Acesso em: 20 nov. 2017.

SILVA JUNIOR, C. A.; NANI, M. R.; TEODORO, P. E.; SILVA, G. F. C. Vegetation Indices for Discrimination of Soybean Areas: A New Approach. **Agronomy Journal**, v. 109, n. 4, p. 1-13, maio. 2017.

SOUZA, C. H. de W.; MERCANTE, E.; JOHANN, J. A.; LAMPARELLI, R. A. C.; URIBE- OPAZO, M. A. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. **International Journal of Remote Sensing**, v. 36, n. 7, p. 1809-1824, 2015. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/01431161.2015.1026956>>. Acesso em: 10 dez. 2017.

WILLMOTT, C. J.; ROBERSON, S. M.; MATSUURA, K. A Refined Index of Model Performance. **International Journal of Climatology**, v. 32, p. 2088-2094, 2012.

ZHANG, H.; LI, Q.; LIU, J.; JSHANG, J.; DU, X.; ZHAO, L.; WANG, N.; DONG, T. Crop Classification and Acreage Estimation in North Korea Using Phenology Features. **Giscience & Remote Sensing**, v. 54, n. 3, p. 1-26, 2017.

ZHONG, L.; HU, L.; YU, L.; GONG, P.; BIGING, G. S. Automated mapping of soybean and corn using phenology. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 119, p. 151-164, 2016.

7 CONSIDERAÇÕES FINAIS

Para o mapeamento de alvos permanentes e áreas agrícolas a metodologia, semiautomática proposta produziu mapeamentos com acurácia espacial excelente, tomando-se como referência as imagens falsa-cor RGB564 do satélite Landsat-8 para os quatro anos-safra (2013/14, 2014/15, 2015/16, 2016/17), apresentando ainda alta correlação com os dados oficiais. De acordo com os métodos de validação, a metodologia proposta obteve êxito na separação de alvos agrícolas e de alvos permanentes.

Para o mapeamento dos alvos floresta/reflorestamento, demais alvos permanentes, culturas semi-perenes, soja, feijão, milho 1ª e 2ª safra e culturas de inverno, os resultados dos mapeamentos de cada ano-safra são considerados bons e excelentes de acordo com o índice Kappa, tendo como base amostras extraídas de imagens falsa-cor RGB564 do satélite Landsat-8. Além disso, as estimativas das áreas agrícolas, em escala municipal, apresentaram de moderada a forte correlação, quando comparadas com dados oficiais.

Conclui-se que as métricas elaboradas de forma automatizada obtiveram êxito no mapeamento de alvos agrícolas, alvos permanentes, culturas semi-perenes, soja/feijão, milho 1ª e 2ª safra e culturas de inverno, podendo estas ser utilizadas em mapeamentos posteriores.

As rotinas desenvolvidas no software R obtiveram êxito na semiautomatização do processo KDD, utilizando classificação com o algoritmo Random Forest. Desse modo, na elaboração de mapeamentos futuros o tempo destes será reduzido.

Desse modo, os resultados obtidos podem ser utilizados para verificação das áreas agrícolas tanto do estado quanto do município, auxiliando em tomadas de decisão e possibilitando realizar estimativas de repasse de ICMS. Além disso, estas informações permitem que o município conteste o órgão oficial quanto à efetiva área de produção agrícola, caso haja divergência com os dados oficiais, tendo em vista que esses dados são utilizados para o cálculo do FPM.

Do mesmo modo, as áreas de florestas mapeadas podem ser usadas para estimar o ICMS ecológico e possibilita aos municípios obter dados da localização dessas áreas, permitindo identificar novas áreas de incidência do benefício, visto que 153 municípios paranaenses não recebem esse recurso.

Futuros trabalhos podem realizar a aplicação das métricas e da metodologia deste trabalho em imagens de menor resolução espacial, com o intuito de melhorar a acurácia dos mapeamentos. Além disso, a metodologia pode ser repetida para outras regiões e estados.

8 ANEXO – TUTORIAL DOS MAPEAMENTOS UTILIZADOS

**Tutorial de mapeamento utilizando software
R**

**Weverton Rodrigo Verica
2018**

Sumário

1	<u>Seleção de dados (Download das imagens)</u>	101
2	<u>Pré-processamento (Aplicação de Filtro de suavização)</u>	102
2.1	<u>Filtro SG</u>	102
3.	<u>Formatação (Script no R)</u>	103
3.1	<u>Diretório e cubo de imagens</u>	104
3.2	<u>Recorte das imagens</u>	104
3.3	<u>Métricas</u>	105
3.3.1	<u>Máximo, Mínimo</u>	106
3.3.2	<u>Amplitude</u>	106
3.3.3	<u>Desvio padrão</u>	106
3.3.4	<u>Média e Coeficiente de variação</u>	107
3.3.5	<u>Data de máximo e mínimo do EVI</u>	107
3.3.6	<u>Diferença das datas</u>	108
3.3.7	<u>Valores de máximo e mínimo da taxa de variação</u>	108
3.3.8	<u>Datas do máximo e mínimo da taxa de variação</u>	109
3.3.9	<u>Inflexão</u>	110
3.5	<u>Diretório de imagens de saída do R</u>	112
4.	<u>Mineração de dados (Classificação)</u>	112
4.1	<u>Conjunto de amostras</u>	113
4.2	<u>Conjunto de treinamento</u>	114
4.3	<u>Classificador</u>	114
5.	<u>Avaliação</u>	115
5.1	<u>Conjunto de validação</u>	115
5.2	<u>Comparação com outros mapeamentos</u>	116

Este tutorial tem como finalidade mostrar o passo a passo da elaboração de um mapeamento com série temporal de imagens MODIS, utilizando o software R e o processo de descoberta de conhecimento em base de dados (KDD).

Como exemplo deste tutorial foi elaborado um mapeamento de área agrícola e alvo permanente utilizando EVI/MODIS para o ano safra 2013/14 no estado do Paraná, sendo este dividido em cinco etapas: seleção de dados, pré-processamento, formatação/transformação, mineração de dados e avaliação/interpretação. Na sequência será descrito detalhadamente cada etapa do mapeamento.

1 Seleção de dados (Download das imagens)

O objetivo desta etapa é selecionar os dados utilizados no processo KDD, ou seja, as imagens do local que será mapeado. Nesta etapa também deve ser determinado o período da série temporal.

No exemplo tem-se que a área a ser mapeada é o Paraná e o período da série temporal é de agosto de 2013 a dezembro de 2014. Desse modo, foram selecionadas todas as imagens de EVI/MODIS dentro desse intervalo de tempo, totalizando 64 datas.

Para o funcionamento do mapeamento é essencial que as imagens tenham a mesma extensão e o mesmo intervalo de tempo entre elas. Para realizar os downloads das imagens pode ser utilizado o site da Embrapa (disponível em: <<https://www.modis.cnptia.embrapa.br>>).

Após realizar o download das imagens, insira uma nova pasta na área de trabalho do seu computador e copie as imagens obtidas nesta pasta.

OBS.: Na pasta onde serão realocadas as imagens não deve conter nenhum outro arquivo.

Na sequência, as imagens devem ser renomeadas de acordo com sua data, ou seja, a imagem mais antiga deve se chamar “D1”, a próxima “D2” e assim por diante. No exemplo, tem-se a seguinte nomenclatura:

Datas	Nome do arquivo
05/08/2013	D1
13/08/2013	D2
21/08/2013	D3
29/08/2013	D4
06/09/2013	D5
14/09/2013	D6
⋮	⋮
19/12/2014	D63
27/12/2014	D64

2 Pré-processamento (Aplicação de Filtro de suavização)

Na sequência se inicia a segunda etapa do KDD, o pré-processamento. Nesta etapa são realizados os ajustes nos dados selecionados, e para isso aplica-se o filtro de suavização Savitzck-Golay (SG). Esse processo pode ser realizado no software Crop-Yield Modeling Platform (Cymp).

2.1 Filtro SG

O software Cymp (Figura 1) é um software livre e de código aberto, sendo este desenvolvido para estimar produtividade de culturas agrícolas em larga escala. Além disso, o software fornece ferramentas que permite o processamento temporal e espacial das imagens.



Figura 41 Tela inicial do software Cymp

Uma dessas ferramentas é o filtro Savitzck-Golay, sendo que, para a aplicação deste, é necessário selecionar a opção “tratamento de dados” e na sequência a opção “filtro Savitzck-Golay”, conforme Figura 2.

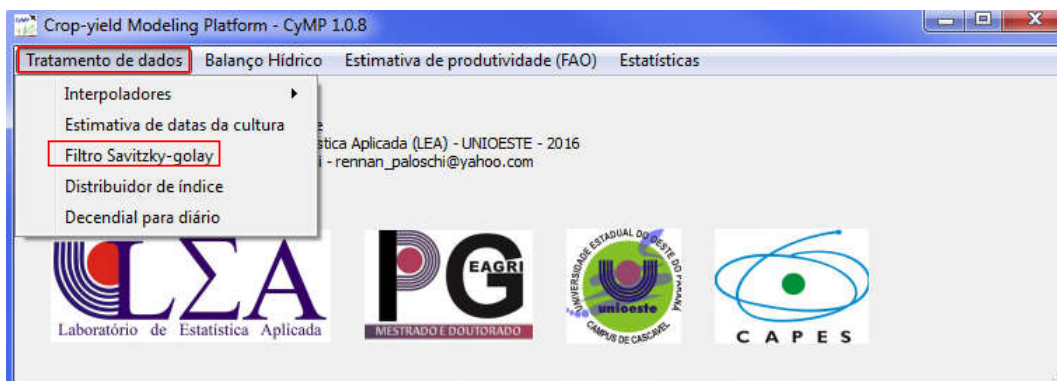


Figura 42 Caminho para utiliza o filtro Savitzck-Golay no Cymp

Após realizar esse procedimento deve-se indicar o local das imagens que serão submetidas ao filtro (neste caso é a pasta criada na seção 1), e o local em que serão armazenadas as novas imagens, conforme Figura 3.

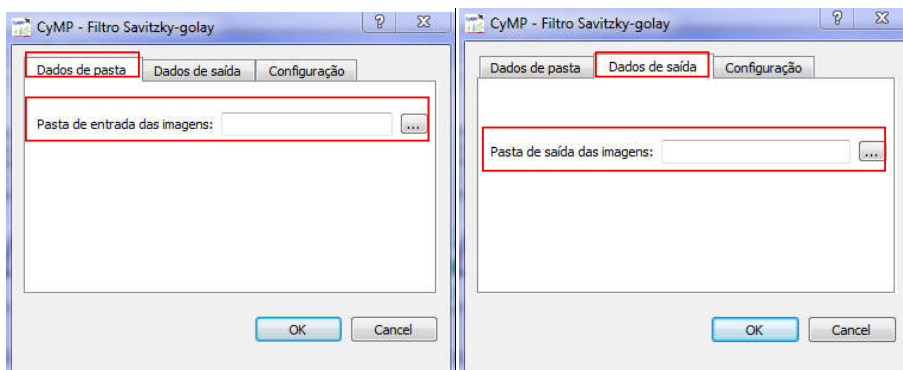


Figura 43 Tela de seleção dos dados de entrada e dados de saída

Por fim, determinam-se os parâmetros para a aplicação do filtro, conforme Figura 4.

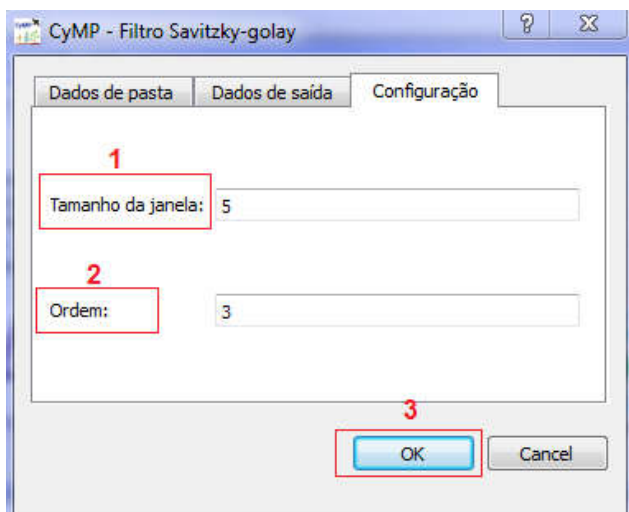


Figura 44 Configurações do filtro

3. Formatação (Script no R)

Para realizar as demais etapas do KDD utilizou-se uma rotina implementada no software R com o intuito de automatizar o mapeamento, usando como dado de entrada imagens filtradas de um ano-safra. Para utilizar esta rotina é necessário possuir os pacotes: caret, rgdal, raster e stats.

3.1 Diretório e cubo de imagens

Na sequência será demonstrado o script utilizado para realizar o mapeamento do ano safra 2013/14, sendo que o caminho do diretório utilizado foi “D:/Meus Documentos/13-14/saida cymp/”, deve ser alterado conforme o local onde foram salvas as imagens filtradas na seção 2.

Para realizar a leitura das imagens na forma de série temporal é necessário construir um cubo de imagens com a função brick, conforme rotina a seguir.

```
require(caret)
require(rgdal)
require(raster)
require(stats)
} Carregando os pacotes

setwd("D:/Meus Documentos/13-14/saida cymp/")
getwd()
} Diretório das imagens

imglist <- list.files(pattern = ".tif")
imglist
} Lista das imagens

img <- stack(list.files(pattern=".tif$"))
img = brick(img)
names(img) <- c(paste0("D", 1:64, coll = "")) #
names(img)
} Elaboração do cubo de imagens

img
} Cubo de imagens (Figura 5)
```

```
> img
class       : RasterBrick
dimensions  : 1872, 2939, 5501808, 64 (nrow, ncol, ncell, nlayers)
resolution  : 0.002245, 0.002245 (x, y)
extent      : -54.61828, -48.02022, -26.71922, -22.51658 (xmin, xmax, ymin, ymax)
coord. ref. : NA
data source : C:\Users\Microtec\AppData\Local\Temp\RtmpawUtZ3\raster\r_tmp_2017-12-03_112322_1040_10269.grd
names       : D1, D2, D3, D4, D5, D6, D7, D8, D9
min values  : -3000.000, -3732.771, -4392.771, -4514.743, -4396.029, -4739.143, -4342.286, -4518.343, -4348.800,
max values  : 8123.629, 8327.029, 9860.514, 9685.200, 9780.229, 10440.829, 10890.714, 11066.171, 10907.143,
```

Figura 45 Informações do cubo de imagem.

- 1 → Extensão das imagens.
- 2 → Número de linhas de cada imagem.
- 3 → Número de coluna de cada imagem.
- 4 → Número de células de cada imagem.
- 5 → Número de imagens contidas no cubo de imagem.
- 6 → Valores de máximo e mínimo de cada imagem.

3.2 Recorte das imagens

Para estudar apenas uma área pré-estabelecida é possível utilizar um arquivo shapefile desta área com o intuito de recortar o cubo de imagem, diminuindo assim o tempo computacional do mapeamento. Esse comando pode ser realizado da seguinte maneira:


```

setwd("D:/Meus Documentos/shapes/perimetro pr/") } local do shape

pr <- shapefile("Perimetro_PR_wgs84.shp") } Carregando o shape
pr

r2 = crop(img, extent(pr)) }
r3 = mask (r2, pr) } Recorando as imagens

```

De modo que "Perimetro_PR_wgs84.shp" é o nome do arquivo shapefile da área que se pretende estudar no exemplo utilizado.

Com o cubo de imagens preparado, o próximo passo é extrair as métricas desse cubo.

3.3 Métricas

As métricas são variáveis elaboradas com a finalidade de auxiliar na separação dos alvos; então, a confecção destas pode variar. No exemplo será mostrado o cálculo de 15 métricas; entretanto, podem ser elaboradas mais destas, conforme a demanda do mapeamento.

Para realizar o cálculo das métricas é fundamental compreender o que está sendo analisado pelo software.

O produto que será utilizado para elaboração das métricas é o perfil espectro-temporal de cada pixel (Figura 6b), extraído de cubo de imagens (figura 6a), sendo que cada pixel do cubo de imagem gera um perfil espectro-temporal.

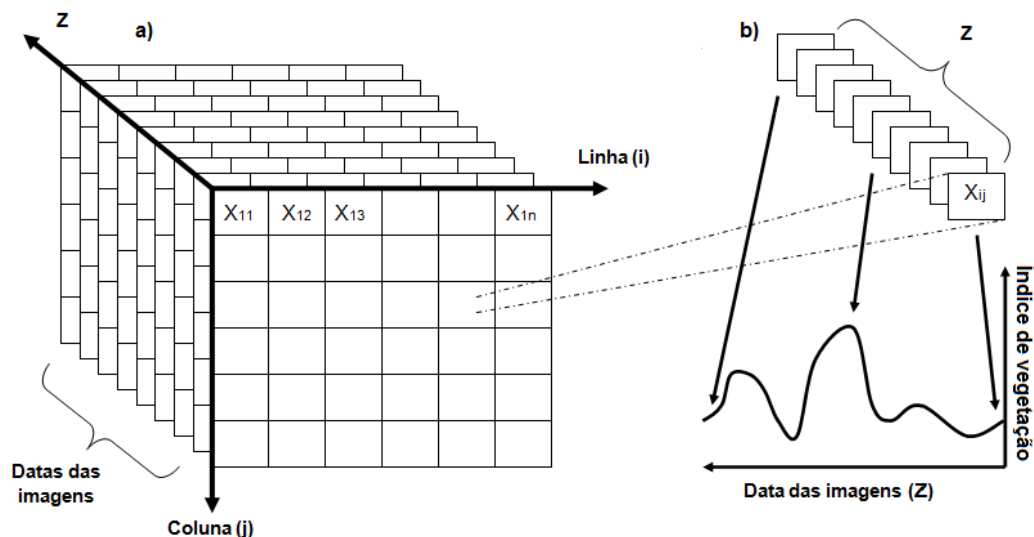


Figura 46 Método de extração dos perfil espectro-temporal de cada um dos pixels do cubo de imagens

Dessa forma, o software R trabalha com as resposta de refletância de cada imagem ao decorrer do tempo, ou seja, um perfil espectro temporal. Visto isso, analisaremos as 15 métricas construídas.

Antes de calcular as métricas, é necessário destinar uma pasta onde estas serão salvas; no exemplo, o local escolhido para o armazenamento foi "C:/Users/weveton.verica/saida r/", lembrando que cada métrica vai gerar um arquivo raster.

```
setwd("C:/Users/weveton.verica/saida r/")
getwd()
```

} Diretório das imagens de saída do R

3.3.1 Máximo, Mínimo

Essas métricas encontram os valores máximos e mínimos do EVI para cada pixel da imagem, por meio dos seguintes comandos:

```
max = max(r3)
min = min(r3)
```

} Métricas de máximo e mínimo

```
plot (max)
plot (min)
```

} Distribuição especial das métricas

```
setwd("D:/Meus Documentos/13-14/saida r/")
writeRaster(max,"F1.tif","GTiff", overwrite=TRUE)
writeRaster(min,"F2.tif","GTiff", overwrite=TRUE)
```

} Salvando as métricas na pasta pré-estabelecida

Sendo que a métrica de máximo EVI foi salva com o nome F1.tif e a métrica de mínimo EVI com o nome F2.tif.

3.3.2 Amplitude

Essa métrica calcula a diferença entre as duas métricas anteriores, da seguinte maneira:

```
amp = max - min
plot (amp)
writeRaster(min,"F3.tif","GTiff", overwrite=TRUE)
```

} Salvando a métrica na pasta pré-estabelecida

3.3.3 Desvio padrão

Para calcular o desvio padrão de cada pixel, primeiro é necessário obter o quadrado da diferença de cada imagem com a média, e esse procedimento é realizado da seguinte forma:

```
sd1 = ((img1 - med)^2)/X
sd2 = ((img2 - med)^2)/X
sd3 = ((img3 - med)^2)/X
⋮
sdX = ((imgX - med)^2)/X
```

} Onde X é o número total de imagens

Com isso pode-se calcular a variância e posteriormente o desvio padrão, da seguinte maneira:

```
var = (sd1 + sd2 + sd3 + ... + sdX)
sd = (var^(1/2))
plot (sd)
writeRaster(sd,"F5.tif","GTiff", overwrite=TRUE)
```

Métrica do desvio padrão, salva como F5.tif

3.3.4 Média e Coeficiente de variação

Estas métricas estatísticas são facilmente calculadas:

```
med = mean(r3)
plot (med)
writeRaster(med,"F4.tif","GTiff", overwrite=TRUE)
```

Métrica da média

```
cv = (sd/med)*100
plot (cv)
writeRaster(cv,"F6.tif","GTiff", overwrite=TRUE)
```

Métrica de Coeficiente de variação

3.3.5 Data de máximo e mínimo do EVI

Para calcular a data que ocorreu o mínimo de EVI na série temporal é necessário avaliar todas as datas separadamente, conforme:

```
v1 = img1/min
v1 [1 == v1] <- 1
v1 [1 != v1] <- 0
v1
```

Analisando a imagem 1, caso o valor do pixel desta imagem seja igual ao valor mínimo deste pixel para todas as imagens atribui-se o valor 1, caso contrário o valor atribuído é 0.

```
v2 = (2*img2)/min
v2 [2 == v2] <- 2
v2 [2 != v2] <- 0
v2
```

Esta análise é repetida para todas as x imagens, sendo que, caso o valor do pixel seja igual ao valor mínimo deste pixel para todas as imagens atribui-se o valor x, caso contrário o valor atribuído é 0.

Após analisarmos todas as imagens (no script de exemplo são 64 imagens), calculamos o máximo dentre os valores encontrados, determinando assim em qual data este valor foi encontrado, sendo que as datas são enumeradas de 1 até o número de imagens (no exemplo é de 1 a 64).

```
dm = max(v1, v2, v3,... , v63, v64)
plot (dm)
writeRaster(dm,"F7.tif","GTiff", overwrite=TRUE)
```

Métrica da data de mínimo EVI

Para o cálculo da data em que ocorreu o máximo EVI, o procedimento é similar, conforme:

```
v1 = img1/max
v1 [1 == v1] <- 1
v1 [1 != v1] <- 0
v1
```

Analisando a imagem 1, caso o valor do pixel desta imagem seja igual ao valor máximo deste pixel para todas as imagens atribui-se o valor 1, caso contrário o valor atribuído é 0.

O procedimento realizado para cálculo da data do mínimo EVI se repete para as demais imagens, alterando apenas o comando “min” para “max”. Após analisar todas as imagens, calcula-se a métrica.

```
dmax = max(v1, v2, v3, ..., v63, v64)
plot (dmax)
writeRaster(dmax,"F8.tif","GTiff", overwrite=TRUE)
```

Métrica da data de máximo EVI

3.3.6 Diferença das datas

Essa métrica calcula a diferença entre as duas datas citadas anteriormente, da seguinte maneira:

```
difd = abs(dmax - dm)
plot (difd)
writeRaster(difd,"F9.tif","GTiff", overwrite=TRUE)
```

Métrica da data da diferença de máximo e mínimo EVI

3.3.7 Valores de máximo e mínimo da taxa de variação

Para isso, primeiramente é necessário construir outro perfil espectro-temporal, sendo que este represente a taxa de variação (1ª derivada).

Primeiramente, é necessário atribuir um nome para cada imagem utilizada, da seguinte maneira:

```
setwd("C:/Users/weveton.verica/saida cymp/")
getwd()
```

Diretório das imagens

```
img1 = raster("D1.tif")
img2 = raster("D2.tif")
img3 = raster("D3.tif")
img4 = raster("D4.tif")
⋮
imgX = raster("DX.tif")
```

Utilização do comando “Raster” para renomear cada uma das imagens da série temporal.

Na sequência, calcula-se a variação das datas, duas a duas:

```
dif1 = img2 - img1
dif2 = img3 - img2
dif3 = img4 - img3
dif4 = img5 - img4
:
dif(X-1) = imgX - img(X-1)
```

Sendo X o número de imagens da série temporal, então o cálculo da diferença será realizado para "X-1" pares de imagens.

Para calcular o valor da máxima inclinação utilizam-se como dados de entrada as variações calculadas anteriormente (dif1, dif2, dif3,..., dif(X-1)), conforme exemplo a seguir (sendo que no exemplo o número de imagens é 64):

```
max1d = max(dif1, dif2, dif3,..., dif62, dif63)
plot (max1d)
writeRaster(max1d,"F10.tif","GTiff", overwrite=TRUE)
```

Métrica de máximo valor de inclinação.

Para o valor da menor inclinação, o procedimento é semelhante, conforme exemplo a seguir:

```
min1d = min(dif1, dif2, dif3,..., dif62, dif63)
plot (min1d)
writeRaster(min1d,"F11.tif","GTiff", overwrite=TRUE)
```

Métrica do menor valor de inclinação.

3.3.8 Datas do máximo e mínimo da taxa de variação

O processo também é similar ao utilizado anteriormente para se obter as datas de máximo EVI e mínimo EVI, bem como o período entre as duas:

```
v1 = dif1/min1d
v1 [1 == v1] <- 1
v1 [1 != v1] <- 0
v1
```

Analisando a imagem dif1, caso o valor do pixel desta imagem seja igual ao valor mínimo deste pixel para todas as imagens atribui-se o valor 1, caso contrário o valor atribuído é 0.

```
v2 = (2*dif2)/min1d
v2 [2 == v2] <- 2
v2 [2 != v2] <- 0
v2
```

Esta análise é repetida para todas as x imagens, sendo que, caso o valor do pixel seja igual ao valor mínimo da taxa de variação para este pixel em todas as imagens atribui-se o valor x, caso contrário o valor atribuído é 0.

Após analisarmos todas as imagens (no script de exemplo são 63 imagens), calculamos o máximo dentre os valores encontrados, determinando assim em qual data este valor foi encontrado, sendo que as datas são enumeradas de 1 até o número de imagens (no exemplo é de 1 a 63).

```
dmin1d = max(t1, t2, t3, ... , t63)
writeRaster(dmin1d,"F12.tif","GTiff", overwrite=TRUE)
```

Métrica da data da variação mínima

Esse procedimento é repetido para calcular a data da máxima diferença; entretanto, nos locais que se usou min1d no script, passa-se a usar max1d para calcular a data de máximo, e o comando para salvar esta métrica é o seguinte:

```
dmax1d = max(t1, t2, t3, ... , t63)
writeRaster(dmax1d,"F13.tif","GTiff", overwrite=TRUE)
```

Métrica da data da variação máxima

Com base nas métricas calculadas anteriormente, calcula-se a décima quarta métrica, onde está representa a diferença entre as duas datas citadas anteriormente, da seguinte maneira:

```
dif1d = abs(dmax1d - dmin1d)
plot (dif1d)
writeRaster(dif1d,"F14.tif","GTiff", overwrite=TRUE)
```

Métrica da diferença entre as datas de variação

3.3.9 Inflexão

Essa métrica soma todos os pontos onde a taxa de variação é igual a zero; para isso, é calculado o produto dos valores de inclinação (dif1, dif2, dif3,..., dif(X-1)), dois a dois, da seguinte maneira:

```
w1 = dif1 * dif2
w2 = dif2 * dif3
w3 = dif3 * dif4
w4 = dif4 * dif5
⋮
w(X-2) = dif(X-2) * dif(X-1)
```

Sendo X o número de imagens da série temporal, então o cálculo do produto das diferenças será realizado para "X-2" pares de imagens

Caso o resultado do produto apresente valor positivo, isso indica que não houve inflexão. Desse modo, esta métrica consiste em somar quantos produtos resultaram em um valor negativo; para isso utilizaram-se as seguintes funções do software R:

```
w1 [0 > w1] <- 1
w1 [0 <= w1] <- 0
```

Verificação se o produto é positivo ou negativo

```
inf = (w1 + w2 + w3 + ... + w61 + w62)
writeRaster(inf,"F15.tif","GTiff", overwrite=TRUE)
plot(inf)
```

Métrica de inflexão

3.4 Reorganizando a extensão das imagens

Caso tenha utilizado um shape de uma área específica para realizar o corte do cubo de imagem (seção 3.2), então deve-se realizar o mesmo procedimento nas métricas 3.3.7 a 3.3.9, pois essas não utilizam o cubo de imagem, e sim uma imagem por vez.

O procedimento para recortar é necessário para deixar todas as imagens geradas com a mesma extensão, e a metodologia para isto é similar à realizada na seção 3.1.

```
i10= raster("F10.tif")
i11= raster("F11.tif")
i12= raster("F12.tif")
i13= raster("F13.tif")
i14= raster("F14.tif")
i15= raster("F15.tif")
```

Renomeia os arquivos rasters das métricas a serem recortadas

```
setwd("D:/Weverton/Shapefiles/Paraná/Perimetro_PR/")
pr <- shapefile("Perimetro_PR_wgs84.shp")
```

Shape da área

```
r20 = crop(i10, extent(pr))
i10 = mask (r20, pr)

r21 = crop(i11, extent(pr))
i11 = mask (r21, pr)

r22 = crop(i12, extent(pr))
i12 = mask (r22, pr)

r23 = crop(i13, extent(pr))
i13 = mask (r23, pr)

r24 = crop(i14, extent(pr))
i14 = mask (r24, pr)

r25 = crop(i15, extent(pr))
i15 = mask (r25, pr)
```

Processo de recorte das imagens utilizando um arquivo shape

```
writeRaster(i10,"F10.tif","GTiff", overwrite=TRUE)
writeRaster(i11,"F11.tif","GTiff", overwrite=TRUE)
writeRaster(i12,"F12.tif","GTiff", overwrite=TRUE)
writeRaster(i13,"F13.tif","GTiff", overwrite=TRUE)
writeRaster(i14,"F14.tif","GTiff", overwrite=TRUE)
writeRaster(i15,"F15.tif","GTiff", overwrite=TRUE)
```

Substitui arquivos por rasters recortados

Garantindo, assim, que todas as imagens tenham as mesmas extensões.

3.5 Diretório de imagens de saída do R

Desse modo, tem-se uma imagem para cada métrica calculada, podendo formar um novo cubo de imagens, composto apenas pelas métricas:

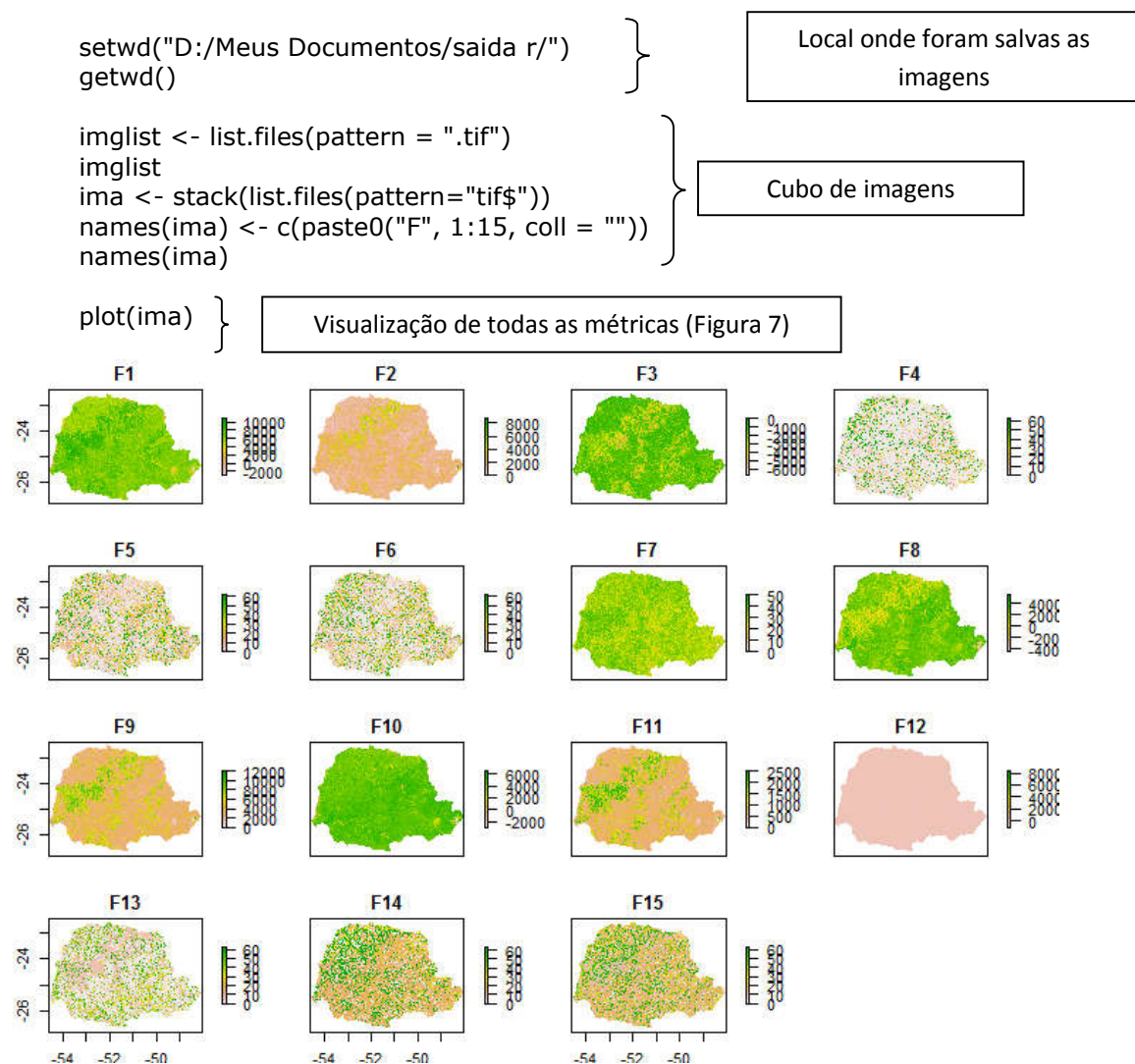


Figura 47 Plot de cada uma das 15 métricas

Esse novo cubo de imagens será utilizado como dado de entrada para o processo de mapeamento, sendo este também realizado no software R.

4. Mineração de dados (Classificação)

Para realizar a tarefa de classificação é necessário possuir um conjunto de amostras com a variável resposta, podendo este ser adquirido por meio de verdade de campo ou por meio de bibliotecas espectro-temporais.

O conjunto de amostra é utilizado para gerar as regras de classificação e, em seguida, aplicam-se estas regras em todos os dados. Além disso, o conjunto de amostra também pode ser utilizado para validar o classificador.

4.1 Conjunto de amostras

Para realizar este mapeamento, o conjunto de amostras será utilizado para classificação e para a avaliação, sendo que, para a elaboração deste conjunto, foi utilizada uma imagem de resolução espacial superior (Landsat-8) para selecionar as amostras de alvo agrícola (soja, milho, cana-de-açúcar, pastagem) e alvo permanente (floresta, corpos hídricos e áreas urbanas). Esse processo foi realizado no software arcgis, gerando um arquivo shapefile com todas as amostras coletadas.

OBS.: Na tabela de atributos do shapefile é necessário ter uma coluna “classe” com a informação de qual classe o pixel do conjunto de amostras pertence

Tendo essas informações, utiliza-se esse shape como entrada no R, da seguinte maneira:

```
trainData <- shapefile("D:/Arq.shp")
trainData
```

Local do arquivo shapefile das amostras

As informações do arquivo com as amostras (traindata) podem ser observadas na Figura 8.

```
> trainData
class      : SpatialPolygonsDataFrame
features   : 3207
extent     : -54.32418, -48.95863, -26.26573, -22.57495 (xmin, xmax, ymin, ymax)
coord. ref.: +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0
variables  : 6
names      : ID, GRIDCODE, Classe, OBJECTID, Shape_Leng, Shape_Area
min values : 0, 0, 1, 0, 0.000000000, 0.000000e+00
max values : 999926, 9982, 8, 9, 0.022449998, 5.040026e-06
```

Figura 48 Conjunto de treinamento

- 1 – Número de pixel do conjunto de amostra
- 2 – Número de variável do conjunto de amostra
- 3 – Variável resposta (coluna que informa qual alvo o pixel possui)

Para buscar a variável resposta (Classe) no arquivo shape, utiliza-se o seguinte comando:

```
responseCol <- "Classe"
responseCol
```

Determina a variável resposta

Para utilizar o conjunto de amostra para classificar e avaliar é necessário dividir o conjunto em dois, um denominado de traindata para gerar as regras do mapeamento e outro denominado de conjunto de validação para avaliar o mesmo.

4.2 Conjunto de treinamento

Para dividir o conjunto de amostras em dois foram utilizadas as proporções de 75% para o conjunto de treinamento e 25% para o conjunto de validação, sendo que a distribuição foi realizada de forma aleatória, conforme exemplo a seguir, com 3207 amostras:

```
CA = (1:3207)
ConjuntoAmostra = sample (CA, (3207), replace = F) } Elaboração do
Conjunto de amostras

conjuntoTreino= ConjuntoAmostra [1:2502]
conjuntoValidação = ConjuntoAmostra [2503:3207] } Elaboração dos
Conjuntos de
treinamento e avaliação

x1 = trainData[conjuntoTreino,]
x2 = trainData[conjuntoValidação,] } Transformando os conjuntos em
arquivos shapefile
```

Após a elaboração do conjunto treinamento este será utilizado como entrada para o classificador.

4.3 Classificador

Primeiramente constrói-se uma tabela que associa cada pixel do conjunto com sua variável resposta, ou seja, com a informação contida no conjunto de amostras.

```
x11 = x1[[responseCol]] } Vetor com a variável resposta

dfAll<-data.frame(matrix(vector(), nrow = 0, ncol = length(names(ima))+1))
dfAll } Elabora uma coluna para cada métrica mais a coluna da variável resposta

#extraí por pixels do img nos locais do shp
for (i in 1:length(unique(x11))) {
  category <- unique(x11)[i]
  categorymap <- x1[x11 == category,]
  dataSet <- extract(ima, categorymap)
  dataSet <- lapply(dataSet, function(x){cbind(x, ID =
as.numeric(rep(category, nrow(x))))})
  df <- do.call("rbind", dataSet)
  dfAll <- rbind(dfAll, df)
}
```

O produto deste comando é uma tabela com todos os pixels do conjunto de amostras e seus respectivos valores das métricas e variável resposta.

```
nsamples <- 2502
dfAll <- subset(dfAll[sample(1:nrow(dfAll), nsamples), ])
dfAll
```

O valor do nsamples pode ser igual ao valor do conjunto de treinamento, ou seja, 2502. Após a elaboração dos dados de entrada é escolhido o algoritmo para classificar, sendo que

no exemplo utilizou-se o Random Forest. Para utilizar essa rotina, primeiramente é elaborado um modelo para a classificação, conforme comando a seguir:

```
modFit_rf <- train(as.factor(ID) ~ F1 + F2 + F3 + F4 + F5 + F6+ F7+ F8+ F9 +
F10+F11 +F12+ F13+ F14 + F15, method = "rf", data = dfAll)
```

Na sequência, utiliza-se este modelo juntamente com o cubo de imagens para realizar a classificação, conforme script a seguir:

```
beginCluster()
preds_rf <- clusterR(ima, raster::predict, args = list(model = modFit_rf))
endCluster()
```

Com isso, pode-se visualizar a classificação realizada e salvá-la:

```
plot(preds_rf) } Visualiza a classificação.
setwd("D:/")
writeRaster(preds_rf,"clas.tif","GTiff", overwrite=TRUE) } Salva a imagem
da classificação
```

Após realizar a classificação, esta deve ser avaliada para verificar sua exatidão, neste tutorial serão abordadas duas formas de avaliação que podem ser feitas por meio do software R.

5. Avaliação

Para realizar a avaliação do mapeamento serão abordadas duas metodologias utilizando conjunto de validação comparando com outros mapeamentos.

5.1 Conjunto de validação

Esse método consiste em utilizar o conjunto de validação formado no tópico 4.2, conforme a seguinte rotina:

```
resposta <- extract(preds_rf, x2)
resposta
x = t(resposta)
for (i in 1:614){ h = x[[i]]}
for (i in 1:614){ h[i] = x[[i]]}
h
x22 = x2[[responseCol]] }
```

Extrai do mapeamento as informações de x2. Lembrando que em 4.2 foi definido x2 como o conjunto de validação

Extrai a verdade de campo do conjunto x2.

```
mc = confusionMatrix(h,x22)
mc
```

Cálculo da matriz de confusão dos valores preditos (h) com a referência (x22).

Ao final dessa rotina tem-se uma matriz de confusão, e por meio desta é possível de calcular Kappa, Exatidão Global, Erro de omissão e Erro de inclusão.

5.2 Comparação com outros mapeamentos

Esse método é utilizado quando possuir mapeamentos previamente realizados da mesma área de estudo, e o objetivo é calcular o erro entre o mapeamento elaborado e o mapeamento já existente. Porém, para isso é necessário possuir o arquivo tiff desses mapeamentos:

```
setwd("D:/Meus Documentos/Denise/Milho/")
getwd()
```

Local do arquivo tiff do mapeamento utilizado na comparação.

```
img1 = raster("Mosaico_Milho_Pr_20132014_0e1.tif")
img1
plot(img1)
```

Carregando e visualizando a imagem do mapeamento.

Para realizar a comparação, é necessário que as extensões das imagens sejam idênticas; portanto, é recomendável realizar o recorte das imagens de referência com o mesmo shapefile utilizado para recortar o presente mapeamento.

```
setwd("D:/Meus Documentos/perimetro pr/")
pr <- shapefile("Perimetro_PR_wgs84.shp")
pr
```

Recortando o mapeamento utilizado como referência

```
r2 = crop(v1,extent(pr))
r3 = mask (r2, pr)
```

Com ambas as imagens na mesma extensão, estas são comparadas:

```
ref = r3[]
pred = v2[]
pred
```

ref é a imagem de referência e
pred é o mapeamento construído

```
mc = confusionMatrix(pred,ref)
mc
```

Cálculo da matriz de confusão

Neste caso, também é gerada uma matriz de confusão e, por meio desta, é possível de calcular o Kappa, Exatidão Global, Erro de Omissão e Erro de Inclusão.