



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2015/06.24.17.55-TDI

**ESTIMATIVA DE PRODUTIVIDADE DA
CANA-DE-AÇÚCAR COM BASE NA ASSOCIAÇÃO
ENTRE SÉRIES DE DADOS MODIS E ESTATÍSTICAS
AGRÍCOLAS**

William Foschiera

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelos Drs. Antônio Roberto
Formaggio, Márcio Pupin de
Mello, aprovada em 30 de junho
de 2015.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3JNHAGS>>

INPE
São José dos Campos
2015

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

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

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas
(CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos
(CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

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

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

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

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

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

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



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2015/06.24.17.55-TDI

**ESTIMATIVA DE PRODUTIVIDADE DA
CANA-DE-AÇÚCAR COM BASE NA ASSOCIAÇÃO
ENTRE SÉRIES DE DADOS MODIS E ESTATÍSTICAS
AGRÍCOLAS**

William Foschiera

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelos Drs. Antônio Roberto
Formaggio, Márcio Pupin de
Mello, aprovada em 30 de junho
de 2015.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3JNHAGS>>

INPE

São José dos Campos

2015

Foschiera, William.

F783e Estimativa de produtividade da cana-de-açúcar com base na associação entre séries de dados MODIS e estatísticas agrícolas / William Foschiera. – São José dos Campos : INPE, 2015.

xxii + 73 p. ; (sid.inpe.br/mtc-m21b/2015/06.24.17.55-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2015.

Orientadores : Drs. Antônio Roberto Formaggio, e Márcio Pupin de Mello.

1. Produtividade agrícola. 2. NDVI. 3. MODIS. 4. Estimativa de produtividade. I. Título.

CDU 528.835:633.61



Esta obra foi licenciada sob uma [Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](#).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#).

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

Dr. Antônio Roberto Formaggio


Presidente / Orientador(a) / INPE / SJCampos - SP

Dr. Márcio Pupin de Mello


Orientador(a) / Boeing Res & Tecno / São José dos Campos - SP

Dra. Ieda Del' Arco Sanches


Membro da Banca / INPE / São José dos Campos - SP

Dr. Alfredo José Barreto Luiz


Convidado(a) / EMBRAPA / Jaguariúna - SP

Este trabalho foi aprovado por:

() maioria simples

☒ unanimidade

Título: "Estimativa de produtividade da cana-de-açúcar com base na associação entre séries de dados MODIS e estatísticas agrícolas."

Aluno (a): **William Foschiera**

São José dos Campos, 30 de Junho de 2015

The most effective learning requires a well-defined task with an appropriate difficulty level for the particular individual, informative feedback, and opportunities for repetition and corrections of errors.

Jean Lave

In Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life, 1988

A meus pais, Antoninho e Rosicler.

AGRADECIMENTOS

Ao Dr. José Carlos Neves Epiphanyo pela orientação no primeiro ano de mestrado e pelos ensinamentos transmitidos, dentro e fora da sala de aula.

Aos Drs. Antônio Roberto Formaggio e Márcio Pupin de Mello que me orientaram sabiamente durante o desenvolvimento deste trabalho.

Ao professor Érico Fernando de O. Martins, por ter sido o grande incentivador desse mestrado.

Ao Dr. Clement Atzberger e à Dra. Anja Klisch pelo fornecimento dos dados da universidade de BOKU.

Aos professores da DSR e DPI que compartilham seus conhecimentos e seu tempo, com o objetivo de formar profissionais melhores.

Aos amigos da pós-graduação Denis, Fábio, Everton, Édipo, Matheus, Vinícius, Thiago, Leandro, Ana Claudia, Lorena, Taíssa, Sílvia e Denise por toda ajuda, pelos grupos de estudo e, principalmente, pelos momentos de descontração, que tornaram a estadia no INPE tão agradável.

A meus pais e minhas irmãs pelo apoio nos momentos em que mais precisei e à Paula que soube entender minha ausência.

Ao Instituto Nacional de Pesquisas Espaciais (INPE) e à Divisão de Sensoriamento Remoto (DSR), pela oportunidade concedida para a realização do mestrado e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa concedida.

RESUMO

Esta pesquisa apresenta um método, para estimativa de produtividade da cana-de-açúcar, baseado na correlação entre dados de sensoriamento remoto e dados oficiais de produtividade. Este método utiliza uma abordagem estatística, para extrair diferentes células de uma série temporal, do Índice de Vegetação (IV) Normalized Difference Vegetation Index (NDVI) do sensor Moderate Resolution Spectroradiometer (MODIS), para serem utilizados como proxies, na estimativa de produtividade da cana-de-açúcar, em escala municipal, utilizando como área de estudo, o estado de São Paulo. Nesta pesquisa foram investigados 361 municípios, que possuíam os dados históricos de produtividade para o período de 2003 a 2013. Duas abordagens foram testadas para escalonar os dados de NDVI e de produtividade: Zscore e Rscore. Ademais, dois métodos foram utilizados como critério de seleção dos proxies: RMSE e ρ de Pearson. Os resultados mostram que a combinação Zscore e RMSE foi mais adequada para o método proposto. Os resultados foram agrupados, de acordo com a quantidade de área colhida em 2012 e por mesorregiões geográficas. Os resultados por grupo não apresentaram correlação entre a precisão das estimativas e quantidade de área colhida. Os resultados por mesorregião foram, em geral, bastante precisos, com diferenças menores que 5% comparado aos dados oficiais. No entanto para as mesorregiões de Presidente Prudente, Vale do Paraíba Paulista, Ribeirão Preto e Marília os erros foram de 12,05%, -11,96%, 7,23% e 7,38%, respectivamente. A precisão das estimativas, por mesorregiões, reforça o potencial deste método, que pode ser operacionalmente utilizado para estimar a produtividade da cana-de-açúcar para grandes regiões. Além disso, o método proposto poderá estimar a produtividade durante a safra, enquanto as estimativas oficiais são, geralmente, divulgadas um ou dois anos após a colheita.

SUGARCANE CROP YIELD ESTIMATION BASED ON MODIS AND OFFICIAL STATISTICS DATA ASSOCIATION

ABSTRACT

This research presents a method for sugarcane crop yield forecast, based on the correlation between remote sensing and official statistics yield data. The method uses a statistical approach to extract different pixels of multitemporal smoothed Normalized Difference Vegetation Index (NDVI) data derived from the Moderate Resolution Spectroradiometer (MODIS) to be used as proxies for sugarcane yield estimation at São Paulo State at municipal scale. In this research, 361 municipalities with yield's historical data from 2003 to 2012 were analyzed. Two different approaches were tested to normalize yield data and NDVI: Zscore and Rscore. In addition, two methods were used as selection criteria: RMSE and Pearson's correlation. Results showed that Zscore and RMSE are best to predict sugarcane yield. The results were grouped according to the amount of area harvested in 2012 and geographic regions. Results for groups based on acreage didn't show correlation between acreage amount and yield estimation accuracy. Results at mesorregions scale were, in general, accurated with small differences (5% or less) compared with official data. However, for Presidente Prudente, Vale do Paraíba Paulista, Ribeirão Preto and Marília, the differences between observed and estimated values was of 12,05%, -11,96%, 7,23% e 7,38%, respectively. The mesorregions results' accuracy reinforce the potential of this method to be operationally used for sugarcane yield forecast over large areas. Moreover, the proposed method may estimate yield early in crop season, whereas official statistics are usually published late after harvest.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 2.1 – Diagrama da composição do produto MOD13Q1, 'n' se refere ao número de observações com qualidade.....	13
Figura 2.2 – Técnica de filtragem do algoritmo de suavização Whittaker. Primeiramente é aplicada uma interação mais suave aos dados, para fechar as lacunas resultantes pela presença de nuvens, conhecido como ajuste ao invólucro superior. Em seguida, são identificados os pontos em que o IV original é menor que o valor filtrado. Para estes pontos, os valores originais são substituídos pelos valores filtrados e o processo de filtragem é repetido novamente. Depois de duas a três iterações, uma curva suavizada é obtida..	16
Figura 2.3 – Esquematização da filtragem dos dados de BOKU. Os produtos mais distantes da extremidade da série temporal tendem a ser mais precisos do que os mais próximos das extremidades. Assim, o produto D tende a ser mais preciso enquanto o produto N tende a ser menos preciso.....	19
Figura 3.1 – Fluxograma do método, apresentando as principais etapas de processamentos. As figuras cilíndricas representam o armazenamento dos dados, tanto originais quanto processados enquanto os paralelogramos representam cada uma das etapas de processamento dos dados.	22
Figura 3.2 – Área de estudo com destaque ao conjunto de 361 municípios analisados neste trabalho, apresentando a localização destes, no estado de São Paulo.....	23
Figura 3.3 – Limites municipais do estado de São Paulo e localização dos <i>minitiles</i> MODIS fornecidos pela universidade de BOKU/IVFL.....	25
Figura 3.4 – Exemplo da seleção dos <i>proxies</i> para estimativa da produtividade da cana-de-açúcar para o ano de 2012 do município de Piracicaba. As linhas pretas representam as variações anuais dos valores escalonados de NDVI de 50 células e a linha vermelha representa o valor escalonado de produtividade da cana-de-açúcar.	30
Figura 3.5 – Recorte da área de abrangência para o município de Bariri – SP,	

com raio de 90 km. Os valores das células dentro da área circular representam os valores do critério de seleção ρ de Pearson. Os valores destas células foram ordenados e os 200 valores mais próximos de 1 são utilizados como <i>proxies</i> para estimar a produtividade.....	31
Figura 3.6 – Diferença entre os valores estimados e observados para todas as semanas do ano de 2012 para cinco municípios do estado de São Paulo. Os melhores resultados ocorrem até o final do mês de março.....	35
Figura 4.1 – Comparativos das quatro combinações entre os métodos de escalonamento e critério de seleção dos proxies para os 361 municípios da área de estudo, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros.	38
Figura 4.2 – Exemplo com dados fictícios no qual são apresentados dois conjuntos de dados. A diferença em cada observação é de 0,05 unidades de medida. O ρ de Pearson, por avaliar a colinearidade dos valores, teve valor igual a um, enquanto que o RMSE, que mediu a magnitude do erro, conseguiu detectar as diferenças entre os valores de cada amostra.	40
Figura 4.3 – Comparativo entre os diferentes tamanhos da área de abrangência, para os 361 municípios da área de estudo, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros.....	41
Figura 4.4 – Erros médios para os quatro diferentes tamanhos das áreas de abrangência.....	42
Figura 4.5 – Comparativo entre os diferentes tempos de integração, com tempos de integração de 1 até 10 semanas para os 361 municípios selecionados nesta pesquisa, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros	43
Figura 4.6 – Erros médios das estimativas calculadas para os 361 municípios selecionados, através dos diferentes tempos de integração.....	44
Figura 4.7 – Comparativo entre os diferentes tempos de atraso (LAG), com valores de 0 semanas até 4 semanas, para os 361 municípios selecionados nesta pesquisa, no qual a linha vermelha representa a mediana dos erros e o	

sinal “+” na cor vermelha representa a média dos erros.	46
Figura 4.8 – Erros médios entre os valores estimados pelo método proposto nesta pesquisa e os dados de produtividade divulgados pelo IBGE para os 361 municípios da área de estudo, ano de 2012, com uso de cada um dos tempos de atraso (LAG) propostos.	46
Figura 4.9 – Comparativo entre os erros relativos dos municípios de cada grupo, separados pelo valor de área colhida. Com exceção do grupo 1, os grupos apresentaram tendência de superestimar o valor de produtividade. A linha vermelha representa a mediana e o sinal “+” vermelho representa a média dos valores.	49
Figura 4.10 – Erros percentuais médios das estimativas de produtividade da cana-de-açúcar para o ano de 2012, na semana do dia 84, por mesorregião geográfica. O erro apresentado é a diferença relativa entre os valores estimados e observados por mesorregião, considerando apenas os municípios utilizados neste trabalho.	50
Figura 4.11 – Estimativa de produtividade de cana-de-açúcar para o ano de 2012 do município de Altinópolis – SP por meio do método proposto.	53
Figura 4.12 – Células selecionadas como <i>proxies</i> da produtividade do município de altinópolis (pontos amarelos).	54
Figura 4.13 – Erros médios (%) entre os valores estimados e observados para os 361 municípios selecionados como área de estudo desta pesquisa.	55

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 2.1 – Produtos gerados para cada data de processamento com base em algumas datas de referência.	18
Tabela 3.1 – Valor absoluto e relativo da área colhida de cana-de-açúcar por grupo.	24
Tabela 3.2 – Configuração dos cenários analisados.....	34
Tabela 4.1 – Resultados com o valor p do teste t pareado para cada combinação de escalonamento e critério de seleção. Em negrito estão destacados os valores que apresentaram diferença estatística a 5% de significância.....	39
Tabela 4.2 – Resultados do teste t pareado comparando os diferentes tamanhos das áreas de abrangência. Em negrito estão destacados os valores que apresentaram diferença estatística a 5% de significância.....	41
Tabela 4.3 – Valor p do teste t pareado aplicado aos tempos de integração de 1 a 10. Os valores destacados em negrito foram os que apresentaram diferença estatística a 5% de significância.....	44
Tabela 4.4 – Valor p do teste t pareado aplicado aos LAGs de 0 a 4. Os valores destacados em negrito foram os que apresentaram diferença estatística a 5% de significância.....	47
Tabela 4.5 – Resultado das estimativas de produtividade da cana-de-açúcar por grupo, para o ano de 2012 utilizando os parâmetros que obtiveram os melhores resultados.	48
Tabela 4.6 – Relação dos resultados obtidos por mesorregião geográfica, incluindo os dados de erro médio (%) e erro médio (t/ha).....	51
Tabela 4.7 – Produtividade divulgada pelo IBGE por ano para os municípios de Irapuru e Pacaembu.....	51
Tabela 4.8 – Produtividade divulgada pelo IBGE, por ano, para alguns municípios da mesorregião de Ribeirão Preto.....	52

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO.....	1
1.1. Objetivo Geral	3
1.2. Objetivos Específicos.....	3
2 FUNDAMENTAÇÃO TEÓRICA.....	5
2.1. Modelos de estimativa de produtividade	5
2.1.1. Modelos agrometeorológicos	5
2.1.2. Modelos agrometeorológicos espectrais.....	6
2.2. Índices de vegetação	8
2.2.1. Correlação entre NDVI e produtividade agrícola.....	9
2.3. A cultura da cana-de-açúcar	10
2.4. Levantamento de produtividade a partir de dados IBGE.....	11
2.5. O sensor MODIS.....	12
2.5.1. Filtragem de séries temporais.....	14
2.6. Produto NDVI do sensor MODIS da universidade de BOKU	17
3 MATERIAIS E MÉTODOS.....	21
3.1. Visão Geral	21
3.2. Materiais	23
3.2.1. Área de estudo.....	23
3.2.2. Dados e recursos de processamento.....	24
3.3. Métodos	26
3.3.1. Método baseado em proxies.....	26
3.3.2. Automatização dos processamentos de obtenção dos proxies	27
3.3.3. Variáveis utilizadas	31
3.3.4. Avaliação dos parâmetros.....	34
3.3.5. Análise das estimativas de produtividade	35
4 RESULTADOS E DISCUSSÃO.....	37
4.1. Escalonamento e critério de seleção	37
4.2. Avaliação do tamanho da área de abrangência.....	40

4.3.	Análise do tempo de integração (T.I.)	43
4.4.	Avaliação do tempo de atraso (LAG)	45
4.5.	Estimativas agrupadas por valor de área colhida	48
4.6.	Estimativas por mesorregiões e por municípios.....	49
5	CONCLUSÃO	57
	REFERÊNCIAS BIBLIOGRÁFICAS	59
	APÊNDICE A – GRÁFICOS E TABELAS AUXILIARES PARA	
	INTERPRETAÇÃO DOS RESULTADOS	67

1 INTRODUÇÃO

A disponibilização de informações em tempo hábil de forma padronizada e contínua sobre as atividades agrícolas de uma região é essencial para o acompanhamento adequado das questões que envolvem a segurança alimentar, o escoamento de safra e a estimativa de preços. Baseado em informações precisas e especializadas, os tomadores de decisão são capazes de decidir de forma rápida e identificar geograficamente regiões com grandes variações na produtividade e na área plantada (ATZBERGER, 2013).

No Brasil, as estimativas agrícolas têm sido realizadas com o uso de métodos classificados como subjetivos, com base em entrevistas realizadas com agentes da cadeia produtiva, sendo os principais organismos oficiais encarregados da geração de estatística agrícolas o Instituto Brasileiro de Geografia e Estatística (IBGE) e a Companhia Nacional de Abastecimento (CONAB). Um dos problemas dos métodos utilizados por estes órgãos é o tempo necessário para a disponibilização das estimativas, que após a safra, podem demorar até dois anos para serem divulgadas. Parte desta demora pode ser atribuída à falta de automatização do método de coleta. Novos métodos devem ser investigados para diminuir o consumo de recursos humanos e financeiros em relação aos métodos não automáticos, além de fornecer estimativas agrícolas eficientes e rápidas.

Segundo Atzberger (2013) um sistema de monitoramento agrícola eficaz deve ser capaz de disponibilizar dados de maneira homogênea e intercambiável, em tempo hábil e com precisão estatística conhecida. Segundo Roughgarden et al. (1991), provavelmente apenas métodos que utilizem dados de sensoriamento remoto orbital, aliados a ferramentas de modelagem serão capazes de prover este tipo de informação em um tempo razoável, sobre grandes áreas, com detalhamento espacial suficiente e com custos razoáveis.

Com o intuito de aprimorar as estatísticas agrícolas do Brasil, o monitoramento da produtividade de culturas agrícolas, por meio de dados de sensoriamento

remoto, tem sido historicamente alvo de grande número de pesquisas. Por exemplo, modelos que se baseiam em dados agrometeorológicos, espectrais ou agrometeorológicos-espectrais foram utilizados com esta finalidade (RUDORFF, 1985; RUDORFF; BATISTA, 1990; SUGAWARA, 2002; LIU et al., 2010; NOGUEIRA, 2014).

Existe, ainda, um interesse crescente pela comunidade de sensoriamento remoto pelo uso de séries temporais visando à análise de alvos terrestres (WARDLOW; EGBERT, 2008; BROWN et al., 2013; SAKAMOTO et al., 2014). Segundo Kastens et al. (2005), a análise das condições ideais de crescimento da vegetação, por exemplo, têm se apresentado como melhor indicador do potencial de produtividade do que a própria cultura investigada.

Kastens et al. (2005) apresentaram um modelo em que são criadas máscaras de cultivo agrícola a partir da correlação entre o índice de vegetação NDVI e dados anuais de produtividade. Estas máscaras de cultivo foram então utilizadas no lugar de mapas de cultivo e inseridas em modelos de estimativa de produtividade que obtiveram resultados semelhantes ao encontrados com máscaras de cultivo. A principal vantagem desse método é que pode ser aplicado a qualquer tipo de cultivo e região geográfica, desde que existam os dados históricos de NDVI e de produtividade.

Mais recentemente, Mello et al. (2014) propuseram um método no qual, a partir da correlação entre uma série filtrada do produto de NDVI do sensor MODIS e dados de produtividade, foram extraídos valores de algumas células que serviram como base para a estimativa de produtividade da cana-de-açúcar, para cinco municípios do estado de São Paulo. Essa pesquisa mostrou que, mesmo quando não utilizadas máscaras de cultivo ou mapas das culturas, é possível obter boas estimativas de produtividade, no entanto a aplicação para uma área mais ampla ainda é necessária.

Com base nos trabalhos de Kastens et al. (2005) e Mello et al. (2014) a presente pesquisa buscou utilizar, de forma semelhante, a correlação entre dados de sensoriamento remoto e dados anuais de produtividade, a fim de

identificar células, no produto NDVI do sensor MODIS, que pudessem ser utilizados para estimar a produtividade da cultura da cana-de-açúcar. Dessa forma, os valores das células que apresentaram a melhor correlação entre o produto NDVI e a produtividade histórica, receberam a denominação de *proxies*.

Assim, essa pesquisa parte da hipótese que produtos NDVI oriundos do sensor MODIS, devidamente processados, podem ser utilizados como representantes da produtividade agrícola da cana-de-açúcar a partir da análise de correlação entre os dados de produtividade e valores de NDVI.

1.1. Objetivo Geral

O objetivo do presente trabalho foi avaliar a eficiência do método de estimativa de produtividade, da cana-de-açúcar, baseado na associação entre a série temporal do produto de NDVI do sensor MODIS e dados de produtividade, para os anos de 2003 a 2012, para os municípios do estado de São Paulo.

1.2. Objetivos Específicos

Este trabalho teve como objetivos específicos:

- a) Avaliar métodos de escalonamento que serão aplicados sobre os valores de produtividade e os valores do índice de vegetação NDVI;
- b) Definir critérios para seleção dos melhores *pixels* para a estimativa de produtividade;
- c) Avaliar diferentes tamanhos das regiões de abrangência de cada município;
- d) Analisar os efeitos decorrentes do uso de produtos NDVI do sensor MODIS de semanas anteriores à semana investigada;

- e) Avaliar a possibilidade de integração entre produtos NDVI do sensor MODIS de diferentes datas;
- f) Validar a metodologia comparando-se os resultados obtidos com os resultados divulgados oficialmente. Os resultados serão comparados por grupos de municípios, de acordo com o valor de área colhida, por mesorregiões geográficas e por municípios.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta uma revisão sobre modelos de estimativa de produtividade, através de dados de sensoriamento remoto. Além disso, são apresentadas informações sobre a cana-de-açúcar e seu cultivo, sobre o método de estimativa oficial da produtividade desta cultura, e sobre o índice de vegetação NDVI. Em sequência, é apresentada uma descrição do sensor MODIS, a importância da filtragem dos dados de séries temporais, e as características do produto NDVI, do sensor MODIS, processado e disponibilizado pelo *Institute of Surveying, Remote Sensing and Land Information* (IVFL), da *University of Natural Resources and Life Sciences* (BOKU).

2.1. Modelos de estimativa de produtividade

A utilização de modelos que estimam a produtividade de safras agrícolas, ganhou força com o advento de métodos que integram dados de sistemas orbitais. Esses sistemas tornam possível o monitoramento contínuo de vastas regiões, com pouca alocação de recursos financeiros. No entanto, até o momento, poucos modelos foram capazes de apresentar boa correlação entre os dados coletados em campo e as estimativas (DOORENBOS; KASSAM, 1979; RUDORFF, 1985; RUDORFF; BATISTA, 1990; SUGAWARA, 2002).

2.1.1. Modelos agrometeorológicos

Diversos trabalhos de estimativa de produtividade agrícola, por meio de modelos agrometeorológicos, foram desenvolvidos com uso do modelo proposto por Doorenbos e Kassam (1979). Este método baseia-se em um fator de sensibilidade a deficiência hídrica, específico para cada cultura agrícola e para cada estágio de desenvolvimento, sendo este fator uma penalização de forma direta em relação ao nível de deficiência hídrica ao qual a planta é submetida.

Outros fatores importantes para a produtividade agrícola, como a relação entre a disponibilidade de água e os solos, processos biológicos e fatores físicos e fisiológicos são complexos de serem modelados, e em alguns casos, não foram considerados no modelo. Alguns autores relatam que os resultados insatisfatórios para a modelagem da produtividade, estão associados a estes fatores, quando não houve qualquer alteração nas medidas de evapotranspiração que justificasse a penalização da produtividade (PEDRO JUNIOR et al., 1984; BERLATO, 1987).

Assad et al. (2007) propuseram um modelo agrometeorológico que integrou dados do sistema SPSoja-Agritempo. Neste modelo foram acrescentados coeficientes empíricos de ajuste, específicos para cada região do país, de forma a adaptar os diferentes cenários regionais, em relação às variedades de cultivares e as particularidades dos sistemas de produção empregados. Este trabalho apresentou boa correlação entre produção estimada e a oficial, sendo que para a Região Sudeste a subestimativa foi inferior a 0,5%. As maiores discrepâncias ocorreram para as safras 2000/2001 e 2001/2002 para São Paulo, e 2001/2002 para Minas Gerais no qual os valores de produtividade superestimam em 9,25% e 7,17% em São Paulo e em 14% para Minas Gerais em relação aos valores divulgados pela Conab, para as respectivas safras.

2.1.2. Modelos agrometeorológicos espectrais

A disponibilidade de dados espacialmente distribuídos e com alta repetitividade de coleta, permitiu o desenvolvimento de modelos que explorassem essas características nos sensores orbitais disponíveis. A integração entre modelos agrometeorológicos, derivados do modelo proposto por Doorenbos e Kassam (1979) e dados espectrais, complementam através do sensoriamento remoto, as informações que não estão disponíveis para o modelo agrometeorológico. Segundo Rudorff e Batista (1990), o componente espectral destes modelos deve representar o efeito coletivo de variáveis como as condições de radiação solar incidente, temperatura, umidade do ar e de disponibilidade hídrica, além

de outras não consideradas, como as práticas de manejo, cultivares e profundidade do sistema radicular.

Recentemente, Morel et al. (2014) analisaram quatro modelos de estimativa de produtividade para a cana-de-açúcar, e concluíram que o modelo empírico, que se baseia no valor acumulado do NDVI através de integrais entre duas colheitas sucessivas, foi o que apresentou os melhores resultados, com RMSE de 10 t/ha. Esse modelo não utilizou máscaras de cultivo, nem restrições relacionadas às etapas de desenvolvimento do cultivo.

Para previsão de safras agrícolas, a abordagem por meio de máscaras para cultivos específicos deveria ser a ideal. Dessa forma, apenas as informações dos índices que estão relacionadas com determinada cultura seriam consideradas. No entanto, quando não são aplicadas, nestas máscaras, restrições em relação ao estágio de desenvolvimento do cultivo, nada impede que baixas respostas dos índices estejam relacionadas à fase de desenvolvimento do cultivo e não às condições fitossanitárias da cultura (BECKER-RESHEF et al., 2010).

Segundo Kastens et al. (2005), muitos autores vêm buscando formas de criar máscaras de cultivo para as imagens de sensoriamento remoto com o intuito de identificar regiões com qualquer tipo de cultivo ou regiões com algum cultivo específico. Esses esforços consideram que através destas máscaras seja possível melhorar os resultados dos modelos de previsão. No entanto, realizar o mapeamento de culturas, principalmente através de imagens de resolução espacial moderada (30-50 metros), é bastante limitado, principalmente devido à ocorrência de nuvens (DEFRIES; TOWNSHEND, 1999; CARRÃO et al., 2008; SUGAWARA et al., 2008). Geralmente, a criação destas máscaras e a definição de restrições, implicam em várias horas de trabalho técnico, pois os métodos de classificação automática ainda não atingiram o grau necessário de confiabilidade.

O trabalho de Mello et al. (2014) buscou estimar a produtividade agrícola da cana-de-açúcar, para cinco municípios do estado de São Paulo, por meio da

correlação entre dados de sensoriamento remoto e a série histórica de produtividade. Algumas células de uma imagem de NDVI foram selecionadas a partir da correlação com os dados de produtividade e o valor de NDVI destas células foi escalonado para valores de produtividade. Uma das vantagens deste método é que máscaras de cultivo, referente à cultura investigada, não são necessárias. Além disso, a seleção de uma parcela de células de uma imagem de sensoriamento remoto, faz com que os dados com baixa correlação com a variável estudada, neste caso a produtividade, sejam desprezados, eliminando-se principalmente as células afetadas pela ocorrência de nuvens.

2.2. Índices de vegetação

Os índices de vegetação (IVs) podem ser definidos como o resultado de transformações entre bandas espectrais e apresentam-se sob a forma de equações matemáticas que utilizam duas ou mais bandas. Muitos estudos utilizaram os IVs para discriminar as características biofísicas do alvo observado, tal como o conteúdo de água, área foliar ou a presença de doenças (ROUSE et al., 1973; PENUELAS et al., 1993; GAO, 1996; MAHLEIN et al., 2013).

Em geral, os IVs são elaborados utilizando as bandas do vermelho na região do visível (*RED*) e do infravermelho próximo (Near Infra-Red - NIR). A região do RED apresenta correlação com a clorofila presente na folha, enquanto o NIR está relacionado com a estrutura foliar (GOEL, 1988). Por essa razão apresentaram boa correlação em estudos que analisaram os parâmetros biofísicos da vegetação (TEILLET et al., 1997; RIZZI et al., 2009; BROWN et al., 2013).

No entanto, devido à resolução espacial dos sensores orbitais, as células de uma imagem de IV tendem a representar uma região da superfície da Terra, que contém mais de um tipo de cobertura (por exemplo, vegetação, sombra, solo, etc). Dessa forma, alguns autores buscaram desenvolver índices que minimizassem os efeitos decorrentes do solo, da sombra, e de outros

componentes, utilizando para isto outras bandas espectrais ou fatores numéricos para compensar os efeitos desses componentes (HUETE, 1988; BARET et al., 1989; HUETE et al., 2002).

Apesar disso, o NDVI (equação 2.1), proposto inicialmente por ROUSE et al. (1973), que utiliza apenas as bandas do NIR e RED, apresentou melhores resultados quando comparados a outros índices mais complexos. O NDVI, por ser um índice normalizado, possui a característica de minimizar os efeitos atmosféricos, apesar de apresentar saturação quando o índice de área foliar (IAF) está elevado e ser sensível para ruídos aditivos (HUETE et al., 2002; ASNER et al., 2003).

O trabalho de Li et al. (2010) mostrou que o produto de NDVI do sensor MODIS apresentou, para uma região da província de Hebei, na China, melhor correlação com dados de campo da cobertura vegetal quando comparado com o EVI do mesmo sensor, utilizando-se para isso uma composição de 32 dias obtida a partir da composição original de 16 dias.

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad 2.1$$

O trabalho realizado por Wardlow e Egbert (2008) mostrou que a série temporal de 15 datas do produto de NDVI do sensor MODIS é uma opção viável para o mapeamento regional de culturas de verão, com precisão relativamente alta (90%) na classificação não supervisionada das classes de cultivo agrícola/não cultivo agrícola e de 84% na classificação supervisionada de áreas cultivadas no estado de Kansas, nos Estados Unidos.

2.2.1. Correlação entre NDVI e produtividade agrícola

Nos últimos 30 anos, muitos pesquisadores utilizaram técnicas de sensoriamento remoto aplicadas às estimativas de produtividade das mais variadas culturas. Wiegand et al. (1979) e Tucker et al. (1980) primeiramente identificaram uma relação entre o NDVI e a produtividade agrícola do trigo, utilizando-se dados de campos experimentais e medidas de campo por meio de

um espectrorradiômetro. Rudorff e Batista (1990) utilizaram os valores de NDVI, como entrada em um modelo agrometeorológico, que conseguiu explicar aproximadamente 70% da variação do trigo para a safra de 1986 no Brasil.

Estes estudos, demonstrando a relação entre NDVI e produtividade agrícola, fundamentam até hoje os trabalhos que utilizam técnicas de sensoriamento remoto para estimativas de produtividade agrícola. Rasmussen (1992) utilizou o valor acumulado do produto de NDVI, do sensor *Advanced Very High-Resolution Radiometer* (AVHRR), e técnicas de regressão estatística para encontrar fortes correlações entre o NDVI acumulado e a produtividade do milho, mas obteve bons resultados apenas para os estádios reprodutivos, sendo que a correlação ruim para os outros estádios, foi atribuída à qualidade dos dados.

Picoli et al. (2009) utilizaram o produto de NDVI do sensor MODIS para estimar a produtividade da cana-de-açúcar, por talhão, na região Centro-Norte do estado de São Paulo. Apesar dos resultados insatisfatórios (o modelo explicou apenas 31% da variabilidade da produtividade e erro quadrático médio de 20,9 t/ha) quando foi estimada a produtividade sem os dados MODIS, o modelo não explicou qualquer variação na produtividade.

Independentemente de suas limitações, o NDVI tem se revelado particularmente favorável ao monitoramento global da vegetação. Os trabalhos apresentados anteriormente mostraram relações entre a produtividade de culturas agrícolas e o NDVI. No entanto, esta relação depende tanto da quantidade de dados quanto da qualidade destes.

2.3.A cultura da cana-de-açúcar

Levando em consideração que às dimensões das áreas plantadas com cana-de-açúcar no estado de São Paulo possuem, em média 10500 ha, o uso de dados e técnicas de sensoriamento remoto, tem potencial favorável para a análise desta cultura. Neste caso, destaca-se a importância de se identificar o calendário agrícola da cultura a ser avaliada, uma vez que é através dele que

se definem as melhores épocas do ano para a aquisição de dados de sensoriamento remoto (SANCHES et al., 2005).

Para a cultura da cana-de-açúcar, há dois principais ciclos distintos de cultivo, sendo o primeiro com duração média de 12 meses (cana de ano) e o segundo com duração média entre 14 e 21 meses (cana de ano-e-meio). A cana-de-açúcar é uma cultura semiperene e seu plantio é feito normalmente no início da estação chuvosa (setembro a novembro) para a cana de ano-e-meio e de janeiro a abril para a cana de ano (SUGAWARA, 2010).

Durante a estação chuvosa o crescimento vegetativo é mais intenso e por essa razão, a demanda por água, radiação e temperatura, é também alta. A partir de abril ocorre o acúmulo de sacarose nos colmos, que é marcado pelo crescimento vegetativo mais discreto. É neste período que ocorre a colheita, chegando até o mês de dezembro no estado de São Paulo (RUDORFF; SUGAWARA, 2007). Por se tratar de um cultivo semiperene, a aquisição de imagens ópticas livres de nuvens é mais fácil quando comparada a culturas de ciclo curto como é o caso da soja e do milho (RUDORFF et al., 2010).

2.4. Levantamento de produtividade a partir de dados IBGE

No Brasil, estatísticas oficiais de produtividade agrícola são realizadas e disponibilizadas periodicamente pelo IBGE e pela CONAB. Para os objetivos deste trabalho, foram utilizados dados oriundos da Pesquisa Agrícola Municipal (PAM) divulgada anualmente pelo IBGE (IBGE, 2013). Este levantamento é realizado por meio do preenchimento de um questionário para cada município, no qual o técnico do IBGE é responsável por entrar em contato com os membros da cadeia produtiva de determinada cultura, como técnicos agrícolas, empresas fornecedoras de insumos, produtores expressivos da região investigada, secretários municipais de agricultura, representantes de órgãos de fomento entre outros. A partir das informações repassadas por estes agentes, o técnico do IBGE é responsável por analisar as informações obtidas e informar o número estimado final ao supervisor da área (IBGE, 2002).

Neste levantamento são obtidas informações sobre 36 tipos de lavoura permanente, além de 32 tipos de lavoura temporária. As informações destes 68 tipos de lavoura são coletadas para todos os municípios do país, o que faz desta pesquisa, a mais completa do setor. Além disso, devido à sua característica de coleta, informações sobre p.e., o suporte creditício e financiamentos concedidos, comercialização, industrialização e demanda de insumos tecnológicos (sementes fiscalizadas, corretivos, fertilizantes, etc.) também são obtidas junto à cadeia de informantes e são consideradas no cômputo das estimativas finais. No entanto, também devido às características deste método de coleta, não é incomum que os dados sejam divulgados em até dois anos após a coleta desta informação, o que pode não ser tempo hábil para análises que demandam maior dinâmica, como escoamento e armazenagem.

2.5.O sensor MODIS

A *National Aeronautics and Space Administration* (NASA), por meio do sensor MODIS, a bordo dos satélites TERRA e AQUA, disponibiliza gratuitamente uma série de produtos para observação da Terra, monitoramento biofísico e do ambiente terrestre (TOWNSHEND; JUSTICE, 2002). Dentre as características desse sensor está a alta resolução temporal (quase diária), que é adequada para estudos que envolvem a dinâmica de cultivos agrícolas. Além disso, a resolução temporal permite aumentar a possibilidade de obtenção de imagens livres de nuvens, o que torna o sensor MODIS bastante adequado quando se deseja obter imagens com qualidade ao longo do tempo. (BERNARDES et al., 2011).

Apesar disso, os dados produzidos por sensores orbitais, como o MODIS, estão sujeitos a efeitos atmosféricos, variações da geometria de aquisição e questões do próprio processamento, que podem afetar o sinal originalmente detectado, fazendo aparecer valores anômalos em uma série de dados. (ATZBERGER; EILERS, 2011). Para parte destes valores anômalos é dada a denominação de ruídos. Tratando-se de séries temporais, os ruídos tornam-se mais perceptíveis quando o fenômeno observado possui um ciclo de repetição.

Por essa razão, produtos de composições de 16 dias foram criados para minimizar os ruídos provenientes do sensor, da presença de nuvens ou efeitos do ângulo de visada. Estes produtos já possuem correção geométrica, radiométrica e atmosférica (VERMOTE et al., 2002). Os produtos MOD13Q1 são produtos de composição de 16 dias de IVs, com resolução espacial de 250 m. A partir dos dados diários obtidos pelo sensor MODIS, são aplicados filtros. Estes filtros são baseados na qualidade, cobertura por nuvens e geometria de visada (HUETE et al., 2002). Apenas células de alta qualidade e livres de nuvens são selecionados inicialmente para a composição. A partir disso, a composição é realizada seguindo o diagrama apresentado na Figura 2.1.

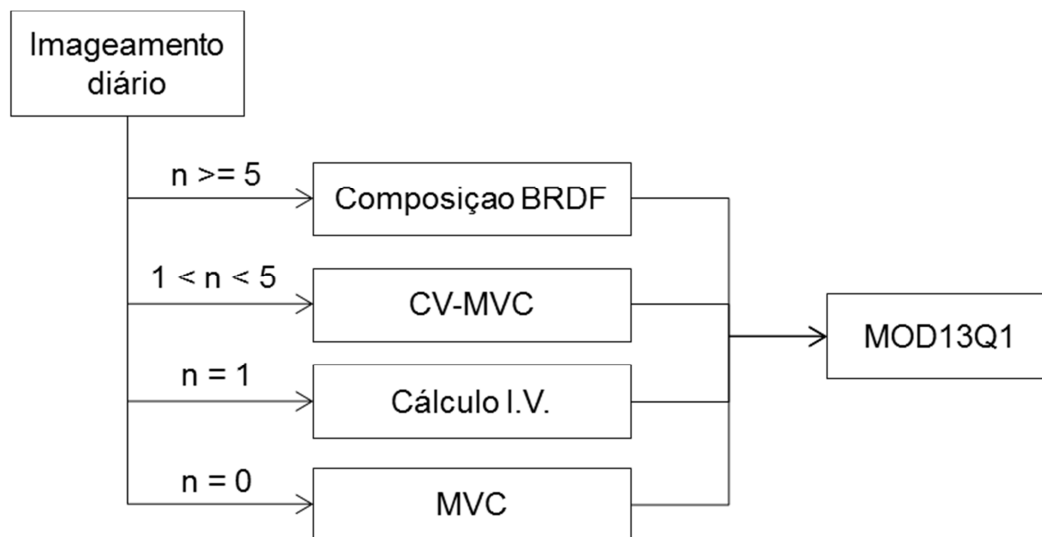


Figura 2.1 – Diagrama da composição do produto MOD13Q1, 'n' se refere ao número de observações com qualidade.
Fonte: adaptado de Huete et al. (2002).

A partir dos dados de imageamento diário, são computadas quantas observações com qualidade foram obtidas. Caso um mínimo de cinco observações sem obtidas, o algoritmo principal é utilizado, no qual é realizada a composição baseada na Função de Distribuição da Reflectância Bidirecional (BRDF). Neste caso, os valores das células com ângulo de visada não perpendicular é computado para a perpendicular, interpolados e por fim, o IV é calculado.

Quando são obtidas entre duas e quatro observações com qualidade, o algoritmo *Constrained View – Maximum Value Composite* (CV-MVC) é aplicado. Neste caso, as células com ângulo de visada menor que 45° (CV) são utilizados e dentre estas células o maior valor de NDVI será selecionado (MVC). Caso exista apenas uma observação de qualidade, então o NDVI é diretamente calculado e por fim, caso nenhuma observação de qualidade seja obtida no período, o NDVI é calculado para todas as observações do período e o maior valor de NDVI será utilizado na composição.

Mesmo após a seleção das células de melhor qualidade para a composição de 16 dias, ruídos ainda são observados (CARREIRAS et al., 2003). Por isso, quando utilizados para análise de séries temporais, o primeiro passo inevitável é a filtragem dos dados (ATKINSON et al., 2012), para que uma série temporal suavizada seja utilizada.

2.5.1. Filtragem de séries temporais

Entre os ruídos encontrados nos produtos de sensoriamento remoto, Hird e McDermid (2009) destacam dois tipos que ocorrem em séries temporais derivadas dos produtos MOD13 e que podem ser minimizados, sendo eles: ruídos minimizáveis por técnicas de preenchimento de lacunas (*gap filling*); e ruídos minimizáveis por técnicas de suavização dos dados (*data smoothing*).

Os ruídos que podem ser corrigidos pelo método de preenchimento de lacunas são resultantes de períodos prolongados ou permanentes de cobertura de nuvens. Em série temporal de NDVI, a presença de nuvens causa quedas abruptas dos valores da série. Enquanto que os ruídos que são corrigidos por meio de algoritmos de suavização são originários dos demais efeitos que podem afetar a obtenção dos dados de sensoriamento remoto orbital, sendo o ângulo de visada o predominante no sensor MODIS (CARREIRAS et al., 2003).

Na maioria dos estudos que utilizam séries temporais de dados de sensores orbitais, a filtragem dos dados se faz necessária devido aos ruídos encontrados na série histórica, sejam esses dados de IV ou não. Após a filtragem destas

séries de dados a estimação de parâmetros fenológicos como o verdor da vegetação, sazonalidade e senescência se torna mais precisa (ATKINSON et al., 2012). Atkinson et al. (2012) analisaram uma série de técnicas utilizadas para filtrar e suavizar séries temporais com o objetivo de estimar parâmetros fenológicos e, após avaliar quatro delas, concluíram que a filtragem através do algoritmo de suavização Whittaker apresentou os melhores resultados em florestas tropicais da Índia. Este algoritmo foi comparado com o ajuste local, função gaussiana assimétrica e função dupla logística para estimar parâmetros fenológicos.

O algoritmo de suavização Whittaker se baseia na penalização dos mínimos quadrados (Figura 2.2), em que séries discretas são ajustadas a dados discretos e a rugosidade da curva suavizada é penalizada (ATZBERGER; EILERS, 2011). Entretanto, o ruído de séries temporais de IV tende, geralmente, a resultados subestimados em decorrência da presença de nuvens e aerossóis, e por esta razão, recomenda-se o uso do ajuste ao invólucro superior (upper envelope). Este ajuste consiste em identificar pontos da série temporal, nos quais os valores observados estão abaixo dos valores filtrados, nestes pontos os valores são substituídos pelos valores filtrados.

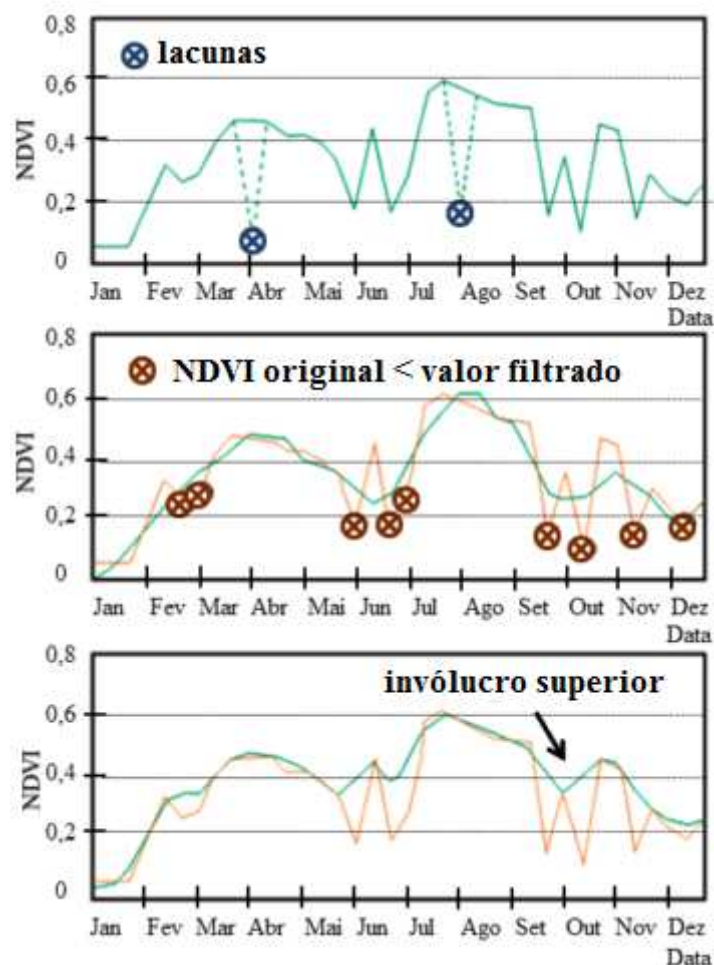


Figura 2.2 – Técnica de filtragem do algoritmo de suavização Whittaker. Primeiramente é aplicada uma interação mais suave aos dados, para fechar as lacunas resultantes pela presença de nuvens, conhecido como ajuste ao invólucro superior. Em seguida, são identificados os pontos em que o IV original é menor que o valor filtrado. Para estes pontos, os valores originais são substituídos pelos valores filtrados e o processo de filtragem é repetido novamente. Depois de duas a três iterações, uma curva suavizada é obtida.

Fonte: adaptado de Chen et al. (2004) e Atzberger e Eilers (2011).

Pela descrição dos procedimentos utilizados pelo algoritmo de suavização Whittaker (Figura 2.2), este algoritmo é bastante adequado à utilização em séries temporais do produto NDVI, uma vez que corrige ruídos causados por nuvens e em seguida aplica um filtro de suavização, que minimiza os demais ruídos da série.

2.6. Produto NDVI do sensor MODIS da universidade de BOKU

O produto NDVI do sensor MODIS utilizado neste trabalho foi disponibilizado pela universidade de BOKU. Apesar de este produto ter como base o mesmo método de composição do produto MOD13Q1 e MYD13Q1, o produto gerado na universidade de BOKU é interpolado e processado em uma base semanal regular, correspondendo às segundas-feiras (KLISCH, 2014). Segundo Vuolo e Mattiuzzi (2012), no processo de filtragem é utilizado o algoritmo de suavização Whittaker (EILERS, 2003). Este algoritmo é aplicado sobre os valores das últimas 25 semanas. Depois de filtrado, é disponibilizado em formato matricial na extensão *Tagged Image File Format for Georeferenced raster imagery* (GeoTIFF), em cenas de $1^\circ \times 1^\circ$, na projeção WGS84 (EPSG:4326) (VUOLO; MATTIUZZI, 2012). Este dado, disponibilizado pela universidade de BOKU, será denominado daqui em diante por *minitile*.

Considerando que a filtragem destes dados é realizada semanalmente e que o produto da semana corrente será filtrado novamente na próxima semana, diferentes produtos foram criados para que cada uma destas filtrações fosse armazenada e identificada, assim, o histórico das 5 primeiras filtrações é armazenado e divulgado em 5 produtos diferentes, disponibilizados semanalmente e assim distribuídos, sendo “s” referente à semana corrente (“atual”):

- Near real-time (N): produto que corresponde ao valor filtrado de NDVI para a semana corrente (s). É o dado da extremidade da série temporal, filtrado pela primeira vez com base na série de 25 semanas anteriores;
- A: Produto que corresponde ao valor filtrado de NDVI para a semana anterior à semana corrente (s-1). É o penúltimo dado da extremidade da série temporal e é filtrado com base nas 24 semanas anteriores e uma semana posterior;

- B: Produto que corresponde ao valor filtrado de NDVI para duas semanas anteriores a semana corrente (s-2). É o antepenúltimo dado da extremidade da série temporal e é filtrado com base nas 23 semanas anteriores e duas semanas posteriores;
- C: Produto que corresponde ao valor filtrado de NDVI para três semanas anteriores da semana corrente (s-3). Segue o mesmo fluxo dos produtos anteriores, mas é filtrado com base nas 22 semanas anteriores e três semanas posteriores;
- D: Produto que corresponde ao valor filtrado de NDVI para quatro semanas anteriores da semana corrente (s-4). Segue o mesmo fluxo dos produtos anteriores, mas é filtrado com base nas 21 semanas anteriores e quatro semanas posteriores.

Na Tabela 2.1 são apresentados os cinco produtos disponibilizados pela universidade de BOKU. São apresentadas também algumas datas de referência e os respectivos produtos que são criados em cada data de processamento.

Tabela 2.1 – Produtos gerados para cada data de processamento com base em algumas datas de referência.

Data referência	Data processamento						
	18/05/15	25/05/15	01/06/15	08/06/15	15/06/15	22/06/15	29/06/15
18/05/15	N	A	B	C	D1	D2	D3
25/05/15		N	A	B	C	D1	D2
01/06/15			N	A	B	C	D1
08/06/15				N	A	B	C
15/06/15					N	A	B
22/06/15						N	A
29/06/15							N

Para entender melhor o processamento de cada um destes produtos, a semana do dia 18/05/2015 será exemplificada. Os dados obtidos pelo sensor MODIS nos sete dias anteriores ao dia 18/05/2015 serão processados na data de referência do dia 18/05/2015, formando uma única imagem composição

destes sete dias, será então filtrado pelo algoritmo Whittaker, receberá a nomenclatura “N” e será armazenado. Esta mesma imagem composição será reprocessada no dia 25/05/2015, será filtrada e o produto receberá a nomenclatura “A” e será armazenado. Este processo será feito até que o produto com a nomenclatura “D” seja gerado, quando isso ocorrer, a filtragem dos dados da data de referência do dia 18/05/2015 estará encerrada. Como a filtragem é feita semanalmente, o produto A de uma determinada semana será o produto B da semana seguinte (Figura 2.3).

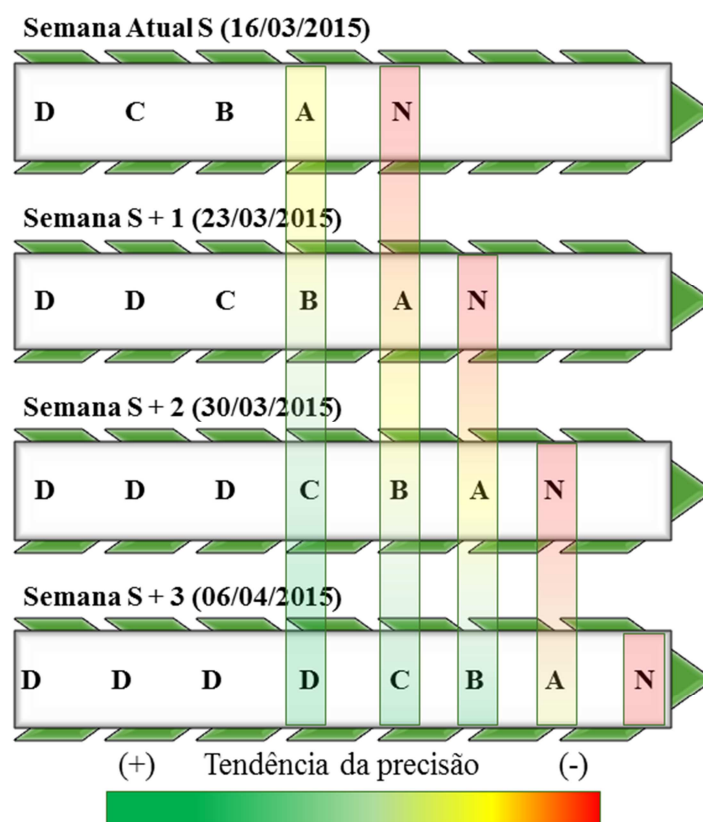


Figura 2.3 – Esquematização da filtragem dos dados de BOKU. Os produtos mais distantes da extremidade da série temporal tendem a ser mais precisos do que os mais próximos das extremidades. Assim, o produto D tende a ser mais preciso enquanto o produto N tende a ser menos preciso.

Dado que a filtragem tende a obter melhores resultados para datas que não correspondam às extremidades da série temporal filtrada, por exemplo, o valor do produto B tende a ser mais preciso que o valor do produto A. Da mesma forma, C da semana corrente passará a ser D na próxima semana, só que com uma melhor filtragem sobre os valores do NDVI. Ou seja, os dados sofrem alterações ao longo do tempo e a divisão em cinco produtos foi idealizada de

forma a manter o histórico dos efeitos da filtragem. Por fim, a filtragem para uma determinada data é encerrada quando se atinge o produto do tipo D.

Nesta pesquisa, não só o valor individual de cada produto, como também o valor médio das células de produtos diferentes foram utilizados. Dessa forma, o termo Tempo de Integração (T.I.) é utilizado para simbolizar quantos produtos foram utilizados no compute de cada valor médio. Por exemplo, quando é utilizado o T.I. 1, apenas o valor de um dos produtos é utilizado e o produto que será utilizado é definido pelo LAG. Quando é utilizado o T.I. 2, dois diferentes produtos serão somados e divididos, obtendo-se o valor médio de cada uma das células. Se for definido um LAG 0 e T.I. 2, significa que a média dos produtos N e A será utilizada nos processamentos.

3 MATERIAIS E MÉTODOS

3.1. Visão Geral

O fluxograma que resume as principais etapas desenvolvidas no presente trabalho é apresentado na Figura 3.1. Os dois principais materiais utilizados neste trabalho foram: 1) imagens com valores de NDVI obtidos a partir de dados do sensor MODIS e processados pela universidade de BOKU; 2) série histórica dos dados de produtividade do IBGE levantados através da pesquisa de Produção Agrícola Municipal (PAM). No entanto utilizaram-se também outros dados de fontes já existentes ou dados derivados destes dados iniciais, conforme descritos na seção 3.3.

A etapa de pré-processamento deste trabalho consistiu no cálculo da produtividade por município, por meio do valor de produção e de área colhida, divulgadas pela PAM (IBGE, 2013), além disso, envolveu a criação do arquivo contendo os centroides dos municípios da área de estudo, no formato *shape*, *datum* SIRGAS 2000 na projeção Cônica Equivalente de Albers, processado através do Sistema de Informação Geográfica (SIG) Qgis (QGIS, 2009).

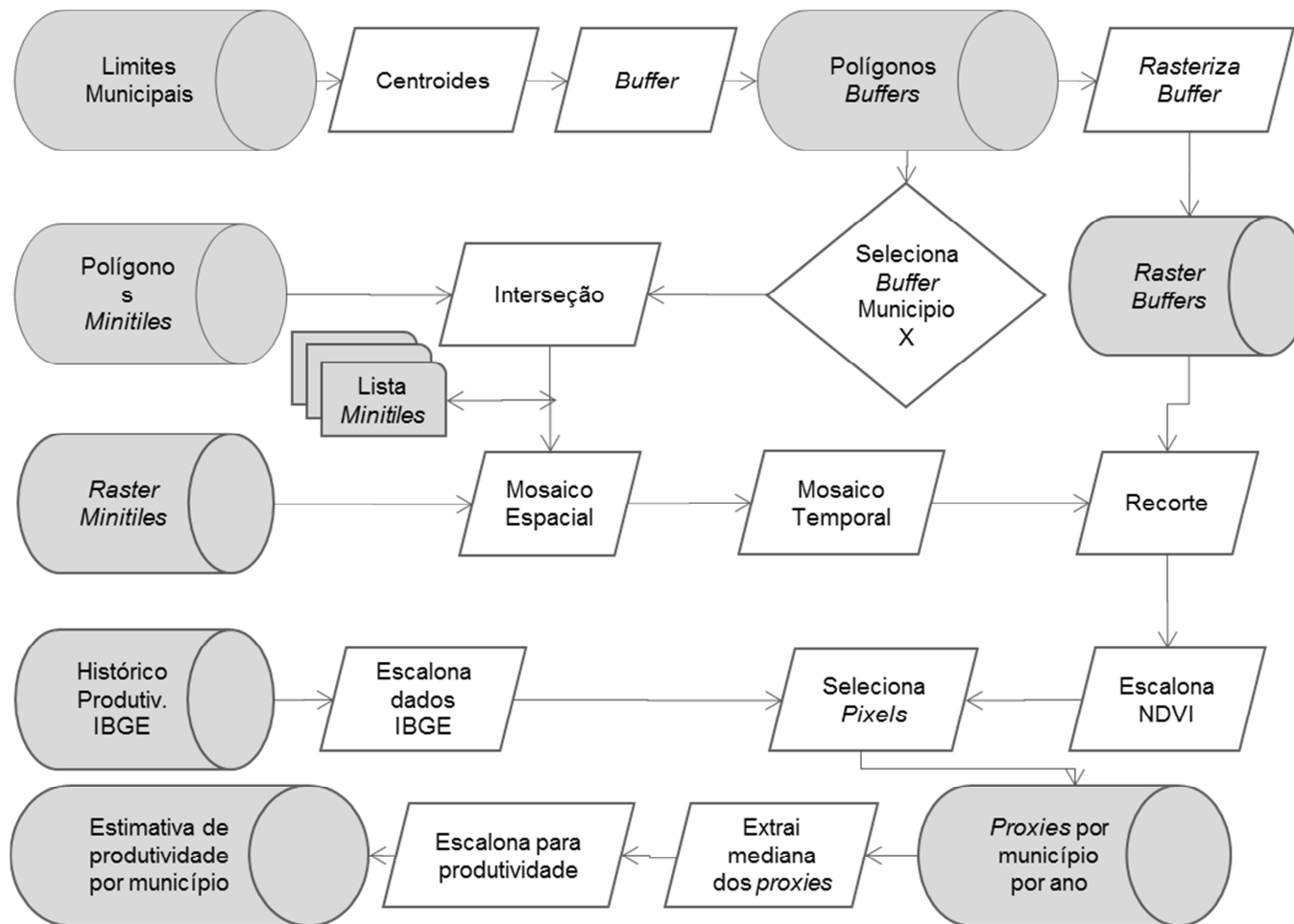


Figura 3.1 – Fluxograma do método, apresentando as principais etapas de processamentos. As figuras cilíndricas representam o armazenamento dos dados, tanto originais quanto processados enquanto os paralelogramos representam cada uma das etapas de processamento dos dados.

3.2. Materiais

3.2.1. Área de estudo

A área de estudo deste trabalho correspondeu a um conjunto de 361 municípios do estado de São Paulo (Figura 3.2). Este conjunto de municípios foi definido em razão da existência da série histórica de área colhida e quantidade produzida da cultura da cana-de-açúcar no período de 2003-2012, no banco de dados do IBGE, que é um dado necessário para este método. Os municípios que não possuíam os dados para algum ano da série histórica de 2003-2012, não foram considerados no trabalho.

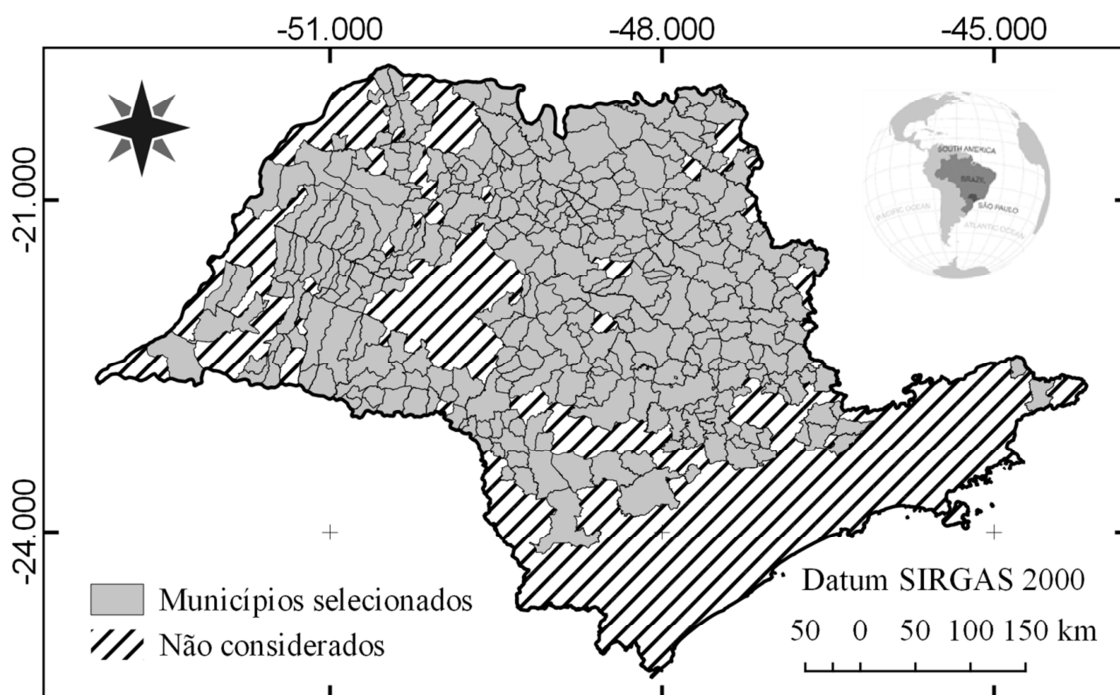


Figura 3.2 – Área de estudo com destaque ao conjunto de 361 municípios analisados neste trabalho, apresentando a localização destes, no estado de São Paulo.

O agrupamento dos municípios, em cinco conjuntos, permitir análises de acordo com diferentes faixas de quantidade de área colhida de cana-de-açúcar. O limiar de área de cada grupo foi obtido por meio do algoritmo de k-médias. Assim, é possível avaliar o desempenho do método tanto para os municípios de menor quanto de maior produção de cana-de-açúcar, de forma

independente. As características de cada um dos grupos são apresentadas na Tabela 3.1.

Tabela 3.1 – Valor absoluto e relativo da área colhida de cana-de-açúcar por grupo.

Grupo	Quantidade de Municípios	Área total (t/Ha)	Área total (%)	Área mínima (Ha)	Área máxima (Ha)
1	52	42.904	0,83%	-	2.166,21
2	114	585.719	11,37%	2.166,21	8.029,56
3	109	1.341.445	26,05%	8.029,56	18.342,40
4	70	1.845.395	35,83%	18.342,40	37.872,15
5	16	796.373	15,46%	37.872,00	-

Fonte: IBGE (2012).

Os resultados também foram agrupados por mesorregiões, de acordo com a divisão geográfica adotada pelo IBGE (IBGE, 2014). Dessa forma, foi possível avaliar se ocorrem variações, entre valores observados e estimados, relacionadas com a dispersão geográfica.

3.2.2. Dados e recursos de processamento

Os dados de produtividade da cana-de-açúcar para o estado de São Paulo coletados por meio da PAM, foram obtidos no site do Sistema IBGE de Recuperação Automática (SIDRA), tabela 1612, para as safras de 2003 a 2012. Foram obtidos também, através do IBGE, os limites municipais oficiais em formato *shape* na projeção SIRGAS 2000 (EPSG: 4674).

A série temporal MODIS foi disponibilizada pela Universidade de BOKU, localizado em Viena, na Áustria. Este dado é disponibilizado semanalmente, em formato matricial e extensão GeoTIFF e em 5 tipos diferentes de produtos, cada um relativo à uma data de filtragem. Foi utilizado um total de 72 *minitiles* que formam a cena da área de estudo. Para cada *minitile* dessa relação, foi obtida a série histórica semanal de 2003-2012. Foram obtidos também os limites de cada *minitile* por meio de um arquivo vetorial, na projeção WGS 84. A relação dos *minitiles* que formam a área de estudo é apresentada na Figura 3.3.

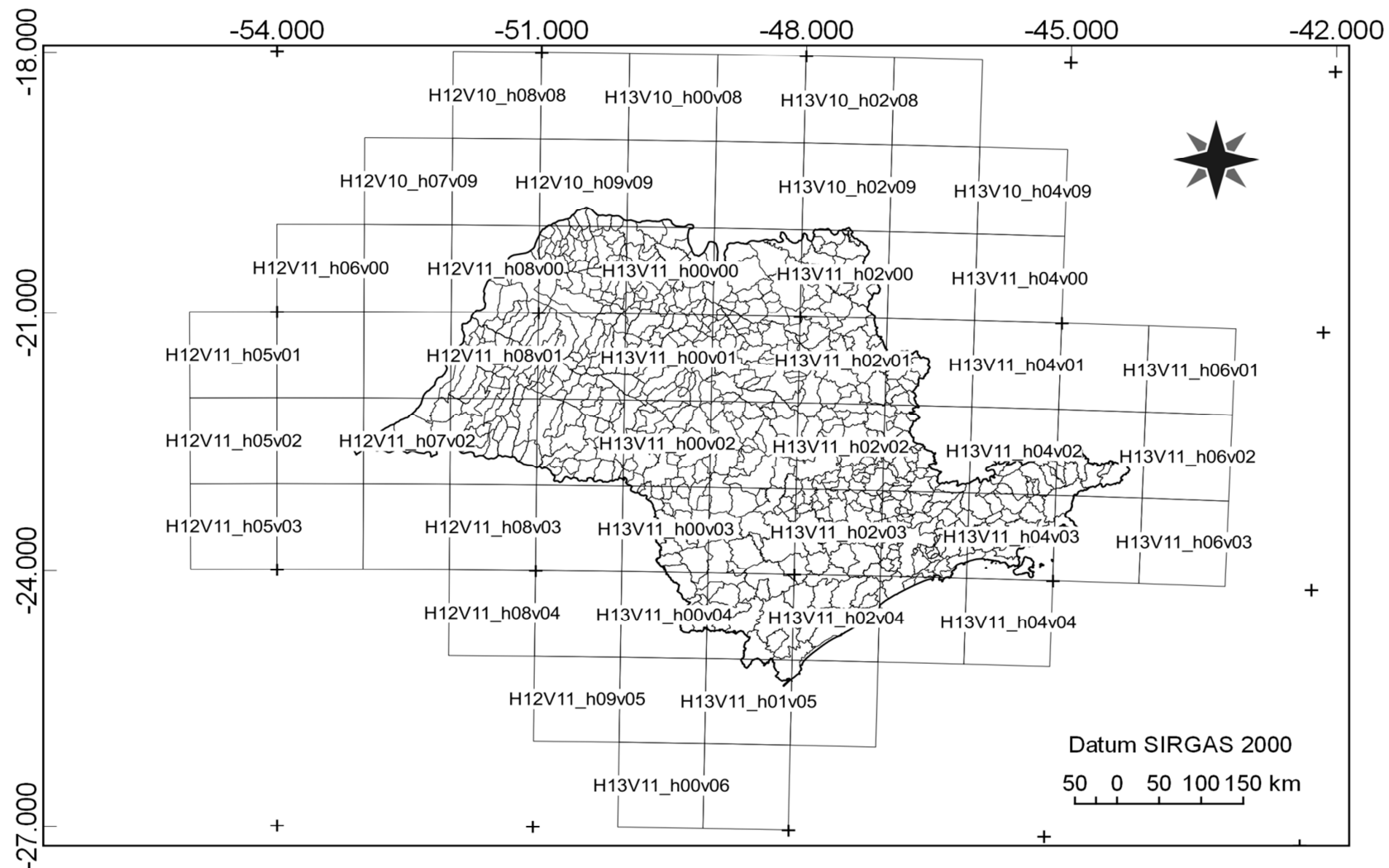


Figura 3.3 – Limites municipais do estado de São Paulo e localização dos *minitiles* MODIS fornecidos pela universidade de BOKU/IVFL.

A partir do dado vetorial de limites municipais foi gerado o arquivo vetorial no formato *shape* com os centroides dos limites municipais do estado de São Paulo utilizando o software Qgis. Os outros arquivos foram gerados por meio de scripts na linguagem de programação Python, na versão 2.7.6 (ROSSUM; DRAKE, 2010). Além das funções nativas da linguagem, foram utilizados os pacotes Scipy (OLIPHANT, 2007), Numpy (VAN DER WALT et al., 2011) e Pandas (MCKINNEY, 2010) para o processamento dos dados tabulares, organização, manipulação e cálculos estatísticos, Matplotlib (HUNTER, 2007) para gerar os gráficos e Gdal (GDAL DEVELOPMENT TEAM, 2015) e Shapely (SEAN GILLIES, 2013) para o processamento de dados vetoriais.

3.3. Métodos

3.3.1. Método baseado em proxies

Esta pesquisa se baseia no trabalho realizado por Kastens et al. (2005) e Mello et al. (2014) no qual, por meio da correlação entre os valores de células do produto de NDVI do sensor AVHRR e valores anuais de produtividade de diversas culturas agrícolas, foram geradas máscaras da área plantada de cada cultura, que serviram como apoio aos modelos de estimativa de produtividade. Através da correlação entre os dados históricos de produtividade e de NDVI, foi possível obter as máscaras para alguns cultivos, que serviram como parâmetro de entrada para o modelo agrometeorológico. Os resultados obtidos pelos modelos, com e sem as máscaras, foram similares.

Na abordagem sugerida na presente pesquisa, a correlação entre os dados históricos de produtividade e os dados de NDVI foi calculada célula a célula, no entanto, ao invés de ser definido um limiar no qual diversas células fossem selecionadas como máscara, apenas os valores das 200 células que melhor tiveram correlação com a produtividade, foram utilizados. Assim, este trabalho não tem como objetivo criar máscaras de cultivo, mas sim estimar a produtividade diretamente, a partir dos valores de NDVI destas células.

Para os 200 valores, das células que apresentaram a melhor correlação com dados de produtividade, foi dada a denominação de *proxies*. Então, a partir do valor escalonado destes *proxies*, foi computado o valor da mediana. O valor da mediana foi escalonado para valores de produtividade e assim, foi obtida a estimativa de produtividade da cana-de-açúcar para cada um dos municípios.

3.3.2. Automação dos processamentos de obtenção dos proxies

Para tornar possível o cálculo da correlação de cada uma das células, foi necessária a automação dos processos. A maior parte do processamento dos dados foi automatizada através de um script desenvolvido na linguagem de programação Python.

A partir do arquivo de centroides, gerado no pré-processamento, o script permite definir diferentes áreas de abrangência, de acordo com a escolha do usuário, que foram geradas para cada município e foram armazenadas como vetores e, também, convertidas para o formato matricial, através da função *rasterize* do pacote GDAL. Este arquivo matricial apresenta células com valor 1 (para a região interna da área de abrangência) e 0 (quando fora desta região). Estes valores foram usados para álgebra de mapas, a fim de recortar a região de interesse de cada município.

Em seguida foi aplicada a função de interseção entre os polígonos da área de abrangência e o arquivo vetorial de limites dos *minitiles*. O resultado é uma lista com o código de cada *minitile* necessário para compor a área de abrangência. Nesse momento o algoritmo define quais *minitiles* serão utilizados para compor cada mosaico, sendo que os parâmetros de entrada serão a lista e o DOY (dia do ano) que será processado, essa etapa foi denominada de mosaico espacial. Ao término do mosaico para todos os DOY entre o DOY_i (dia do ano inicial) e DOY_f (dia do ano final) é realizado o empilhamento destes mosaicos, formando um único arquivo matricial com 10 bandas, esse processo foi denominado de mosaico temporal.

Estes mosaicos foram recortados através do processo de álgebra de mapas, usando o arquivo matricial da área de abrangência, multiplicado pelas bandas

do arquivo matricial de NDVI, assim foram eliminadas as células fora da área de abrangência.

Por fim, estes dados são escalonados usando-se o Zscore (Equação 3.1) ou o Rscore (Equação 3.2), dependendo da escolha feita no início do processamento. O conjunto de dados V representa o conjunto de valores de NDVI para a série histórica analisada. No entanto, a série histórica de dados do IBGE também é normalizada pelas mesmas funções, assim V também pode ser entendido como a série histórica de dados do IBGE.

$$Z = \frac{v - \mu(V)}{\sigma(V)} \quad (3.1)$$

Onde: Z é o valor escalonado através do Zscore; v é o valor NDVI a ser escalonado; μ é a média dos valores; σ é o desvio padrão; V é o conjunto de dados observados;

$$R = \frac{v - \min(V)}{\max(V) - \min(V)} \quad (3.2)$$

Onde: R é o valor escalonado através do Rscore e min e max referem-se, respectivamente, ao mínimo e ao máximo valor observado no conjunto de dados V .

Depois de escalonar os valores de NDVI dos *minitiles*, são então escalonados os valores de produtividade de cada município, em ambos os casos utiliza-se uma ou outra equação descrita anteriormente. A conversão para uma mesma escala é necessária para que a comparação entre os dados de NDVI e de produtividade seja possível por meio do RMSE.

Depois de escalonados, foi calculado o RMSE ou ρ de Pearson, conforme o caso, para cada uma das células. Os critérios de seleção dos *proxies* foram: a) menor RMSE; b) ρ de Pearson mais próximo a 1.

O RMSE (Equação 3.3) é uma medida da diferença entre valores preditos por um modelo e os valores realmente observados (RICE, 2001). Este método não é escalonado e não é possível comparar seu valor em diferentes conjuntos de

dados. No entanto, o valor de RMSE pode ser utilizado para distinguir o desempenho de um modelo dentro de um período de calibração ou para comparar os desempenhos individuais de um modelo, sendo o valor ideal igual a zero (inexistência de erros entre o modelo e os dados observados).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2} \quad (3.3)$$

Onde: RMSE é o valor da raiz do erro médio quadrático; y_i é o valor observado na posição i e \hat{y}_i é o valor modelado na posição i .

O ρ de Pearson (Equação 3.4) é a proporção da variabilidade de um conjunto de dados que pode ser explicada por meio de um modelo estatístico (RICE, 2001). É comum encontrarmos na literatura a denominação de coeficiente de correlação ou coeficiente de correlação de Pearson. Este valor representa o quanto da variabilidade dos dados é explicada por meio do modelo proposto.

$$\rho = \frac{\sum_{i=1}^n (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^n (x_i - \hat{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \hat{y})^2}} \quad (3.4)$$

Onde: ρ é o valor do coeficiente de correlação de Pearson (ρ de Pearson), x_i e y_i assumem os valores medidos das variáveis x e y , respectivamente e \hat{x} e \hat{y} são as médias aritméticas das observações em x e y .

O cálculo do RMSE e ρ de Pearson foi realizado célula a célula para cada semana de cada ano, para cada município selecionado. O cálculo realizado célula a célula é apresentado na Figura 3.4, utilizando-se 50 células como exemplo. A partir dessa comparação é obtido o valor do RMSE ou do ρ de Pearson para cada uma das células da cena, que são armazenados em formato matricial, extensão GeoTIFF, conforme apresentado na Figura 3.5. Os valores deste arquivo matricial são ordenados, e as 200 células que apresentaram os melhores valores (menor RMSE ou ρ de Pearson mais próximo de 1) foram selecionados.

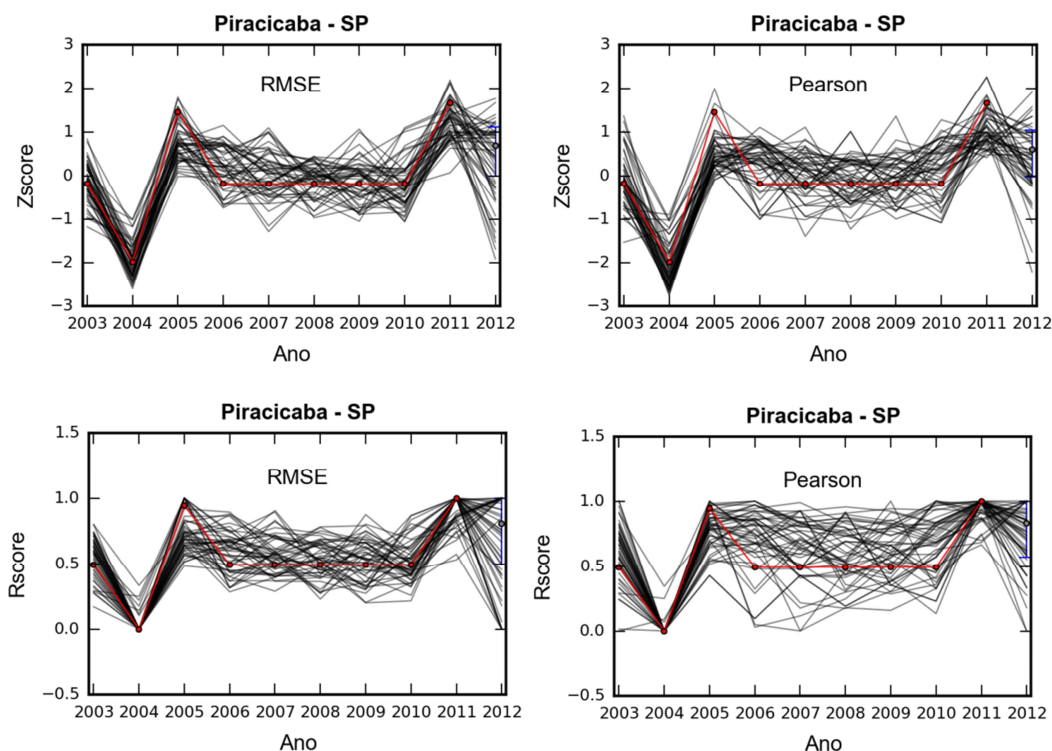


Figura 3.4 – Exemplo da seleção dos *proxies* para estimativa da produtividade da cana-de-açúcar para o ano de 2012 do município de Piracicaba. As linhas pretas representam as variações anuais dos valores escalonados de NDVI de 50 células e a linha vermelha representa o valor escalonado de produtividade da cana-de-açúcar.

Os dados referentes aos *proxies* (apenas 200 melhores valores) foram armazenados em tabelas, no formato .csv. Nesta tabela constam os valores de RMSE ou ρ de Pearson de cada *proxy*, suas coordenadas de latitude e longitude, o código IBGE do município, o dia do ano que foi processado, o ano e os valores escalonados das bandas MODIS. Um exemplo desta tabela pode ser visto no Apêndice A.3.

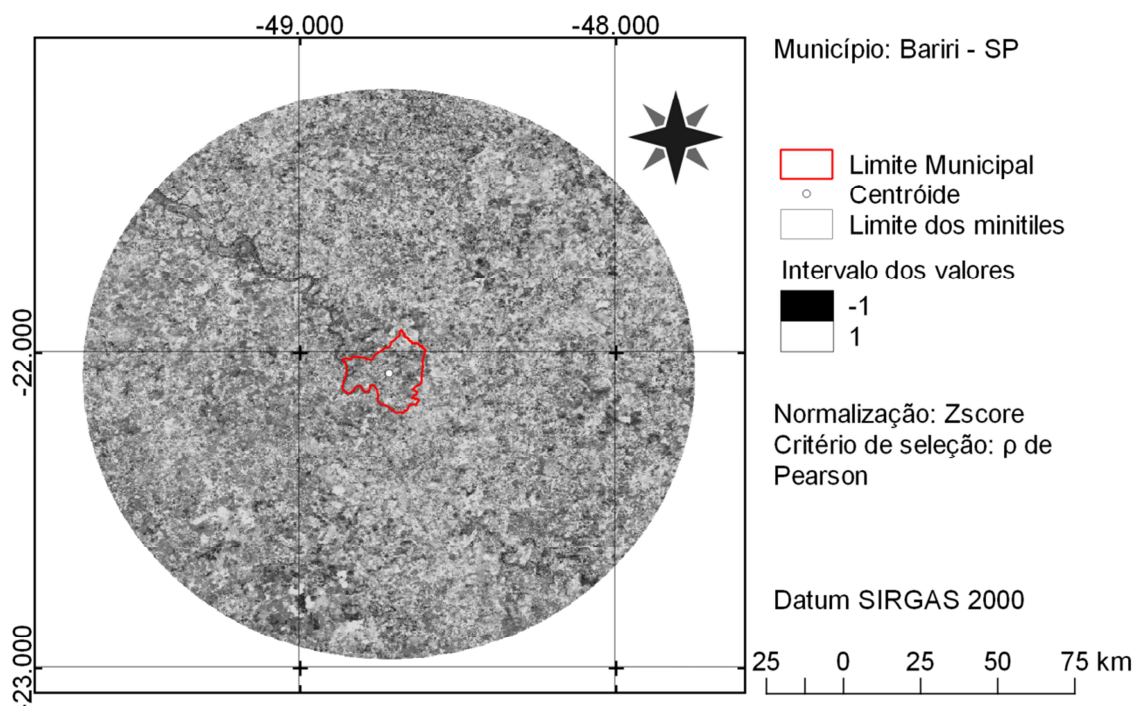


Figura 3.5 – Recorte da área de abrangência para o município de Bariri – SP, com raio de 90 km. Os valores das células dentro da área circular representam os valores do critério de seleção p de Pearson. Os valores destas células foram ordenados e os 200 valores mais próximos de 1 são utilizados como *proxies* para estimar a produtividade.

3.3.3. Variáveis utilizadas

Apesar do processo de obtenção dos *proxies* ter sido automatizado, alguns parâmetros de entrada precisaram ser definidos a fim de tornar o método adaptável para outras regiões e outras culturas. Assim, todos os processamentos foram realizados com base em alguns parâmetros de entrada, que definem o cenário a ser processado em busca de tais *proxies*. Estes parâmetros são apresentados a seguir.

- a) Municípios: municípios que serão processados em cada cenário, podendo ser um município ou um grupo. Informa-se aqui o código de sete dígitos de identificação do(s) município(s) seguindo o padrão adotado pelo IBGE. Este parâmetro foi criado para que, este método, possa ser adaptado para outras regiões;

- b) Ano inicial: refere-se ao primeiro ano da série de dados a ser processado. No presente trabalho o ano inicial considerado foi o de 2003;
- c) Ano final: refere-se ao último ano da série de dados que será processado. No presente trabalho o ano final considerado foi 2012;
- d) Dia do ano inicial (DOYi): dentro do período de anos é necessário determinar para qual ou para quais dias do ano os dados serão processados. Como o produto NDVI de BOKU é disponibilizado semanalmente (seção 2.5), é possível escolher qualquer uma dentre as 52 semanas de um ano para o processamento. Assim, neste parâmetro define-se a partir de qual dia do ano inicial será realizado o processamento;
- e) Dia do ano final (DOYf): este parâmetro é complementar ao DOYi (item “d”, acima). Deve-se informar aqui até qual dia do ano final será processado;
- f) Cultura: refere-se ao tipo de cultura a ser utilizada no processamento. O presente trabalho restringiu-se apenas à cana-de-açúcar, mas poderiam ser consideradas, com as devidas adaptações, outras culturas. Alterando-se este parâmetro, altera-se a cultura que será extraída do histórico de produtividade do IBGE;
- g) Tempo de atraso (LAG): este parâmetro refere-se ao tipo de produto disponibilizado pela universidade de BOKU. Como descrito no item 2.5 do presente documento, os *minitiles* são disponibilizados em cinco diferentes tipos. Assim, um LAG igual a 0, significa que o processamento será realizado usando-se o produto N; para um LAG igual a 1 significa que será utilizado o produto A. O valor do LAG vai de 0 (quando se utiliza o produto N) até 4 (quando se utiliza o produto D);

- h) Tempo de integração (T.I.): o processamento pode ser realizado usando-se apenas um determinado produto, mas também pode ser feito utilizando-se a média de mais de um produto, que chamamos aqui de “tempo de integração”. Por exemplo, para um T.I. igual a 1 e LAG igual a 0, apenas o produto N será utilizado na análise. Para um T.I. igual a 2 e LAG igual a 0, tanto o produto N quando o produto A serão utilizados na análise. Quando o tempo de integração é maior que 1, os produtos são somados e o valor médio de cada célula é então utilizado na análise. O T.I. pode receber até o valor 10, significando que os produtos das últimas 10 semanas serão utilizados. O tipo de produto utilizado é dependente do valor do T.I. e do LAG. As combinações possíveis de T.I. e LAG podem ser verificadas no Apêndice A.1;
- i) *Buffer*: refere-se à área de abrangência que delimita a máxima distância, na qual as células do produto disponibilizado pela universidade de BOKU serão processadas. Os tamanhos da área de abrangência, testados neste trabalho, foram círculos com raios de 30, 60, 90 e 120 km, tomando o centroide de cada município como o centro;
- j) Método de escalonamento: os dados de produtividade da cana-de-açúcar e os valores de NDVI obtidos a partir dos *minitiles* apresentam unidades de medida em diferentes escalas. Para permitir análises correlativas entre estas variáveis, ambas foram convertidas para uma mesma escala. Para tanto, estes valores foram escalonados utilizando-se os métodos de Zscore e Rscore, sendo posteriormente selecionado o método que produziu os melhores resultados;
- k) Critério de seleção: os *proxies* são obtidos através da comparação entre a série histórica dos valores do produto de NDVI do sensor MODIS disponibilizado pela universidade de BOKU e os correspondentes valores de produtividade do IBGE, por município. Os valores das células que apresentam melhor concordância entre o

comportamento dos valores de NDVI e dos dados de produtividade são então escolhidos como *proxies* da produtividade. Para obter-se as células com melhor concordância, foram utilizados o RMSE e o ρ de Pearson;

- I) Quantidade de proxies: por último, definiu-se a quantidade de células a serem utilizados como *proxies* da produtividade. Por exemplo, quando este critério é definido como 100, as 100 células que melhor acompanharam as nuances do comportamento dos dados de produtividade foram selecionados como *proxies*. Nesta pesquisa, foram considerados 200 *proxies*.

3.3.4. Avaliação dos parâmetros

O termo “cenário” define um conjunto de parâmetros que foram adotados durante o processamento. Com isso, entende-se que apenas os parâmetros variáveis são analisados para cada cenário, sendo que os demais estão fixados e não influenciam os resultados. Essa definição possibilitou que cada um dos objetivos específicos deste trabalho fosse testado individualmente. Os cenários apresentados na Tabela 3.2 apresentam cada objetivo específico (Seção 1.2) e os respectivos parâmetros adotados para cada cenário

Tabela 3.2 – Configuração dos cenários analisados

Objetivo Específico	Parâmetros						
	Dia do ano	LAG	IT	Buffer (km)	Qtde Proxies	Método normalização	Critério Seleção
Definir Zscore ou Rscore e Pearson ou RMSE	84	0	1	60	200	Rscore x Zscore	Pearson x RMSE
Tamanho Buffer	84	0	1	30, 60, 90, 120	200	Zscore	RMSE
Produto MODIS, Integração	84	0	1, 2, 4, 6, 8, 10	60	200	Zscore	RMSE
Produto MODIS, Atraso	84	0, 1, 2, 3, 4	1	60	200	Zscore	RMSE

A escolha do DOY dos cenários se deu a partir da análise dos resultados obtidos previamente por Mello et al. (2014) (Figura 3.6). Nessa figura é possível

observar que as melhores estimativas foram obtidas até o final do mês de março. Por essa razão, para os cenários que foram analisados para uma única data, o DOYi e DOYf foi definido como 84.

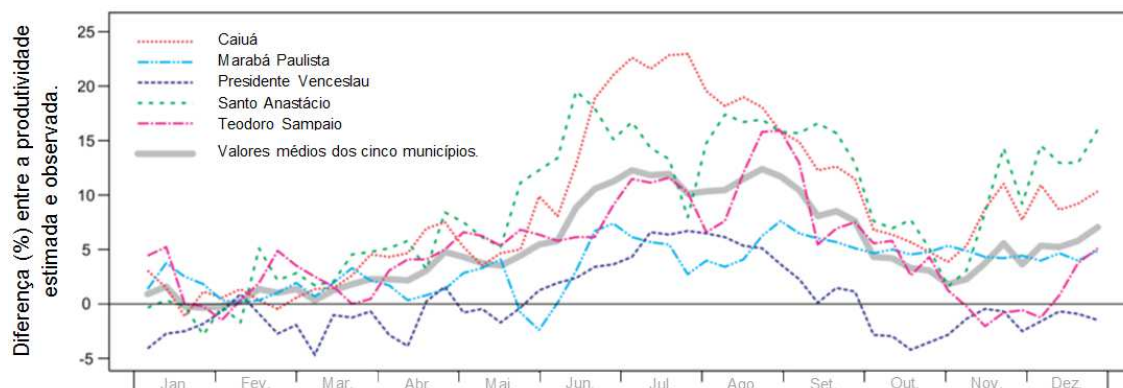


Figura 3.6 – Diferença entre os valores estimados e observados para todas as semanas do ano de 2012 para cinco municípios do estado de São Paulo. Os melhores resultados ocorrem até o final do mês de março. Fonte: adaptado de Mello et al. (2014).

3.3.5. Análise das estimativas de produtividade

Depois de definidos os cenários para a avaliação de cada um dos objetivos específicos foram então geradas estimativas de produtividade para o ano de 2012 utilizando-se cada um dos cenários descritos anteriormente (Tabela 3.2). Dessa maneira, foi possível avaliar quais parâmetros influenciaram de forma positiva nos resultados finais, bem como realizar as devidas análises estatísticas para que melhorias significativas pudessem ser identificadas.

Com o intuito de avaliar a possível influência da densidade de produção e da regionalidade, os resultados obtidos por municípios foram separados por grupos, de acordo com dois critérios independentes. No primeiro, o agrupamento foi realizado com base no valor de área colhida de cana-de-açúcar em 2012. Os limiares de separação de cada um dos grupos foram definidos utilizando o método de K-médias (SCHOWENGERDT, 1997). Com isso, foram criados cinco grupos (Tabela 3.1). No segundo agrupamento, os municípios foram separados de acordo com a mesorregião a que pertencem. Por essa razão, os resultados são apresentados por mesorregião geográfica

(IBGE, 2014). Assim, o primeiro agrupamento permitiu que diferentes variações relacionadas ao tamanho da área colhida fossem analisadas enquanto que o segundo agrupamento permitiu verificar as diferenças relacionadas com a dispersão geográfica.

Ao término do processamento das estimativas, fez-se necessário avaliar a qualidade das estimativas geradas. Desta forma, para esta validação, foram comparados os valores obtidos pelas estimativas com os valores divulgados pelo IBGE, para o ano de 2012.

4 RESULTADOS E DISCUSSÃO

A base do método que foi proposta nesta pesquisa, foi de encontrar, em imagens multitemporais do produto de NDVI do sensor MODIS, células que possam servir como representantes (*proxies*) das variações climáticas que afetam o NDVI e permitem inferir sobre a produtividade da cana-de-açúcar para um dado ano-safra de interesse. Se, por exemplo, um órgão oficial necessitar da previsão da produtividade da cana-de-açúcar para um município, uma mesorregião ou para todo o Estado de São Paulo, o método em proposição possibilitará a obtenção de uma estimativa da produtividade via *proxies* antes da divulgação das estimativas oficiais.

Os resultados apresentados nesta seção seguem a ordem dos cenários estudados, conforme mostrado na Tabela 3.2 e, ao final, são apresentados os resultados obtidos por meio dos melhores ajustes encontrados. Cada cenário objetivou avaliar alguns parâmetros do método, conforme a sequência de objetivos específicos listada no item 1.2 do presente documento, sendo os resultados apresentados nas seções 4.1 a 4.4. Nas seções 4.5, 4.6 e 4.7 são apresentados os valores das estimativas para o ano de 2012, agrupados por quantidade de área colhida e por mesorregiões, assim como por municípios.

4.1. Escalonamento e critério de seleção

Os métodos de escalonamento testados foram o Zscore (Eq. 3.1) e o Rscore (Eq. 3.2). A avaliação foi feita através do teste t pareado, sendo a hipótese nula de que não há diferença entre as médias dos valores observados e estimados por ambos os métodos de escalonamento (CRAMÉR, 1946). Da mesma maneira foram analisados dois critérios de seleção sendo eles o RMSE e o p de Pearson.

As etapas de análise dos parâmetros de escalonamento e dos critérios de seleção foram realizadas conjuntamente, testando-se as combinações possíveis entre eles. Para isso, os dados foram escalonados utilizando-se o Zscore e os *proxies* selecionados usando-se o menor RMSE e o p de Pearson

mais próximo de 1. Em seguida, os dados foram escalonados utilizando-se o Rscore e os *proxies* foram novamente selecionados utilizando-se os dois critérios de seleção. A Figura 3.4 apresenta os exemplos da estimativa de produtividade para o município de Piracicaba para o ano de 2012 utilizando-se as quatro combinações e na Figura 4.1 são apresentados os valores médios, medianas e variâncias das estimativas de cada uma destas combinações, por meio do boxplot, para os 361 municípios selecionados para a execução deste trabalho.

Na Figura 4.1 é possível notar que a variância dos resultados obtidos por meio do RMSE é menor do que a variância dos erros obtidos por meio do p de Pearson. Da mesma forma, os resultados das estimativas obtidas quando utilizada o escalonamento Zscore tiveram menor variância em relação às obtidas por meio do Rscore.

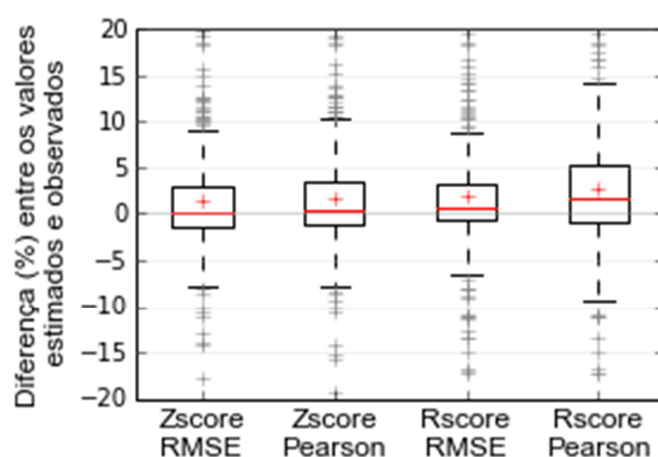


Figura 4.1 – Comparativos das quatro combinações entre os métodos de escalonamento e critério de seleção dos *proxies* para os 361 municípios da área de estudo, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros.

Ainda, para definir qual destas combinações obteve os melhores resultados, foram realizadas análises por meio do teste t pareado com base nos valores médios dos erros. Na Tabela 4.1, apresentam-se os erros médios entre os valores observados e estimados utilizando-se cada uma das combinações de escalonamento e critério de seleção, bem como o valor p, resultado do teste t, para cada uma das combinações possíveis.

Tabela 4.1 – Resultados com o valor p do teste t pareado para cada combinação de escalonamento e critério de seleção. Em negrito estão destacados os valores que apresentaram diferença estatística a 5% de significância.

Normalização	Critério Seleção	Média	Valor p			
			Zscore RMSE	Zscore Pearson	Rscore RMSE	Rscore Pearson
Zscore	RMSE	1,46	-	~ 0,00	0,060	~ 0,00
Zscore	Pearson	1,70	~ 0,00	-	0,318	~ 0,00
Rscore	RMSE	1,97	0,060	0,318	-	~ 0,00
Rscore	Pearson	2,74	~ 0,00	~ 0,00	~ 0,00	-

Analisando-se estes valores, percebe-se que o menor erro médio apresentado foi para a combinação Zscore + RMSE, com diferenças significativas à 5% quando testado contra as combinações Zscore + p de Pearson e Rscore + p de Pearson. Apesar dos resultados obtidos pelas combinações Zscore + RMSE e Rscore + RMSE não apresentarem diferenças significativas, optou-se por utilizar o Zscore por ter apresentado valor médio dos erros menor quando comparado ao Rscore.

Segundo Kastens et al. (2005), o p de Pearson mede a colinearidade entre os dados enquanto o RMSE mede a magnitude do erro e por essa razão pode não ter sido tão eficaz na seleção dos *proxies* quando comparado ao RMSE. Um exemplo que demonstra a diferença entre estes dois critérios é apresentado na Figura 4.2.

Assim, estes resultados mostram que, em geral, os erros obtidos por meio do escalonamento Zscore e do critério de seleção RMSE, foram mais próximos de zero, o que indica que estes dois parâmetros possibilitam melhores estimativas de produtividade para a cana-de-açúcar do que os demais métodos apresentados.



Figura 4.2 – Exemplo com dados fictícios no qual são apresentados dois conjuntos de dados. A diferença em cada observação é de 0,05 unidades de medida. O ρ de Pearson, por avaliar a colinearidade dos valores, teve valor igual a um, enquanto que o RMSE, que mediu a magnitude do erro, conseguiu detectar as diferenças entre os valores de cada amostra.

4.2. Avaliação do tamanho da área de abrangência

É oportuno indicar que o tamanho da área de abrangência representa diretamente a quantidade de células que foram processadas e consequentemente, reflete no tempo de processamento. Além disso, uma área de abrangência com maior raio significa que células mais distantes do centroide do município serão selecionadas. Essa distância do centroide do município implica diretamente na correlação entre a produtividade observada neste município e os valores do produto NDVI do sensor MODIS, pois quanto maior o número de células analisadas, maior a chance de células que apresentaram maior correlação com os dados de produtividade serem selecionadas. Para a análise dos diferentes tamanhos da área de abrangência foram utilizadas circunferências criadas a partir do centroide de cada município com raios de 30 km, 60 km, 90 km e 120 km.

Assim, partiu-se de dois pressupostos: a) um maior número de células analisadas (área de abrangência maior) significa que haverá a possibilidade de selecionar células com valores mais próximos do ideal (zero para RMSE) do que uma área de abrangência menor; b) os valores de NDVI das células mais

distantes do centroide, mesmo que apresentem boa correlação com a produtividade histórica do município, podem não representar exatamente as mesmas condições climáticas que o município estudado.

A avaliação para definir qual a melhor área de abrangência, foi realizada através do teste t pareado. Buscou-se identificar quais tamanhos da área de abrangência apresentaram os menores erros, entre os valores observados e estimados, e que fossem estatisticamente significantes. A Figura 4.3 apresenta os resultados gerais, levando em consideração os 361 municípios analisados, sem estratificação por grupos.

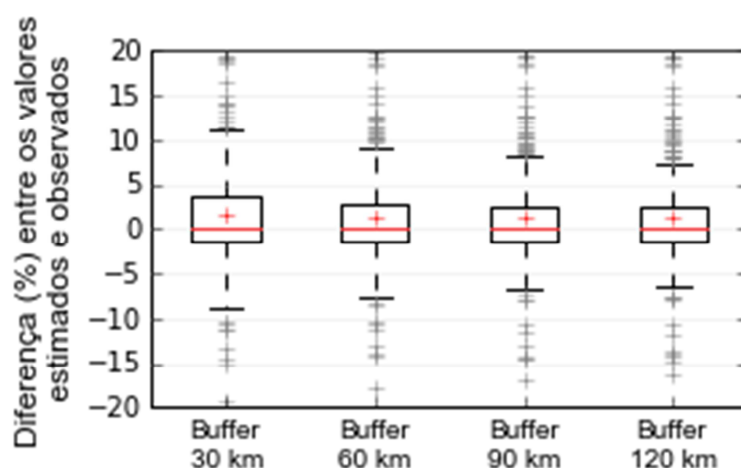


Figura 4.3 – Comparativo entre os diferentes tamanhos da área de abrangência, para os 361 municípios da área de estudo, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros.

Os resultados do teste t pareado são apresentados na Tabela 4.2. Os valores médios dos erros que apresentaram diferença à 5% de significância estão destacados em negrito.

Tabela 4.2 – Resultados do teste t pareado comparando os diferentes tamanhos das áreas de abrangência. Em negrito estão destacados os valores que apresentaram diferença estatística a 5% de significância

	Média	Buffer (km)			
		30	60	90	120
30	1,63	-	0,01	~0,00	~0,00
60	1,46	0,01	-	~0,00	0,01
90	1,32	~0,00	~0,00	-	0,84
120	1,31	~0,00	0,01	0,84	-

As áreas de abrangência de 30 km, 60 km, 90 km apresentaram diferença estatística significativa, sendo o melhor resultado obtido quando utilizado o valor de 90 km. A área de abrangência com raio de 120 km não apresentou diferença significativa quando comparado ao raio de 90 km. A Figura 4.4 apresenta os valores médios, dos erros das estimativas, considerando os 361 municípios, calculados utilizando-se cada um dos tamanhos da área de abrangência.

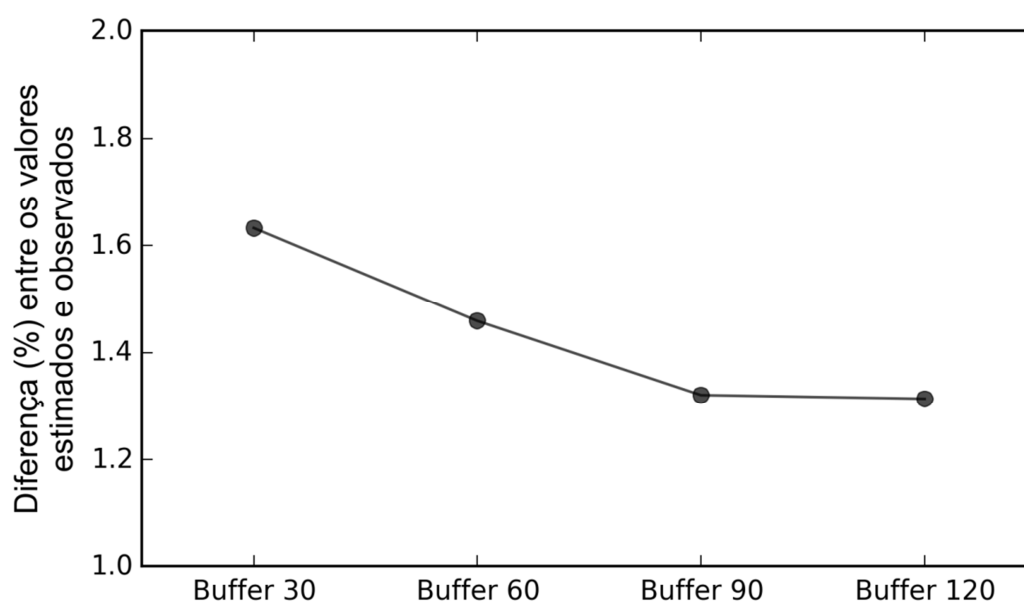


Figura 4.4 – Erros médios para os quatro diferentes tamanhos das áreas de abrangência.

Os resultados obtidos nas análises estatísticas da área de abrangência, mostraram que as melhores estimativas, foram obtidas utilizando-se o raio de 90 km, sendo que o raio de 120 km não melhorou as estimativas obtidas.

No entanto, ressalta-se aqui que, para critério de avaliação dos demais parâmetros, utilizou-se o raio de 60 km, uma vez que a redução de 90 km para 60 km implica na redução para aproximadamente 44% do número de células analisadas em cada parâmetro, que reduziu o tempo de cada processamento em aproximadamente 4 horas para cada processamento.

4.3. Análise do tempo de integração (T.I.)

O tempo de integração (T.I.) tem relação direta com a qualidade e a quantidade de produtos da série temporal de NDVI disponibilizados pela universidade BOKU que serão utilizados (Figura 2.1). O tempo de integração determina a quantidade de produtos que serão utilizados para o cálculo do valor médio de cada célula. Esse parâmetro influencia diretamente no tempo necessário para divulgação das estimativas, uma vez que a integração de uma quantidade maior de produtos implica no uso de produtos que precisam de dados de mais de uma semana de processamento diferente (Tabela 2.1).

A Figura 4.5 apresenta a comparação entre diferentes T.I.s tendo como base o erro relativo das estimativas de produtividade dos 361 municípios selecionados para o ano de 2012 e dia do ano 84.

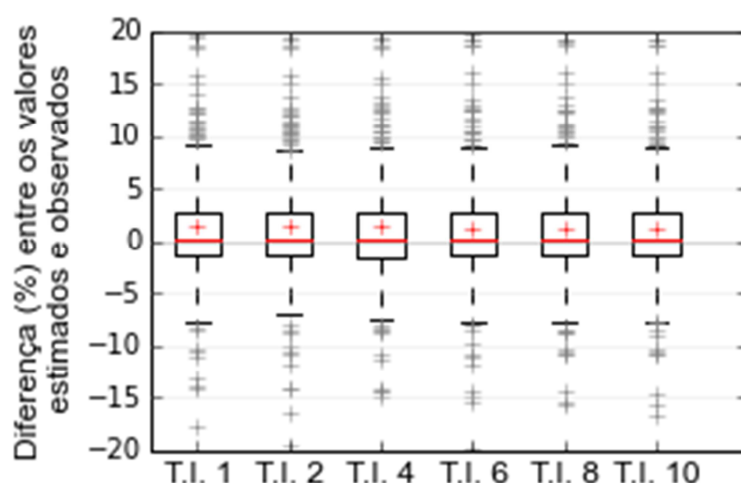


Figura 4.5 – Comparativo entre os diferentes tempos de integração, com tempos de integração de 1 até 10 semanas para os 361 municípios selecionados nesta pesquisa, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros

Os erros médios das estimativas de produtividade utilizando os diferentes T.I.s avaliados nesta pesquisa são apresentados com maiores detalhes na Figura 4.6 e os resultados dos testes t aplicados a estas médias, são apresentados na Tabela 4.3.

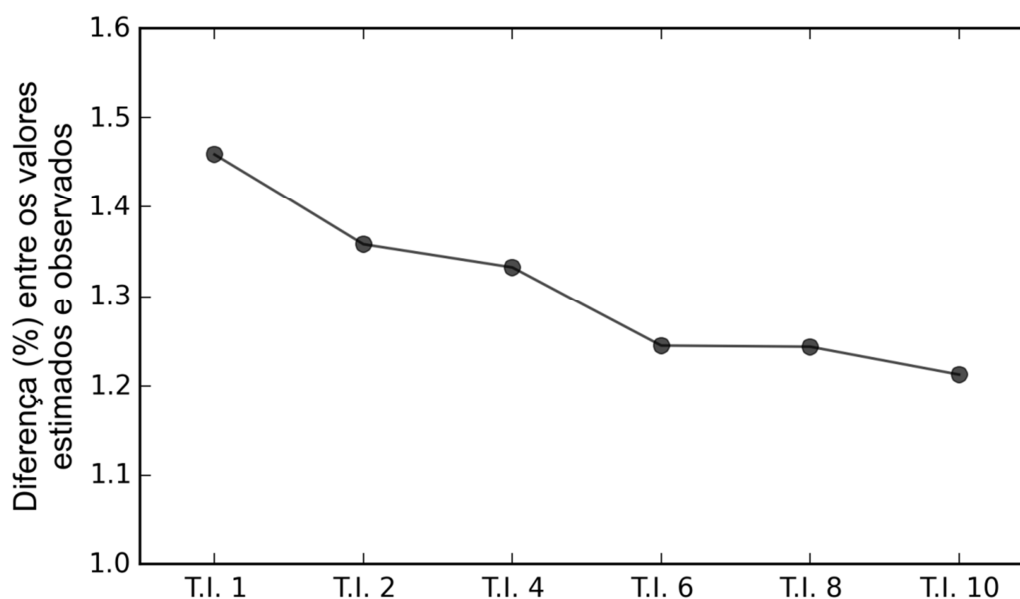


Figura 4.6 – Erros médios das estimativas calculadas para os 361 municípios selecionados, através dos diferentes tempos de integração.

Os valores obtidos (Figura 4.6) mostram que há diminuição nos erros obtidos por meio de cada um dos T.I.s investigados nesta pesquisa. A Tabela 4.3 mostra os resultados dos testes t aplicados para estas médias, que apontam que a partir do T.I. 6 não há diferença estatística com 5% de significância.

Tabela 4.3 – Valor p do teste t pareado aplicado aos tempos de integração de 1 a 10. Os valores destacados em negrito foram os que apresentaram diferença estatística a 5% de significância.

	Média	Tempo de integração					
		1	2	4	6	8	10
1	1,46	-	0,02	0,04	~0,00	~0,00	~0,00
2	1,36	0,02	-	0,53	0,01	0,01	~0,00
4	1,33	0,04	0,53	-	0,01	0,01	~0,00
6	1,25	~0,00	0,01	0,01	-	0,96	0,27
8	1,24	~0,00	0,01	0,01	0,96	-	0,12
10	1,21	~0,00	~0,00	~0,00	0,27	0,12	-

De acordo com os resultados apresentados na Tabela 4.3, os tempos de integração de seis semanas ou mais apresentam resultados com diferenças estatisticamente não significativas. No entanto, o T.I. 6 implica em um atraso de 35 dias para a estimativa dos dados, uma vez que exige a integração dos produtos N, A, B, C, D1 e D2 (Apêndice A.1). Os resultados dos testes também mostram que não há diferença significativa entre utilizar um tempo de

integração T.I. 2 e T.I. 4. De fato, o T.I. 2 implica em um atraso de apenas 7 dias entre o processamento nos dados da universidade de BOKU e a divulgação, pois necessita apenas dos produtos N e A, o que pode tornar sua utilização mais adequada.

Como o método proposto tem a finalidade de ser utilizado em aspecto operacional, o T.I. 2 seria mais adequado, quando considerado o tempo para divulgação, enquanto que o T.I. 6 é o mais adequado em relação à precisão das estimativas. No entanto, os resultados que serão apresentados na seção 4.4, mostram que a influência do parâmetro de tempo de atraso (LAG) é mais determinante nos resultados finais e por essa razão o T.I. 1 foi definido em todos os processamentos.

4.4.Avaliação do tempo de atraso (LAG)

O tempo de atraso foi um parâmetro criado para definir qual produto NDVI do sensor MODIS disponibilizado pela universidade de BOKU será utilizado dentro do método proposto. Além da possibilidade de integração de diferentes produtos (tempo de integração, seção 4.3), há a possibilidade de atrasar as estimativas em razão do uso de um produto que melhore os resultados finais. Por exemplo, caso o produto N não apresente bons resultados, as estimativas podem ser realizadas utilizando o produto A, ou B, ou qualquer outro que melhore os resultados obtidos consideravelmente. No entanto, cada escolha dessas afetará o tempo de divulgação em 7 dias.

Para avaliar os resultados obtidos por meio dos diferentes tempos de atraso, foi realizado o teste t pareado sobre os valores médios das estimativas calculadas para cada um dos tempos de atraso propostos nesta pesquisa. A Figura 4.7 apresenta os boxplots comparando os resultados obtidos por cada um dos tempos de atraso.

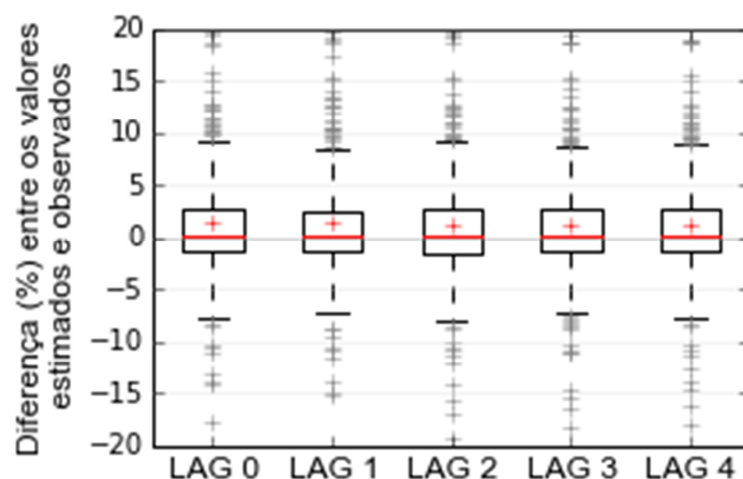


Figura 4.7 – Comparativo entre os diferentes tempos de atraso (LAG), com valores de 0 semanas até 4 semanas, para os 361 municípios selecionados nesta pesquisa, no qual a linha vermelha representa a mediana dos erros e o sinal “+” na cor vermelha representa a média dos erros.

Os valores médios, dos erros entre as estimativas obtidas por este método e os valores divulgados pelo IBGE, são apresentados na Figura 4.8. Os resultados dos testes estatísticos são apresentados na Tabela 4.4.

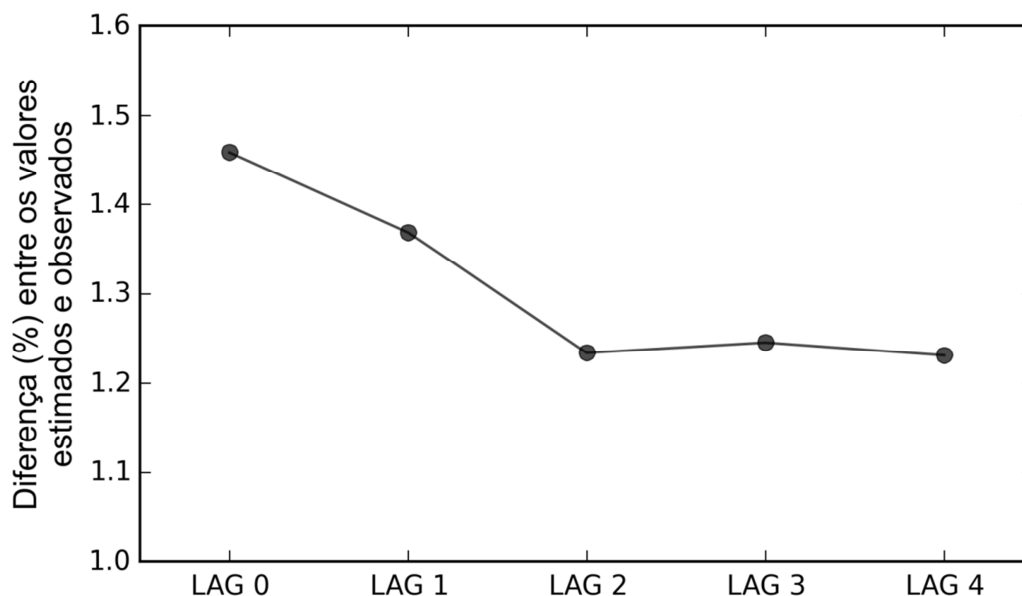


Figura 4.8 – Erros médios entre os valores estimados pelo método proposto nesta pesquisa e os dados de produtividade divulgados pelo IBGE para os 361 municípios da área de estudo, ano de 2012, com uso de cada um dos tempos de atraso (LAG) propostos.

O valor p do teste t pareado aplicado às médias dos diferentes tempos de atraso (Tabela 4.4) mostraram que os tempos de atraso LAG 2, LAG 3 e LAG 4 foram os que, em média, obtiveram os menores erros, com significância de 5%.

Esse resultado reforça o que foi encontrado para o T.I., uma vez que o tempo de atraso também mostrou que os produtos N e A não apresentaram as melhores estimativas.

Tabela 4.4 – Valor p do teste t pareado aplicado aos LAGs de 0 a 4. Os valores destacados em negrito foram os que apresentaram diferença estatística a 5% de significância.

	Média	Tempo de atraso				
		LAG 0	LAG 1	LAG 2	LAG 3	LAG 4
LAG 0	1,46	-	0,12	~0,00	~0,00	~0,00
LAG 1	1,37	0,12	-	0,01	0,04	0,03
LAG 2	1,23	~0,00	0,01	-	0,82	0,96
LAG 3	1,25	~0,00	0,04	0,82	-	0,74
LAG 4	1,23	~0,00	0,03	0,96	0,74	-

É importante notar também que o LAG 2 não apresentou diferenças estatísticas à 5% de significância quando comparado ao LAG 3 e LAG 4, além do mais, o LAG 2, dentre os três LAGs com diferenças significativas, é o que implica em um menor tempo entre o processamento dos dados e a divulgação, lembrando que o LAG 2 implica na utilização do produto B, enquanto o LAG 3 utiliza o produto C e o LAG 4 o produto D.

Quando comparados os valores médios da Tabela 4.3 com os valores da Tabela 4.4, percebe-se que o erro médio do LAG 2 é menor quando comparado ao erro médio de qualquer um dos T.I.s. Por essa razão, optou-se então por utilizar o T.I. 1 e LAG 2 para o processamento dos resultados das seções 4.5, 4.6 e 4.7.

De fato, a integração por meio do cálculo da média entre diferentes produtos não se mostrou um bom fator para melhorar a estimativa de produtividade. O trabalho de Morel et al. (2014), utilizando 64 áreas de estudo em *Reunion Island* (ilha pertencente à França) mostrou que, o valor acumulado do NDVI é um bom estimador da biomassa da cana-de-açúcar. Assim, a utilização do valor médio pode não ter sido um bom critério e futuramente deverá ser substituído pelo valor acumulado.

4.5. Estimativas agrupadas por valor de área colhida

A partir dos resultados obtidos para cada parâmetro, foi criado um novo cenário, que utilizou os melhores valores dos parâmetros, encontrados nos testes anteriores. Assim, este cenário ótimo utiliza T.I. 1 e LAG 2, ou seja, apenas o produto B será utilizado, mas sem ser integrado com nenhum outro produto. Foram utilizados também como parâmetros: a) escalonamento por meio do Zscore; b) critério de seleção RMSE e raio da área de abrangência de 90 km. As estimativas obtidas neste cenário são apresentadas na Tabela 4.5.

Tabela 4.5 – Resultado das estimativas de produtividade da cana-de-açúcar por grupo, para o ano de 2012 utilizando os parâmetros que obtiveram os melhores resultados.

Grupo	Quantidade de Municípios	Área total (ha)	Erro médio (%)	Erro médio (t/ha)
1	52	42.904	-2,22	1,71
2	114	585.719	6,61	5,51
3	109	1.341.445	4,76	3,94
4	70	1.845.395	3,84	3,18
5	16	796.373	5,81	4,93

Os resultados obtidos por grupo tinham o objetivo de mostrar se municípios com maior valor de área colhida da cana-de-açúcar apresentariam melhores resultados em relação àqueles com menores áreas. No entanto, a Figura 4.9 mostra que o Grupo 1 (municípios com área colhida até 42.904 ha) foi o que apresentou a menor variância nos resultados.

Esperava-se que os municípios com a maior quantidade de área colhida fossem os que apresentariam as melhores estimativas, porém cinco dos 16 municípios pertencentes ao grupo 5 apresentaram erros superiores a 10 % sendo eles Barretos (28,07%), Olímpia (26,13%), Jabotical (20,04%), Guaíra (19,83%) e Jaú (11,16%). Os resultados para os demais municípios podem ser vistos no Apêndice A.2.

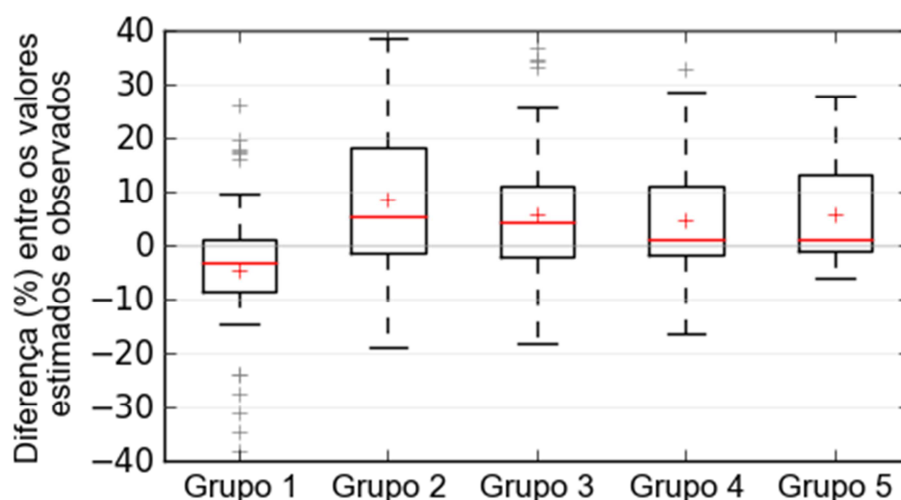


Figura 4.9 – Comparativo entre os erros relativos dos municípios de cada grupo, separados pelo valor de área colhida. Com exceção do grupo 1, os grupos apresentaram tendência de superestimar o valor de produtividade. A linha vermelha representa a mediana e o sinal “+” vermelho representa a média dos valores.

Os erros nas estimativas para estes cinco municípios afetaram o resultado geral do grupo e o mesmo efeito foi observado nos demais grupos. Alguns municípios tiveram diferenças maiores que 40% entre os valores observados e estimados da produtividade e por essa razão afetaram os resultados de seus respectivos grupos.

Assim, nada se pode dizer sobre a influência da variação na quantidade de área colhida dos municípios em relação a precisão das estimativas por meio deste método.

4.6. Estimativas por mesorregiões e por municípios

Os resultados das estimativas para o ano de 2012, por municípios, foram agrupados por mesorregião (Figura 4.10). Importante lembrar que, os resultados agrupados por mesorregiões não são a média aritmética dos erros de cada município da mesorregião, mas sim a média ponderada pela área. Ou seja, municípios com maior valor de área colhida têm maior representatividade nos resultados finais da mesorregião.

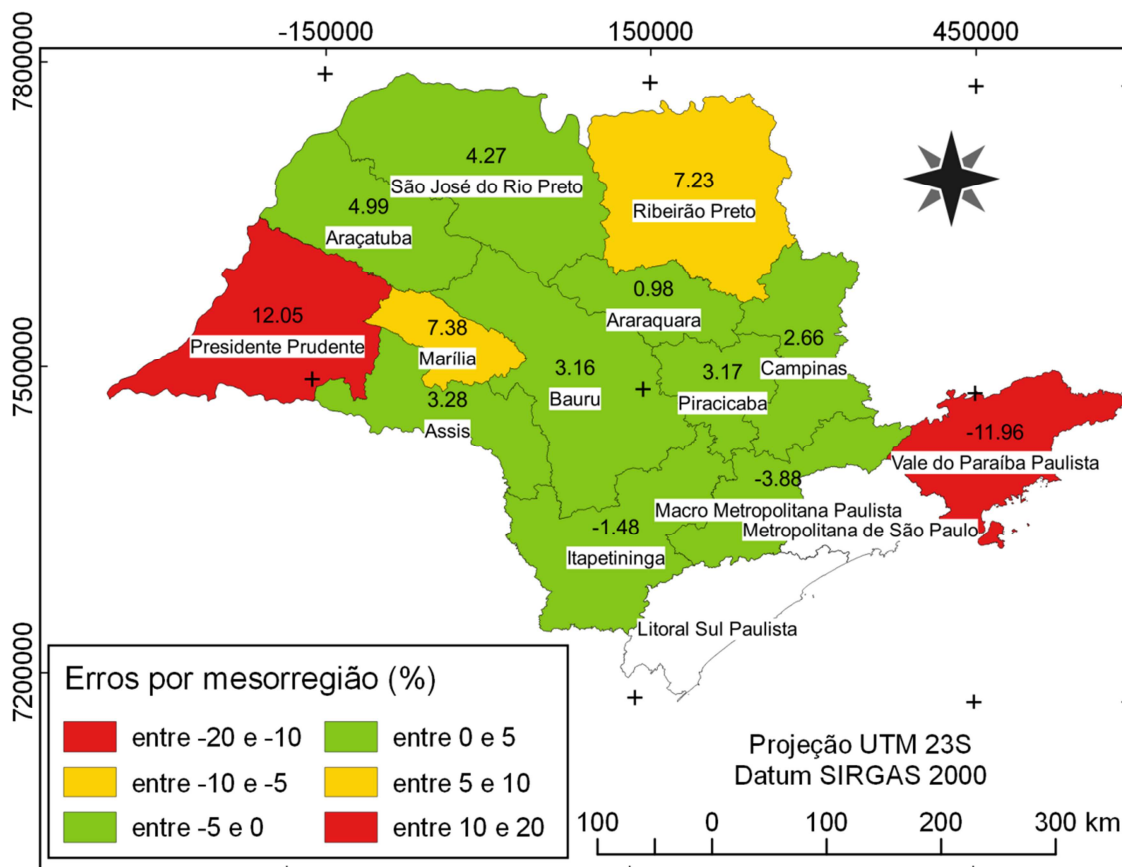


Figura 4.10 – Erros percentuais médios das estimativas de produtividade da cana-de-açúcar para o ano de 2012, por mesorregião geográfica. O erro apresentado é a diferença relativa entre os valores estimados e observados por mesorregião, considerando apenas os municípios utilizados neste trabalho.

Os resultados agrupados por mesorregiões mostraram que, os maiores erros ocorreram nas mesorregiões de Presidente Prudente e Vale do Paraíba Paulista. Para Presidente Prudente, o valor de produtividade foi superestimado em 12,05% em relação ao valor divulgado pelo IBGE, que representa aproximadamente 8,46 t/ha. Nesta mesorregião o erro está relacionado aos resultados obtidos para os municípios de Irapuru e Pacaembu, nos quais foram observadas diferenças de 49,67% e 45,65%, respectivamente, quando comparados aos dados divulgados pelo IBGE. No Vale do Paraíba Paulista, no qual a subestimativa foi de -11,96%, o que representa -7,21 t/ha, a maior diferença observada foi para o município de Queluz, no qual a produtividade da cana-de-açúcar foi subestimada em 27,39%. Além disso, a mesorregião do Vale do Paraíba Paulista teve uma amostra muito pequena, pois apenas 2 municípios (Tabela 4.6) foram utilizados nesta pesquisa.

Tabela 4.6 – Relação dos resultados obtidos por mesorregião geográfica, incluindo os dados de erro médio (%) e erro médio (t/ha).

Mesorregião	Quantidade de municípios	Área total (ha)	Erro médio (%)	Erro médio (t/ha)
São José do Rio Preto	72	794.174	4,27	3,46
Ribeirão Preto	61	1.259.604	7,23	5,72
Araçatuba	27	486.878	4,99	3,94
Bauru	35	426.164	3,16	2,50
Araraquara	20	359.605	0,98	0,82
Piracicaba	25	305.550	3,17	2,47
Campinas	32	225.161	2,66	2,23
Presidente Prudente	31	341.311	12,05	8,46
Marília	4	24.146	7,38	5,31
Assis	32	327.141	3,28	2,59
Itapetininga	8	45.670	-1,48	-1,25
Macro Metropolitana Pta.	12	16.366	-3,88	-3,08
Vale do Paraíba Pta.	2	66	-11,96	-7,22
Litoral Sul Paulista	0	-	-	-
Metropolitana de São Paulo	0	-	-	-

Em razão dos valores das estimativas serem obtidos a partir do valor da mediana das células do produto NDVI, variações extremas nos valores de NDVI não são percebidos. Assim, quando há variações bruscas nos valores de produtividade divulgados pelo IBGE (Tabela 4.7), este método não conseguiu detectar estas nuances e por isso, são nestes casos, que as piores estimativas são obtidas.

Tabela 4.7 – Produtividade divulgada pelo IBGE por ano para os municípios de Irapuru e Pacaembu.

	Ano									
Município	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Irapuru	116,5	100,0	77,5	130,0	120,0	100,0	110,0	110,0	110,0	70,0
Pacaembu	41,9	87,0	77,5	120,0	120,0	90,0	100,0	110,0	82,0	60,0

Observando os resultados por mesorregião é possível verificar que, para o ano de 2012, o método tendeu a subestimar a produtividade da cana-de-açúcar. Apesar de que, em geral, os erros por mesorregiões ficaram sempre menores que 5% em relação às estimativas oficiais.

A mesorregião de Ribeirão Preto, que representa a maior área colhida dentre todas as regiões, apresentou um erro de 7,23% em comparação aos dados divulgados pelo IBGE para os mesmos municípios. Os municípios que tiveram os piores resultados, para esta mesorregião, foram Altinópolis (-17,89%), Serra Azul (-10,36%), Santa Rita do Passa Quatro (36,82%), Ribeirão Corrente (33,40%), Santa Cruz da Esperança (29,69%) e Barretos (28,07%). Os resultados divulgados da produtividade histórica destes municípios são apresentados na Tabela 4.8.

Tabela 4.8 – Produtividade divulgada pelo IBGE, por ano, para alguns municípios da mesorregião de Ribeirão Preto.

Município	Ano									
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Altinópolis	65,0	65,0	69,5	75,0	75,0	75,0	70,0	80,0	78,0	87,0
Serra Azul	70,0	70,0	69,5	70,0	74,8	70,0	74,9	75,0	75,0	80,0
Santa Rita do Passa Quatro	90,0	80,0	84,4	85,0	85,0	80,0	80,0	80,0	85,0	60,0
Ribeirão Corrente	140,0	100,0	118,7	100,0	100,0	100,0	100,0	100,0	90,0	75,0
Santa Cruz da Esperança	70,0	70,0	69,5	75,0	74,9	75,0	80,0	80,0	56,0	56,0
Barretos	90,0	90,0	92,0	90,0	90,0	90,0	90,0	90,0	85,0	70,0

Os municípios de Santa Rita do Passa Quatro, Ribeirão Corrente, Santa Cruz da Esperança e Barretos, tiveram mudanças abruptas nos valores divulgados de produtividade entre os anos de 2010/2011 ou 2011/2012 e, conforme explicado anteriormente, essas nuances não são detectadas em sua totalidade pelo método apresentado. Com relação ao município de Altinópolis, o comportamento dos *proxies* para o ano estimado (2012) é apresentado na Figura 4.11.

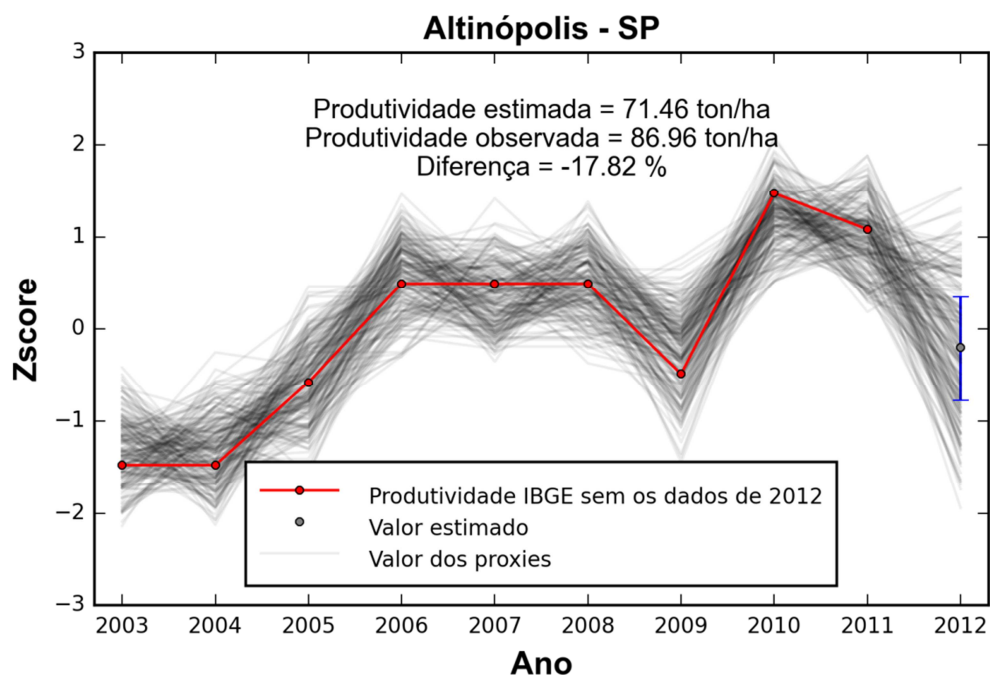


Figura 4.11 – Estimativa de produtividade de cana-de-açúcar para o ano de 2012 do município de Altinópolis – SP por meio do método proposto.

Para o caso apresentado na Figura 4.11 foi possível verificar que os *proxies* tiveram a tendência de subestimar a produtividade, uma vez que o valor de NDVI destes *proxies* foi menor para o ano estimado. Com o intuito de identificar a classe de cobertura destes *proxies*, a tabela contendo as informações por *proxy* (Apêndice A.3) foi aberta no software Qgis e o *plugin openLayers* foi utilizado para apresentar a camada *Google Sattelite*.

Na Figura 4.12 é possível verificar algumas classes de cobertura nas quais foram selecionados os *proxies*. O método de correlação apresentado nesta pesquisa, de fato, selecionou em sua maioria, células pertencentes às classes de cobertura vegetal. No entanto, para alguns municípios foram selecionadas células de área urbana e hidrografia.

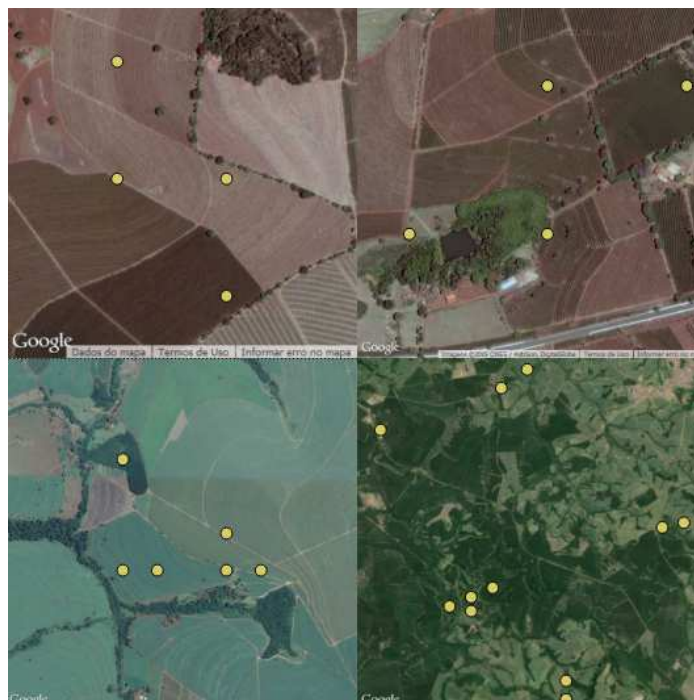


Figura 4.12 – Células selecionadas como *proxies* da produtividade do município de Altinópolis (pontos amarelos).

Como esta técnica utiliza-se da correlação entre os dados históricos de produtividade de uma região e os valores de NDVI, quando a produtividade histórica do município não sofreu variações, a melhor correlação entre os dados de produtividade e NDVI é mais provável nas células temporalmente estáveis da área de abrangência.

Os valores dos erros para cada um dos 361 municípios da área de estudo podem ser verificados no Apêndice A.2. A média dos resultados por municípios foi de 5,39% com $\sigma = 13,5$. No entanto, os maiores erros ultrapassaram 30% em relação ao valor observado pelo IBGE, sendo que 20 municípios tiveram erros superiores a 30% e três municípios tiveram erros inferiores a -30%.

A distribuição geográfica das diferenças pode ser verificada na Figura 4.13, em que é apresentado o mapa coroplético das diferenças observadas entre a produtividade estimada pelo método proposto e a produtividade divulgada pelo IBGE, para o estado de São Paulo.

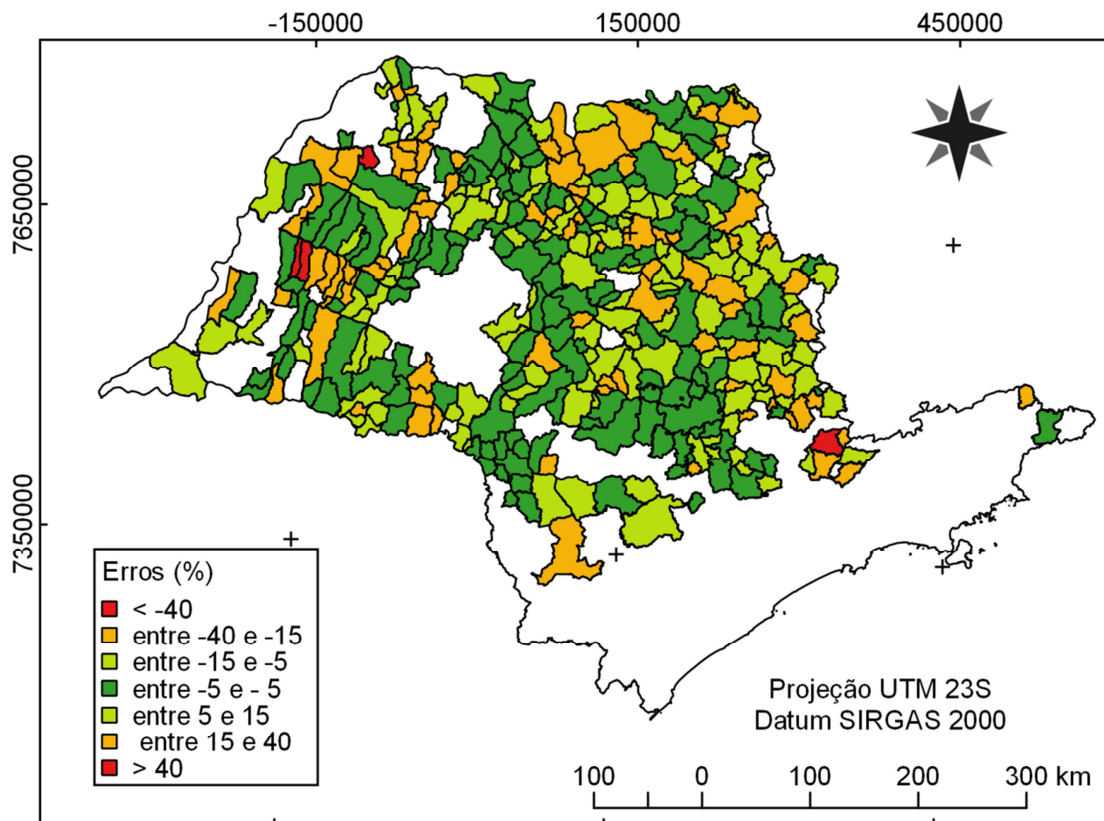


Figura 4.13 – Erros médios (%) entre os valores estimados e observados para os 361 municípios selecionados como área de estudo desta pesquisa.

O trabalho de Morel et al. (2014) mostra que modelos empíricos podem obter melhores resultados quando comparados com modelos agrometeorológicos. Ademais, o modelo empírico é o que necessita de um menor número de dados de entrada o que facilita o ajuste deste tipo de modelo para variadas regiões. Desse modo, o método proposto nesta pesquisa pode ser aplicado em diversos pares de cultura/região precisando de mínimos ajustes nos dados de entrada.

É importante lembrar que, variações na produtividade de alguns municípios não puderam ser explicadas, e por isso, a inserção de novas variáveis e restrições que auxiliem este método, são necessárias. Além disso, este método não substituirá os modelos convencionais de levantamento, uma vez que necessita dos dados destes levantamentos, como parâmetros de entrada. No entanto, poderá servir, futuramente, como uma maneira de obterem-se estimativas de produtividade para algumas culturas, antes que os resultados oficiais sejam divulgados.

5 CONCLUSÃO

O objetivo inicial deste trabalho foi obtido, uma vez que o método apresentado conseguiu estimar a produtividade da cana-de-açúcar para o estado de São Paulo. Além disso, nesta pesquisa, foi desenvolvido e implementado o algoritmo que permitirá a replicação e aprofundamento dos estudos sobre este método em outras áreas de estudo e para outras culturas.

A integração de diferentes produtos de NDVI, do sensor MODIS, disponibilizados por BOKU, não teve relevância neste trabalho. Talvez outros métodos, que não seja a média simples de diferentes produtos, como por exemplo, a integral abaixo da curva de NDVI, poderão obter melhores relações com a produtividade.

O LAG mostrou-se um parâmetro determinante na precisão dos resultados e, conforme esperado, os produtos NDVI de uma série temporal filtrada, possuem melhores resultados quanto mais distante da extremidade da série temporal. No caso do dado utilizado nesta pesquisa, a partir da terceira filtragem (produto B) o produto de NDVI atingiu os melhores valores para o método proposto.

Os resultados por grupos, separados pela quantidade de área colhida de cana-de-açúcar em 2012, não apresentaram diferenças à 5% de significância. Assim, não foram detectadas diferenças gerais, que pudessem ser atribuídas à quantidade de área colhida. A divisão, em um maior número de grupos, pode ser necessária para que essas diferenças possam ser detectadas.

As estimativas por mesorregião apresentaram bons resultados, em geral, com diferenças menores que 5%, no entanto o resultado de alguns municípios foi insatisfatório. Há necessidade de aplicar algumas restrições no algoritmo, de forma a detectar os possíveis casos em que o método proposto não seja suficiente para estimar a produtividade.

Finalmente, considerando que para cobrir a área de estudo, foram necessários 72 *minitiles* para cada semana e que para cada ano são 52 semanas, para os 10 anos analisados na presente pesquisa (2003 – 2012) foi necessário o

processamento de 37440 imagens. Dessa maneira, a automatização do processo foi fundamental para tornar o objetivo possível. Apesar de que, em trabalhos de sensoriamento remoto, não seja incomum a necessidade de processamento de grandes quantidades de dados, os avanços no poder de processamento dos computadores (principalmente em computadores pessoais) tornou possível analisar, e processar, grandes volumes de dados, como é o caso deste trabalho. Isso tem relação direta com a expansão, nos últimos anos, dos trabalhos que envolvem a análise de longas séries temporais.

Diante dos resultados apresentados neste trabalho, recomenda-se para trabalhos futuros:

- a) Que máscaras sejam criadas com o intuito isolar regiões de hidrografia e núcleos urbanos, para evitar a seleção de células com classes de coberturas indesejáveis;
- b) Nesta pesquisa foi investigado o uso apenas da variável espectral, assim, recomenda-se a integração dos *proxies* obtidos neste trabalho como dados de entrada em modelos agrometeorológicos;
- c) Estimar novos resultados com um número menor de *proxies*, pois se espera que assim, as estimativas acompanhem melhor as variabilidades da produtividade;
- d) Estimar a produtividade da cana-de-açúcar para outras regiões do país e para outras culturas. Assim, será possível avaliar o comportamento deste método, quando aplicado às mais diferentes combinações de cultivo e região.

REFERÊNCIAS BIBLIOGRÁFICAS

- ASNER, G. P.; HICKE, J. A.; LOBELL, D. B. Per-pixel analysis of forest structure. In: WULDER, M. A.; FRANKLIN, S. E. (Org.). **Remote sensing of forest environment: concepts and case studies**. California: Kluwer Academic Publishers, 2003. 209–254 p.
- ASSAD, E. D.; MARIN, F. R.; EVANGELISTA, S. R.; PILAU, F. G.; FARIAS, J. R. B.; PINTO, H. S.; JÚNIOR, J. Z. Sistema de previsão da safra de soja para o brasil. **Pesquisa agropecuária brasileira**, v. 42, n. 1, p. 615–625, 2007.
- ATKINSON, P. M.; JEGANATHAN, C.; DASH, J.; ATZBERGER, C. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. **Remote sensing of environment**, v. 123, p. 400–417, ago. 2012.
- ATZBERGER, C. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. **Remote sensing**, v. 5, n. 2, p. 949–981, 2013.
- ATZBERGER, C.; EILERS, P. H. C. A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of south america. **International journal of digital earth**, v. 4, n. 5, p. 365–386, 2011.
- BARET, F.; GUYOT, G.; MAJOR, D. Tsavi: a vegetation index which minimizes soil brightness effects on lai and apar estimation. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (**IGARSS'89**), 1989, Vancouver, Canada. **Proceedings...** . Vancouver: IEEE, 1989. p. 1355–1358.
- BECKER-RESHEF, I.; JUSTICE, C.; SULLIVAN, M.; VERMOTE, E.; TUCKER, C.; ANYAMBA, A.; SMALL, J.; PAK, E.; MASUOKA, E.; SCHMALTZ, J.; HANSEN, M.; PITTMAN, K.; BIRKETT, C.; WILLIAMS, D.; REYNOLDS, C.; DOORN, B. Monitoring global croplands with coarse resolution earth observations: the global agriculture monitoring (glam) project. **Remote sensing**, v. 2, n. 6, p. 1589–1609, 18 jun. 2010.
- BERLATO, M. A. **Modelo de relação entre o rendimento de grão de soja no rio grande do sul e disponibilidade hídrica**. 1987. Dissertação (Mestrado em Meteorologia) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. Sao Jose dos Campos. 1987.
- BERNARDES, T.; ADAMI, M.; FORMAGGIO, A. R.; MOREIRA, M. A.; FRANÇA, D. DE A.; NOVAES, M. R. Imagens mono e multitemporais modis para estimativa da área com soja no estado de mato grosso. **Pesquisa agropecuária brasileira**, v. 46, n. 1, p. 1530–1537, 2011.
- BROWN, J. C.; KASTENS, J. H.; COUTINHO, A. C.; VICTORIA, D. D. C.; BISHOP, C. R. Classifying multiyear agricultural land use data from mato

grosso using time-series modis vegetation index data. **Remote sensing of environment**, v. 130, p. 39–50, mar. 2013.

CARRÃO, H.; GONÇALVES, P.; CAETANO, M. Contribution of multispectral and multitemporal information from modis images to land cover classification. **Remote sensing of environment**, v. 112, n. 3, p. 986–997, mar. 2008.

CARREIRAS, J. M. B.; PEREIRA, J. M. C.; SHIMABUKURO, Y. E.; STROPPIANA, D. Evaluation of compositing algorithms over the brazilian amazon using spot-4 vegetation data. **International journal of remote sensing**, v. 24, n. 17, p. 3427–3440, 2003.

CHEN, J.; JÖNSSON, P.; TAMURA, M.; GU, Z.; MATSUSHITA, B.; EKLUNDH, L. A simple method for reconstructing a high-quality ndvi time-series data set based on the savitzky-golay filter. **Remote sensing of environment**, v. 91, n. 3-4, p. 332–344, 2004.

CRAMÉR, H. **Mathematical methods of statistics**. 19. ed. Princeton: University Press, 1946.

DEFRIES, R. S.; TOWNSHEND, J. R. G. Global land cover characterization from satellite data: from research to operational implementation? **Global ecology and biogeography**, v. 8, p. 367–379, 1999.

DOORENBOS, J.; KASSAM, A. **Yield response to water**. Rome: FAO-Plant production and protection, 1979. Disponível em: <<http://www.cabdirect.org/abstracts/19796730583.html>>. Acesso em: 6 fev. 2014.

EILERS, P. H. C. A perfect smoother. **Analytical chemistry**, v. 75, n. 14, p. 3631–6, 15 jul. 2003.

GAO, B. C. Ndwí—a normalized difference water index for remote sensing of vegetation liquid water from space. **Remote sensing of environment**, From Duplicate 2 (NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space - Gao, Bo-cai), v. 58, n. 3, p. 257–266, dez. 1996.

GDAL DEVELOPMENT TEAM. **Gdal - geospatial data abstraction library, version 1.11.2**. [S.l.: s.n.] . 2015. Disponível em: <<http://www.gdal.org>>.

GOEL, N. S. Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. **Remote sensing reviews**, v. 4, n. 1, p. 1–212, jan. 1988.

HIRD, J. N.; MCDERMID, G. J. Noise reduction of ndvi time series: an empirical comparison of selected techniques. **Remote sensing of environment**, v. 113, n. 1, p. 248–258, 2009.

HUETE, A. R. A soil-adjusted vegetation index (savi). **Remote sensing of environment**, v. 25, p. 295–309, 1988.

HUETE, A. R.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P. Overview of the radiometric and biophysical performance of the modis vegetation indices. **Remote sensing of environment**, v. 83, p. 195–213, 2002.

HUNTER, J. D. Matplotlib: a 2d graphics environment. **Computing in science and engineering**, v. 9, n. 3, p. 99–104, 2007.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Pesquisas agropecuárias**. 2. ed. Rio de Janeiro: [S.N.], 2002. v. 6. . Disponível em: <ftp://ftp.ibge.gov.br/Producao_Agricola/Metodologia_da_pesquisa/PesquisasAgropecuarias2002.pdf>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Divisão territorial brasileira**. Disponível em: <ftp://geoftp.ibge.gov.br/organizacao_territorial/divisao_territorial/2014/>. Acesso em: 15 jan. 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Produção agrícola municipal 2012**. Disponível em: <<http://sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&c=1612>>. Acesso em: 15 jan. 2014.

KASTENS, J.; KASTENS, T.; KASTENS, D.; PRICE, K.; MARTINKO, E.; LEE, R. Image masking for crop yield forecasting using avhrr ndvi time series imagery. **Remote sensing of environment**, v. 99, n. 3, p. 341–356, 2005.

KLISCH, A. **Time series data of a given pixel**. 2014. [mensagem pessoal]. Mensagem recebida por <wfoschiera@dsr.inpe.br> em: 16 dez. 2014.

LI, Z.; LI, X.; WEI, D.; XU, X.; WANG, H. An assessment of correlation on modis-ndvi and evi with natural vegetation coverage in northern hebei province, china. **Procedia environmental sciences**, v. 2, p. 964–969, jan. 2010.

LIU, J.; PATTEY, E.; MILLER, J. R.; MCNAIRN, H.; SMITH, A.; HU, B. Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. **Remote sensing of environment**, v. 114, n. 6, p. 1167–1177, jun. 2010.

MAHLEIN, A.-K.; RUMPF, T.; WELKE, P.; DEHNE, H.-W.; PLÜMER, L.; STEINER, U.; OERKE, E.-C. Development of spectral indices for detecting and identifying plant diseases. **Remote sensing of environment**, v. 128, p. 21–30, jan. 2013.

MCKINNEY, W. Data structures for statistical computing in python. In: PROCEEDINGS OF THE 9TH PYTHON IN SCIENCE CONFERENCE. 2010, Austin, USA. **Proceedings...** Austin, USA: [S.N.], p. 51–56, 2010.

MELLO, M. P.; ATZBERGER, C.; FORMAGGIO, A. R. Near real time yield estimation for sugarcane in brazil combining remote sensing and official statistical data. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS 2014), 34., 2014, Quebec, Canada. **Proceedings...** Quebec, Canada: IEEE, 2014. p. 10–13.

MOREL, J.; TODOROFF, P.; BÉGUÉ, A.; BURY, A.; MARTINÉ, J.; PETIT, M. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: a case study on reunion island. **Remote sensing**, v. 6, n. 7, p. 6620–6635, 2014.

NOGUEIRA, S. M. C. **Aplicação de um modelo agrometeorológico-espectral e de variáveis meteorológicas do modelo ETA para estimar a produtividade do trigo**. 2014. 105 p. (sid.inpe.br/mtc-m21b/2014/05.22.18.12-TDI). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014. Disponível em: <<http://urlib.net/8JMKD3MGP5W34M/3GBRAHL>>. Acesso em: 10 jul. 2015.

OLIPHANT, T. E. Python for scientific computing. **Computing in science and engineering**, v. 9, n. 3, p. 10–20, 2007.

PEDRO JUNIOR, M. J.; CAMARGO, M. B. P.; MIRANDA, M. A. C.; BRUNINI, O.; ORTOLANI, A. A.; ALFONSI, R. R. Teste de modelo agrometeorológico para estimativa da produtividade da cultura da soja de ciclo precoce. In: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 3., 1983, Campinas. **Anais...** Campinas: Sociedade Brasileira de Agrometeorologia, 1983. p. 11–17. (verificar a data deste Congresso)

PENUELAS, J.; FILELLA, I.; BIEL, C.; SERRANO, L.; SAVÉ, R. The reflectance at the 950–970 nm region as an indicator of plant water status. **International journal of remote sensing**, v. 14, n. 10, p. 1887–1905, jul. 1993.

PICOLI, M. C. A.; RUDORFF, B. F. T.; RIZZI, R.; GIAROLLA, A. Índice de vegetação do sensor modis na estimativa da produtividade agrícola da cana-de-açúcar. **Bragantia**, v. 68, n. 3, p. 789–795, 2009.

QGIS DEVELOPMENT TEAM. **Qgis geographic information system**. [S.l.: s.n.]. 2009. Disponível em: <<http://qgis.osgeo.org>>.

RASMUSSEN, M. S. Assessment of millet yields and production in northern burkina faso using integrated ndvi from the avhrr. **International journal of remote sensing**, v. 13, n. 18, p. 3431–3442, 1992.

RICE, J. A. **Mathematical statistics and data analysis**. 3. ed. [S.l.]: Duxbury Press, 2001. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0534399428>>.

RIZZI, R.; RISSO, J.; EPIPHANIO, R. D. V.; RUDORFF, B. F. T.; FORMAGGIO, A. R.; SHIMABUKURO, Y. E.; FERNANDES, S. L. Estimativa da área de soja no mato grosso por meio de imagens modis. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, , 14. (SBSR), 2009, Natal. **Anais...**São José dos Campos: INPE, 2009. p. 387-394. DVD, On-line. ISBN 978-85-17-00044-7. (INPE-16002-PRE/10611). Disponível em:<<http://urlib.net/dpi.inpe.br/sbsr@80/2008/11.16.18.50.57>>. Acesso em: 24 jun. 2015.

ROSSUM, G. VAN; DRAKE, F. L. **Python tutorial**. Amsterdan, 2010. Disponível em: <<http://docs.python.org/tutorial/>>.

ROUGHGARDEN, J.; RUNNING, S.; MATSON, P. A. What does remote sensing do for ecology? **Ecology**, v. 72, n. 6, p. 1918–1922, 1991.

ROUSE, J.; HAAS, R.; SCHELL, J. **Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation**. Maryland: [S.N.], 1973. Disponível em: <<http://library.wur.nl/WebQuery/clc/154154>>. Acesso em: 4 fev. 2014.

RUDORFF, B. F. T. **Dados landsat na estimativa da produtividade agrícola da cana-de-açúcar**. 1985. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1985.

RUDORFF, B. F. T.; BATISTA, G. Yield estimation of sugarcane based on agrometeorological-spectral models. **Remote sensing of environment**, v. 33, n. 3, p. 183–192, 1990.

RUDORFF, B. F. T.; DE AGUIAR, D. A.; DA SILVA, W. F.; SUGAWARA, L. M.; ADAMI, M.; MOREIRA, M. A. Studies on the rapid expansion of sugarcane for ethanol production in são paulo state (brazil) using landsat data. **Remote sensing**, v. 2, p. 1057–1076, 2010.

RUDORFF, B. F. T.; SUGAWARA, L. M. Mapeamento da cana-de-açúcar na região centro-sul via imagens de satélites. **Informe agropecuário**, v. 28, n. 241, p. 79–86, 2007.

SAKAMOTO, T.; GITELSON, A. A.; ARKEBAUER, T. J. Near real-time prediction of u.s. corn yields based on time-series modis data. **Remote sensing of environment**, v. 147, p. 219–231, maio 2014.

SANCHES, I. D.; EPIPHANIO, J. C. N.; FORMAGGIO, A. R. Culturas agrícolas em imagens multitemporais do satélite landsat. **Agric. São Paulo**, v. 52, p. 83–96, 2005.

SCHOWENGERDT, R. A. **Remote sensing** - models and methods for image processing. 3. ed. Burlington: Academic Press, 1997. 522 p. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780123694072500012>>.

SEAN GILLIES. **The shapely user manual, version 1.3**. [S.l: s.n.] . 2013. Disponível em: <<http://toblerity.org/shapely/manual.html>>.

SUGAWARA, L. M. **Avaliação de modelo agrometeorológico e imagens noaa/avhrr no acompanhamento e estimativa de produtividade de soja no estado do paraná**. 2002. 181 p. (INPE-8702-TDI/794). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2001. Disponível em: <<http://urlib.net/dpi.inpe.br/lise/2002/11.18.18.05>>. Acesso em: 24 jan. 2015.

SUGAWARA, L. M. **Variação interanual da produtividade agrícola da cana-de-açúcar por meio de um modelo agrônômico**. 2010. 114 p. (sid.inpe.br/mtc-m19/2010/10.08.14.39-TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010. Disponível em: <<http://urlib.net/8JMKD3MGP7W/38CSKHB>>. Acesso em: 24 jun. 2015. SUGAWARA, L. M.; RUDORFF, B. F. T.; ADAMI, M. Viabilidade de uso de imagens do landsat em mapeamento de área cultivada com soja no estado do paraná. **Pesquisa agropecuária brasileira**, v. 43, n. 12, p. 1777–1783, 2008.

TEILLET, P. M.; STAENZ, K.; WILLIAM, D.; STAENX, K.; WILLIAMS, D. J. Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. **Remote sensing of environment**, v. 61, p. 139–149, 1997.

TOWNSHEND, J. R. .; JUSTICE, C. O. Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. **Remote sensing of environment**, v. 83, n. 1-2, p. 351–359, nov. 2002.

TUCKER, C. J.; HOLBEN, B. N.; ELGIN, J. H. Relationship of spectral data to grain yield variation. **Photogrammetric engineering and remote sensing**, v. 45, p. 657–666, 1980.

VAN DER WALT, S.; COLBERT, S. C.; VAROQUAUX, G. The numpy array: a structure for efficient numerical computation. **Computing in science and engineering**, v. 13, n. 2, p. 22–30, 2011.

VERMOTE, E. F.; EL SALEOUS, N. Z.; JUSTICE, C. O. Atmospheric correction of modis data in the visible to middle infrared: first results. **Remote sensing of environment**, v. 83, n. 1-2, p. 97–111, nov. 2002.

VUOLO, F.; MATTIUZZI, M. Data service platform for modis vegetation indices time series processing at boku vienna: current status and future perspectives. Proceedings of SPIE. v. 8538, 2012, Proc. of SPIE, v. 8538 85380A-1 doi: 10.1117/12.974857. Available at: <http://proceedings.spiedigitallibrary.org/> on 12/17/2012 Terms of Use: <http://spiedigitallibrary.org/terms>.

WARDLOW, B.; EGBERT, S. Large-area crop mapping using time-series modis 250 m ndvi data: an assessment for the us central great plains. **Remote sensing of environment**, v. 112, p. 1096–1116, 2008.

WIEGAND, C. L.; RICHARDSON, A. J.; KANEMASU, E. T. Leaf area index estimates for wheat from landsat and their implications for evapotranspiration and crop modeling¹. **Agronomy journal**, v. 71, n. 2, p. 336, 1979.

APÊNDICE A – GRÁFICOS E TABELAS AUXILIARES PARA INTERPRETAÇÃO DOS RESULTADOS

Tabela A.1 Tipos de produtos utilizados no processamento de acordo com as diferentes combinações dos valores dos parâmetros de entrada LAG e T.I.

T.I.	LAG 0	LAG 1
1	N	A
2	N + A	A + B
1	N + A + B + C	A + B + C + D1
3	N + A + B + C + D1 + D2	A + B + C + D1 + D2 + D3
5	N + A + B + C + D1 + D2 + D3 + D4	A + B + C + D1 + D2 + D3 + D4 + D5
7	N + A + B + C + D1 + D2 + D3 + D4 + D5 + D6	A + B + C + D1 + D2 + D3 + D4 + D5 + D6 + D7

T.I.	LAG 2	LAG 3
1	B	C
2	B + C	C + D1
1	B + C + D1 + D2	C + D1 + D2 + D3
3	B + C + D1 + D2 + D3 + D4	C + D1 + D2 + D3 + D4 + D5
5	B + C + D1 + D2 + D3 + D4 + D5 + D6	C + D1 + D2 + D3 + D4 + D5 + D6 + D7
7	B + C + D1 + D2 + D3 + D4 + D5 + D6 + D7 + D8	C + D1 + D2 + D3 + D4 + D5 + D6 + D7 + D8 + D9

T.I.	LAG 1
1	D
2	D1 + D2
1	D1 + D2 + D3 + D4
3	D1 + D2 + D3 + D4 + D5 + D6
5	D1 + D2 + D3 + D4 + D5 + D6 + D7 + D8
7	D1 + D2 + D3 + D4 + D5 + D6 + D7 + D8 + D9 + D10

Tabela A.2 – Erros por município e grupo para o ano de 2012.

Cód. Mun.	Erro (%)	Grupo	Cód. Mun.	Erro (%)	Grupo
3507605	-46,98	Grupo 1	3533304	2,78	Grupo 1
3556354	-38,2	Grupo 1	3512308	2,82	Grupo 1
3504107	-34,64	Grupo 1	3520509	3,45	Grupo 1
3532405	-31,01	Grupo 1	3541109	5,76	Grupo 1
3541901	-27,39	Grupo 1	3518859	9,89	Grupo 1
3555307	-23,63	Grupo 1	3540853	16,23	Grupo 1
3537107	-23,61	Grupo 1	3506003	17,42	Grupo 1
3538600	-14,51	Grupo 1	3518008	17,57	Grupo 1
3554755	-10,82	Grupo 1	3519055	17,9	Grupo 1
3512209	-10,36	Grupo 1	3510906	19,83	Grupo 1
3545209	-10,35	Grupo 1	3551603	26,28	Grupo 1
3508405	-9,69	Grupo 1	3511508	-18,84	Grupo 2
3539400	-8,48	Grupo 1	3549706	-15,78	Grupo 2
3508702	-8,39	Grupo 1	3522406	-15,67	Grupo 2
3525201	-8,27	Grupo 1	3514304	-12,29	Grupo 2
3552106	-7,62	Grupo 1	3528858	-12,26	Grupo 2
3535903	-7,59	Grupo 1	3515194	-11,88	Grupo 2
3542404	-7,46	Grupo 1	3523800	-11,21	Grupo 2
3515152	-6,53	Grupo 1	3543600	-8,79	Grupo 2
3536570	-6,36	Grupo 1	3505807	-8,31	Grupo 2
3557303	-5,19	Grupo 1	3556909	-8,02	Grupo 2
3535804	-5	Grupo 1	3510153	-7,24	Grupo 2
3514502	-4,79	Grupo 1	3550506	-7,01	Grupo 2
3549607	-4,11	Grupo 1	3536901	-6,84	Grupo 2
3524006	-3,22	Grupo 1	3515608	-6,52	Grupo 2
3502606	-3,11	Grupo 1	3556404	-5,82	Grupo 2
3548302	-2,78	Grupo 1	3542107	-5,27	Grupo 2
3551207	-1,28	Grupo 1	3546702	-5,18	Grupo 2
3553005	-0,91	Grupo 1	3519808	-5,07	Grupo 2
3554201	-0,56	Grupo 1	3547007	-4,73	Grupo 2
3556958	-0,54	Grupo 1	3545159	-4,32	Grupo 2
3538808	-0,5	Grupo 1	3544707	-4,24	Grupo 2
3533809	-0,35	Grupo 1	3504305	-3,41	Grupo 2
3508306	-0,12	Grupo 1	3530300	-2,43	Grupo 2
3502200	0,13	Grupo 1	3509809	-2,38	Grupo 2
3527009	0,25	Grupo 1	3532900	-2,38	Grupo 2
3554607	0,8	Grupo 1	3529500	-2,28	Grupo 2
3528601	0,97	Grupo 1	3523602	-1,68	Grupo 2
3525854	1,26	Grupo 1	3543808	-1,46	Grupo 2
3523909	1,27	Grupo 1	3511607	-1,28	Grupo 2
3510302	2,6	Grupo 1	3526407	-1,24	Grupo 2

(Continua...)

Cód. Mun.	Erro (%)	Grupo	Cód. Mun.	Erro (%)	Grupo
3530409	-1,18	Grupo 2	3553658	9,78	Grupo 2
3506300	-1,06	Grupo 2	3540259	10,14	Grupo 2
3553807	-0,92	Grupo 2	3546207	10,4	Grupo 2
3503406	-0,71	Grupo 2	3548005	11,39	Grupo 2
3534757	-0,71	Grupo 2	3515202	11,76	Grupo 2
3527702	-0,6	Grupo 2	3548401	11,78	Grupo 2
3515301	-0,44	Grupo 2	3525409	11,91	Grupo 2
3553609	0,76	Grupo 2	3547700	12,23	Grupo 2
3502408	1,04	Grupo 2	3511904	12,36	Grupo 2
3508900	1,16	Grupo 2	3512704	12,65	Grupo 2
3507704	1,28	Grupo 2	3551900	13,87	Grupo 2
3504602	1,46	Grupo 2	3509304	16,23	Grupo 2
3507456	1,56	Grupo 2	3530805	16,48	Grupo 2
3507308	2,18	Grupo 2	3514700	17,36	Grupo 2
3535705	2,38	Grupo 2	3520004	17,61	Grupo 2
3551504	2,5	Grupo 2	3535101	18,58	Grupo 2
3524709	2,61	Grupo 2	3540705	18,67	Grupo 2
3521101	2,75	Grupo 2	3510104	19,53	Grupo 2
3503356	2,81	Grupo 2	3516507	19,66	Grupo 2
3515806	2,94	Grupo 2	3529609	21,08	Grupo 2
3539905	3,25	Grupo 2	3508108	21,1	Grupo 2
3541406	3,57	Grupo 2	3514957	21,76	Grupo 2
3553104	4,42	Grupo 2	3522000	22,95	Grupo 2
3542503	4,45	Grupo 2	3549250	23,37	Grupo 2
3546504	4,61	Grupo 2	3539707	23,55	Grupo 2
3549805	4,69	Grupo 2	3537701	24,31	Grupo 2
3503802	5,23	Grupo 2	3513306	24,62	Grupo 2
3532702	5,49	Grupo 2	3515129	26,43	Grupo 2
3555703	5,59	Grupo 2	3545100	27,1	Grupo 2
3536000	5,82	Grupo 2	3537008	27,62	Grupo 2
3522604	5,88	Grupo 2	3546256	29,69	Grupo 2
3543204	6,6	Grupo 2	3519501	30,09	Grupo 2
3537156	6,62	Grupo 2	3534609	30,9	Grupo 2
3518503	7,06	Grupo 2	3503109	31,07	Grupo 2
3511300	7,14	Grupo 2	3501905	31,43	Grupo 2
3515186	7,6	Grupo 2	3527405	32,71	Grupo 2
3553203	8,16	Grupo 2	3532868	33,07	Grupo 2
3514924	9,13	Grupo 2	3543105	33,4	Grupo 2
3521150	9,2	Grupo 2	3502002	34,51	Grupo 2
3520806	9,6	Grupo 2	3527256	37,32	Grupo 2

(Continua...)

Cód. Mun.	Erro (%)	Grupo	Cód. Mun.	Erro (%)	Grupo
3551306	37,6	Grupo 2	3502309	-0,38	Grupo 3
3501608	38,68	Grupo 2	3527900	-0,21	Grupo 3
3521606	49,67	Grupo 2	3544400	-0,2	Grupo 3
3518909	73,61	Grupo 2	3500709	-0,06	Grupo 3
3501004	-17,89	Grupo 3	3501707	0,03	Grupo 3
3512407	-15,88	Grupo 3	3541059	0,43	Grupo 3
3521507	-12,29	Grupo 3	3507506	0,55	Grupo 3
3551405	-10,36	Grupo 3	3550407	0,91	Grupo 3
3549508	-9,92	Grupo 3	3503703	1,14	Grupo 3
3519303	-9,29	Grupo 3	3534005	1,31	Grupo 3
3526902	-8,72	Grupo 3	3554409	1,63	Grupo 3
3507803	-6,68	Grupo 3	3512506	1,78	Grupo 3
3516853	-6,48	Grupo 3	3541505	2,02	Grupo 3
3540408	-6,44	Grupo 3	3533007	2,1	Grupo 3
3531308	-5,98	Grupo 3	3539004	3,35	Grupo 3
3538907	-5,93	Grupo 3	3540804	4,04	Grupo 3
3517505	-5,5	Grupo 3	3511409	4,12	Grupo 3
3545803	-5,47	Grupo 3	3544004	4,18	Grupo 3
3532157	-4,42	Grupo 3	3543907	4,43	Grupo 3
3505104	-4,35	Grupo 3	3521408	4,63	Grupo 3
3532603	-4,28	Grupo 3	3504008	4,69	Grupo 3
3517109	-4,09	Grupo 3	3532504	5,26	Grupo 3
3542909	-3,92	Grupo 3	3514007	5,44	Grupo 3
3557204	-3,91	Grupo 3	3530904	5,58	Grupo 3
3552601	-3,55	Grupo 3	3506201	5,97	Grupo 3
3540606	-3,16	Grupo 3	3540903	5,99	Grupo 3
3544806	-2,84	Grupo 3	3556800	6,52	Grupo 3
3552700	-2,71	Grupo 3	3521804	7,29	Grupo 3
3511706	-2,59	Grupo 3	3547601	7,48	Grupo 3
3530102	-2,22	Grupo 3	3546900	7,67	Grupo 3
3519105	-2,12	Grupo 3	3532801	8,63	Grupo 3
3554706	-1,82	Grupo 3	3519204	8,73	Grupo 3
3500303	-1,79	Grupo 3	3554508	9,11	Grupo 3
3519907	-1,28	Grupo 3	3525607	9,23	Grupo 3
3520905	-0,93	Grupo 3	3544202	9,34	Grupo 3
3504800	-0,83	Grupo 3	3503000	9,64	Grupo 3
3555604	-0,79	Grupo 3	3530706	9,72	Grupo 3
3556008	-0,57	Grupo 3	3503604	9,83	Grupo 3
3555356	-0,5	Grupo 3	3516101	10	Grupo 3
3542701	-0,46	Grupo 3	3536307	10,14	Grupo 3

(Continua...)

Cód. Mun.	Erro (%)	Grupo	Cód. Mun.	Erro (%)	Grupo
3534302	10,26	Grupo 3	3545605	-5,67	Grupo 4
3534708	10,26	Grupo 3	3514106	-5,15	Grupo 4
3511102	10,35	Grupo 3	3542206	-4,84	Grupo 4
3515509	10,64	Grupo 3	3507407	-4,14	Grupo 4
3528809	10,65	Grupo 3	3519600	-3,85	Grupo 4
3528007	11,06	Grupo 3	3517703	-2,71	Grupo 4
3533254	11,44	Grupo 3	3510807	-2,63	Grupo 4
3522307	12,09	Grupo 3	3546306	-1,93	Grupo 4
3539608	12,35	Grupo 3	3510005	-1,6	Grupo 4
3531506	12,57	Grupo 3	3534203	-1,6	Grupo 4
3514908	12,79	Grupo 3	3548906	-1,55	Grupo 4
3528700	13,97	Grupo 3	3526506	-1,5	Grupo 4
3522901	14,54	Grupo 3	3513108	-1,41	Grupo 4
3528106	15,49	Grupo 3	3510401	-1,34	Grupo 4
3519402	16,01	Grupo 3	3536703	-1,28	Grupo 4
3505302	17,61	Grupo 3	3520103	-1,07	Grupo 4
3538105	18,22	Grupo 3	3521903	-1,03	Grupo 4
3514601	18,59	Grupo 3	3548054	-0,79	Grupo 4
3505609	18,63	Grupo 3	3553401	-0,73	Grupo 4
3508207	18,84	Grupo 3	3537305	-0,4	Grupo 4
3549102	19,12	Grupo 3	3502101	-0,19	Grupo 4
3529807	19,31	Grupo 3	3535002	0,23	Grupo 4
3552304	20,1	Grupo 3	3550100	0,28	Grupo 4
3509106	20,57	Grupo 3	3526001	0,44	Grupo 4
3511201	20,57	Grupo 3	3536604	0,44	Grupo 4
3526704	23,3	Grupo 3	3539301	0,49	Grupo 4
3532058	23,71	Grupo 3	3551702	1,01	Grupo 4
3506508	25,97	Grupo 3	3553708	1,01	Grupo 4
3516903	33,29	Grupo 3	3512001	1,39	Grupo 4
3528304	34,33	Grupo 3	3539202	1,86	Grupo 4
3500105	34,89	Grupo 3	3546405	2,82	Grupo 4
3547502	36,82	Grupo 3	3543402	3,08	Grupo 4
3534906	45,65	Grupo 3	3518602	3,31	Grupo 4
3500907	-16,13	Grupo 4	3504404	3,44	Grupo 4
3530508	-11,45	Grupo 4	3525102	3,67	Grupo 4
3541703	-9,78	Grupo 4	3531407	4,1	Grupo 4
3525706	-9,49	Grupo 4	3540200	5,96	Grupo 4
3506706	-7,67	Grupo 4	3554300	6,06	Grupo 4
3511003	-6,64	Grupo 4	3505203	6,57	Grupo 4
3544905	-5,86	Grupo 4	3543709	7,8	Grupo 4

(Continua...)

Cód. Mun.	Erro (%)	Grupo
3553955	7,95	Grupo 4
3507902	9,02	Grupo 4
3539509	10,27	Grupo 4
3506102	10,63	Grupo 4
3502804	10,81	Grupo 4
3553302	11,29	Grupo 4
3550902	11,37	Grupo 4
3521309	11,39	Grupo 4
3513702	12,68	Grupo 4
3509403	13,64	Grupo 4
3503307	13,81	Grupo 4
3512100	14,3	Grupo 4
3527603	15,44	Grupo 4
3549409	17,25	Grupo 4
3517901	17,84	Grupo 4
3524204	18,56	Grupo 4
3503208	20,65	Grupo 4
3516002	21,79	Grupo 4
3537404	24,17	Grupo 4
3532207	24,62	Grupo 4
3529203	26,6	Grupo 4
3517802	28,62	Grupo 4
3535309	33,06	Grupo 4
3522703	-5,9	Grupo 5
3526803	-5,05	Grupo 5
3533502	-3,98	Grupo 5
3535507	-3,76	Grupo 5
3505906	-0,14	Grupo 5
3538709	0,46	Grupo 5
3524105	0,59	Grupo 5
3518206	1,08	Grupo 5
3556305	1,47	Grupo 5
3531902	2,29	Grupo 5
3529708	3,7	Grupo 5
3525300	11,16	Grupo 5
3517406	19,83	Grupo 5
3524303	20,04	Grupo 5
3533908	26,13	Grupo 5
3505500	28,07	Grupo 5

Tabela A.3 – Exemplo de armazenamento dos *proxies* em formato tabular. Cada linha representa um *proxy* do município de Adamantina (Código 3500105) para o ano 2012 e DOYi 84. As colunas b1, ..., b10, representam os valores escalonados das 10 bandas do produto NDVI.

						Valores normalizados de cada banda do produto NDVI										
Código		Ano		Valor												
IBGE	Latitude	Longitude	DOY	estimado	RMSE	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	
3500105	-21,67969	-51,31138	84	2012	0,29	-0,58	-0,10	-0,10	-0,26	2,16	0,71	0,38	0,22	-0,91	-1,55	
3500105	-20,84487	-50,78906	84	2012	0,29	0,28	0,28	0,02	-0,24	1,60	1,16	0,28	-0,33	-1,03	-1,90	
3500105	-20,92746	-51,22656	84	2012	0,31	0,96	0,08	-0,14	-1,01	1,83	0,74	-0,14	-0,57	-0,14	-1,67	
3500105	-22,15513	-50,80915	84	2012	0,33	1,39	-0,32	-0,61	-0,34	1,42	0,62	0,27	0,03	-0,53	-1,92	
3500105	-20,81808	-50,79129	84	2012	0,33	0,98	-0,04	0,01	-0,52	1,14	1,10	-0,24	-0,16	0,01	-2,27	
3500105	-21,53906	-51,52344	84	2012	0,34	0,93	0,45	0,12	-0,85	1,10	0,93	0,12	-0,04	-0,69	-2,15	
3500105	-22,04799	-50,93862	84	2012	0,35	0,00	-0,41	0,21	0,10	1,54	0,72	0,00	0,41	-0,31	-2,37	
3500105	-20,89174	-51,03237	84	2012	0,35	-0,06	0,06	-0,63	-0,51	2,23	1,08	-0,29	0,06	-0,63	-1,31	
3500105	-21,69531	-50,26004	84	2012	0,36	0,30	-0,20	0,05	-0,33	1,42	1,67	0,05	-0,33	-0,95	-1,71	
3500105	-21,45201	-50,72210	84	2012	0,37	0,50	0,06	-0,39	-0,17	1,85	0,95	-0,39	-0,39	0,28	-1,96	
3500105	-22,34040	-51,00558	84	2012	0,37	-0,12	-0,12	-0,42	-0,73	1,74	1,12	0,81	0,19	-0,73	-1,66	
3500105	-21,17522	-50,59933	84	2012	0,37	1,12	0,46	0,04	-0,04	1,54	0,29	-0,04	-0,46	-0,87	-2,03	
3500105	-22,32031	-51,02121	84	2012	0,38	-0,06	0,43	-0,06	0,10	1,41	1,25	0,10	-0,55	-0,39	-2,19	
3500105	-20,78683	-51,16183	84	2012	0,38	-0,26	0,26	-0,49	-0,87	2,23	1,02	-0,19	-0,26	-0,11	-1,32	