



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2014/03.30.15.33-TDI

**JANELA DE REGRESSÃO: UMA ANÁLISE ESPACIAL
E TEMPORAL PARA ESTIMAR VALORES DE NDVI
CLASSIFICADOS COM BAIXA QUALIDADE EM
SÉRIES TEMPORAIS MODIS**

Júlio César de Oliveira

Tese de Doutorado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelo Dr. José Carlos Neves Epiphanyo, aprovada em 31 de março de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3G3872B>>

INPE
São José dos Campos
2014

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

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):

Presidente:

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

Membros:

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

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

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (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)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

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

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

EDITORAÇÃO ELETRÔNICA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

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



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2014/03.30.15.33-TDI

**JANELA DE REGRESSÃO: UMA ANÁLISE ESPACIAL
E TEMPORAL PARA ESTIMAR VALORES DE NDVI
CLASSIFICADOS COM BAIXA QUALIDADE EM
SÉRIES TEMPORAIS MODIS**

Júlio César de Oliveira

Tese de Doutorado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelo Dr. José Carlos Neves Epiphânio, aprovada em 31 de março de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3G3872B>>

INPE
São José dos Campos
2014

Dados Internacionais de Catalogação na Publicação (CIP)

Oliveira, Júlio César.

Ol4j Janela de Regressão: uma análise espacial e temporal para estimar valores de NDVI classificados com baixa qualidade em séries temporais MODIS / Júlio César de Oliveira. – São José dos Campos : INPE, 2014.

xxiv + 93 p. ; (sid.inpe.br/mtc-m21b/2014/03.30.15.33-TDI)

Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

Orientador : Dr. José Carlos Neves Epiphânio.

1. Qualidade do dado. 2. Janela espaço temporal. 3. Redução de ruídos. 4. MODIS. 5. Série temporal. I.Título.

CDU 528.88



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

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

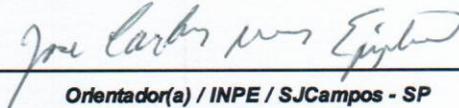
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Doutor(a)** em
Sensoriamento Remoto

Dr. Camilo Daleles Rennó



Presidente / INPE / SJC Campos - SP

Dr. José Carlos Neves Epiphânio



Orientador(a) / INPE / SJC Campos - SP

Dr. Laércio Massaru Namikawa



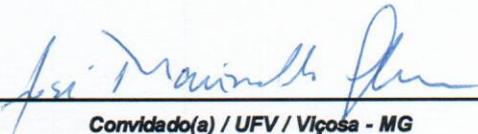
Membro da Banca / INPE / São José dos Campos - SP

Dr. Ramon Moraes de Freitas



Convidado(a) / Camargo-Schubert Eng / Curitiba - PR

Dr. José Marinaldo Gleriani



Convidado(a) / UFV / Viçosa - MG

Este trabalho foi aprovado por:

maioria simples

unanimidade

Aluno (a): **Júlio César de Oliveira**

São José dos Campos, 31 de Março de 2014

“Sucesso é a soma de pequenos esforços repetidos o tempo todo”.

Robert Collier

A meus pais Laerte de Oliveira (in memoriam) e Maria S. Ribeiro.

A meus irmãos.

À Fernanda e ao Davi pela fonte de inspiração.

AGRADECIMENTOS

Agradeço à Universidade Federal de Viçosa (UFV) pela oportunidade e apoio para a realização deste trabalho.

Ao Departamento de Engenharia Civil da UFV, principalmente aos colegas do setor de Engenharia de Agrimensura, que assumiram compromissos durante o meu afastamento. À Margarida e Suely, funcionárias do setor de capacitação e treinamento da UFV, pela atenção e colaboração neste período.

Ao Dr. José Carlos Neves Epiphanyo, por este período de crescimento profissional, pela exemplar orientação oferecida e por todo o apoio e conselhos que enriqueceram este documento. Além disso, pela amizade e admiração pessoal e profissional que nasceu durante esses anos de trabalho.

Ao Programa de Pós-Graduação em Sensoriamento Remoto do Instituto Nacional de Pesquisas Espaciais (INPE), pela oportunidade de adquirir tamanho conhecimento, bem como pelo apoio e infraestrutura oferecidos para a realização desta pesquisa.

À Fundação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo auxílio financeiro.

Aos colegas Kleber, Tiago, Juliana, Gabriel, Francielle, André, aos membros da turma PGSER2010 e tantos outros que durante todo este período participaram de alguma forma, seja auxiliando neste trabalho ou nas horas de lazer. Em especial ao Pupin pela ajuda oferecida e tempo dedicado.

A minha família, que sempre me apoiou e incentivou na busca dos meus objetivos.

A Deus, por sua eterna orientação.

Minha eterna gratidão a todos que fizeram parte desta jornada.

RESUMO

Imagens MODIS são amplamente utilizadas para análise de vários fenômenos terrestres, como a fenologia da vegetação, mudanças no uso e cobertura do solo e monitoramento do desmatamento. Em geral, os produtos MODIS destinados às análises multitemporais são compostos por mosaicos dos melhores pixels adquiridos ao longo de certo período de tempo (tipicamente duas semanas). Entretanto, a análise das séries temporais é prejudicada pela presença de pixels com baixa qualidade ao longo da composição de dados. Na presente pesquisa é apresentada uma metodologia para redução de ruídos em séries temporais que considera a qualidade da informação disponibilizada para um ponto de interesse, bem como a relação no espaço e no tempo para os vizinhos. Esta metodologia, chamada de Janela de Regressão (JR), estima o valor do ponto de interesse baseado em análises de regressão de dados selecionados por meio de uma janela espaço-temporal. A partir do produto MOD13Q1 foram selecionadas três áreas testes com o objetivo de representar a diversidade de uso e cobertura do solo presente no município de Ipuã/SP. Em cada área teste foi avaliado o desempenho do método proposto, e de mais outras técnicas de filtragem (4253H twice, Mean Value Iteration e Savitzky–Golay), na redução de ruídos simulados ao longo do tempo e do espaço. A simulação de ruídos consistiu em introduzir níveis de ruídos com base em quatro tipos de amostragem de dados (tempo; tempo e espaço; grupo de pixels ou cluster; e sequência de datas) sobre o conjunto de imagens de cada área teste. A medida de dispersão Mean Absolute Percentage Error (MAPE) e o índice Akaike Information Criterion (AIC) indicaram que o método JR apresentou um desempenho superior às demais técnicas avaliadas nas estimativas de novos valores de NDVI. As técnicas de filtragem Janela de Regressão e Savitzky–Golay forneceram os melhores ajustes na redução dos ruídos de baixa intensidade (ruídos de 10%) para todos os tipos de amostragem. Para os ruídos simulados com os níveis de 30% e 50% a técnica de filtragem JR apresentou desempenho superior à obtida pelas demais técnicas analisadas. As discrepâncias entre o dado original e filtrado obtidas pela JR foram consistentes independentemente do nível de ruído simulado, pois esta técnica não utiliza a informação de baixa qualidade (ou com ruído) nos procedimentos para a estimativa de novos valores. Porém, observou-se uma queda de desempenho na estimativa de novos valores em áreas com maior diversidade de uso e cobertura do solo, o que indica a necessidade de cautela no emprego da metodologia proposta em áreas heterogêneas. Os filtros que necessitam da definição de limiares ou janela de análises podem ter desempenhos distintos em função do comportamento da série temporal a ser analisada ou do nível de acerto do usuário em definir tais parâmetros. Os resultados obtidos nesta pesquisa indicam bom potencial da metodologia para reduzir o ruído em séries temporais NDVI/MODIS; no entanto, mais investigações são necessárias para a redução de ruído por meio de análise espaço-temporal.

WINDOW REGRESSION: A SPATIAL-TEMPORAL ANALYSIS TO ESTIMATE PIXELS CLASSIFIED AS LOW-QUALITY IN MODIS NDVI TIME SERIES

ABSTRACT

MODIS data is largely used in multitemporal analysis of various Earth related phenomena, such as vegetation phenology, land use/land cover change, deforestation monitoring and time series analysis. In general, the MODIS products used to undertake multitemporal analysis are composite mosaics of the best pixels over a certain period of time (typically two weeks). However, it is common the presence of bad pixels in the composition which affect the time series analysis. We present a filtering methodology that takes the pixel position (location in space) and time (location in the temporal data series) to define a new value for the bad pixel. This methodology, called Window Regression (WR), estimates the value of the point of interest based on regression analysis of the data selected by a spatial-temporal window. From the product MOD13Q1 intensities of noise were simulated over time and space, and both the method presented and other techniques (4253H twice, Mean Value Iteration (MVI) and Savitzky–Golay (SG)) were evaluated using the Mean Absolute Percentage Error (MAPE) and Akaike Information Criterion (AIC). The simulated noises in time series were scaled into three intensities. The Window Regression and Savitzky–Golay filtering techniques provided the best fits for the noise reduction at low intensity (10% noise), according to the MAPE and AIC values. For the 30% and 50% noise simulations, the WR presented the best performance when compared to the other techniques. It is obvious that the noise intensity in the vegetation indices affects the results obtained by the 4253H twice, MVI and SG filtering techniques. On the other hand, it does not occur in the WR filtering technique, since the NDVI value of the pixel, along with the pixels considered as of low quality, are not used in the procedure for estimating the new value of the pixel of interest. The proposed method uses spatial-temporal analysis to estimate new NDVI value for a given pixel, and the analysis between neighboring pixels will be applied with caution in heterogeneous areas, or in more fragmented landscapes. The results obtained in this research indicate good potential of the method to reduce noise in MODIS NDVI time series data; however, more investigations are needed for noise reduction of NDVI time series from spatial-temporal analysis

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1.1 - Perfil Temporal e Valores de Qualidade para pixels vizinhos.	3
Figura 3.1 - Fluxograma para reconstrução de série temporal com base no método do valor médio da iteração.	12
Figura 4.1 – Localização geográfica da área de estudo.	15
Figura 4.2 - Mapa de referência de campo do município de Ipuã para o mês de Agosto de 2003.....	17
Figura 5.1 - Esquema metodológico das principais etapas realizadas na presente pesquisa.....	20
Figura 5.2 – Máscara de qualidade simulada ao longo de uma série temporal.	22
Figura 5.3 – Janela espaço-temporal para o pixel (3,3) na data 6 ($i = 6$).	23
Figura 5.4 – Análise do número e intervalo de pixels classificados como de baixa qualidade ao longo do perfil temporal.	23
Figura 5.5 – Perfil temporal do valores de NDVI para os pixels (3,3) e (3,4).	24
Figura 5.6 - Análise de regressão para o pixel (3,3) _i com uma janela temporal de (i-5 a i+5). Os círculos representam dados com qualidade satisfatória para as datas analisadas. Os triângulos significam que o dado foi considerado inválido para a regressão por possuir baixa qualidade em pelo menos uma data. O triângulo branco é o ponto de interesse.	27
Figura 6.1 – Etapas para simulação de dados e avaliação do desempenho das técnicas de filtragem na estimativa de novos valores..	29
Figura 6.2 – Cobertura do solo do município de Ipuã entre os anos 2003 – 2010.	30
Figura 6.3 – Localização das áreas de teste.....	31

Figura 6.4 – Ilustração dos tipos de amostragem simuladas para avaliar as técnicas de filtragem utilizadas neste estudo.....	32
Figura 6.5 – Perfil Temporal de NDVI original e degradado com 3 níveis de ruídos para o pixel (3,3).....	34
Figura 7.1 - Espacialização da porcentagem das datas obtidas ao longo de um ano (23 imagens) em que os pixels foram classificados com valores 2 ou 3 no produto <i>reliability</i>	39
Figura 7.2 – Porcentagem da área de estudo distribuída nos intervalos de qualidades (< 5%, 5 a 10%, 10 a 15% e 15 a 20% de datas com baixa qualidade) da série temporal para cada ano safra.	39
Figura 7.3 – Porcentagem de pixels da área de estudo em relação ao Máximo tamanho de <i>gap</i> obtido ao longo de um ano.....	40
Figura 7.4 – Imagens <i>reliability</i> referente ao conjunto de dados (44 datas entre Janeiro de 2003 e Dezembro de 2004) que compõem a Área “A”.....	42
Figura 7.5 – Exemplo da avaliação de desempenho do filtro 4253H twice sobre um perfil temporal simulado com ruído de 50%.	43
Figura 7.6 – Acurácia das técnicas de redução de ruídos (4253H twice – MVI – SG – JR) para a amostragem “A” com base no MAPE (erro percentual médio absoluto) e valores médios de AIC obtido entre 1000 iterações.....	44
Figura 7.7 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações realizadas na amostragem “II”..	46
Figura 7.8 - Acurácia das técnicas de redução de ruídos (4253H twice – MVI – SG – JR) para a amostragem “II” com base no valor médio do AIC (Akaike Information Criterion) obtido entre 1000 iterações.....	46
Figura 7.9 - Boxplot (no formato de violino) do erro percentual médio absoluto para as 1000 iterações realizadas na amostragem “III”..	50

Figura 7.10 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações realizadas na amostragem “IV”	53
Figura 7.11 - Acurácia das técnicas de redução de ruídos (4253H twice – MVI – SG – JR) para a amostragem “II” com base no valor médio do AIC obtido entre 1000 iterações sobre as áreas teste “B” e “C”	56
Figura A.1 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “II” realizadas na Área “B” ..	77
Figura A.2 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “II” realizadas na Área “C” ..	78
Figura B.1 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “III” realizadas na Área “B” .	79
Figura B.2 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “III” realizadas na Área “C” .	80
Figura C.1 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “IV” realizadas na Área “B”	81
Figura C.2 - Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “IV” realizadas na Área “C”	82

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 2.1 – Nome, período e resolução espacial dos produtos de índice de vegetação das plataformas Terra e Aqua.	7
Tabela 7.1 – Análise descritiva dos valores de MAPE para as 1000 iterações realizadas na amostragem "I"	44
Tabela 7.2 – Desempenho das técnicas de redução de ruídos com base no Teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações	48
Tabela 7.3 – Desempenho das técnicas de redução de ruídos com base no Teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações - Amostragem "III"	51
Tabela 7.4 – Desempenho das técnicas de redução de ruídos com base no teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações - Amostragem "IV"	54
Tabela 7.5 – Desempenho das técnicas de redução de ruídos com base no Teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações sobre as áreas "B" e "C"	57
Tabela 7.6 – Desempenho das técnicas de redução de ruídos com base no teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações sobre as áreas "B" e "C"	59
Tabela 7.7 – Desempenho das técnicas de redução de ruídos com base no teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações sobre as áreas "B" e "C"	60
Tabela D.1 – Valores de p-valor para o teste de Mann-Whitney entre os clusters simulados para a área teste "A"	84

Tabela D.2 – Valores de p-valor para o teste de Mann-Whitney entre os <i>gaps</i> simulados para a área teste “A”	85
-------------------------------------------------------------------------------------------------------------------------	----

LISTA DE SIGLAS E ABREVIATURAS

Pág.

AIC	Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average
AVHRR	Advanced Very High Resolution Radiometer
BISE	Best Index Slope Extraction
CATI	Coordenadoria de Assistência Técnica Integral
CV-MVC	Constrained View Angle - Maximum Value Compositing
EOS	Earth Observing System
FFT	Fast Fourier Transform
IV	Índice de Vegetação
JR	Janela de Regressão
LP-DAAC	Land Processes Distributed Active Archive Center
MAPE	Mean Absolute Percentage Error
MODIS	Moderate Resolution Imaging Spectroradiometer
MVC	Maximum Value Composite
MVI	Mean Value Iteration
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
QA	Quality Assessment
SDS	Science Datasets
SG	Savitzky-Golay

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1. Justificativa	2
1.2. Hipótese.....	5
1.3. Objetivo	6
2 PRODUTO ÍNDICE DE VEGETAÇÃO DO SENSOR MODIS.....	7
3 REDUÇÃO DE RUÍDOS EM ANÁLISE DE SÉRIES TEMPORAIS.....	9
3.1. Composição de Máximo Valor (MVC – Maximum Value Composite).....	10
3.2. Filtro 4253H twice	10
3.3. Valor Médio da Iteração (MVC - Mean Value Iteration).....	11
3.4. Filtro Savitzky-Golay	13
4 ÁREA DE ESTUDO.....	15
5 METODOLOGIA.....	19
5.1. Análise de Dados e Obtenção da Máscara de Qualidade	21
5.2. Janela de Regressão - Método para estimar novos valores.....	21
5.2.1. Regressão linear e estimativa de um novo valor	24
5.2.2. Estimativa dos valores de NDVI dos pixels com baixa qualidade.....	26
6 SIMULAÇÃO DE DADOS PARA ANÁLISE.....	29
6.1. Série temporal de referência.....	29
6.2. Degradação da série temporal	31
6.3. Reconstrução da série temporal	34
6.4. Análise dos resultados.....	35

7	RESULTADOS E DISCUSSÃO	37
7.1.	Análise da base de dados	37
7.2.	Avaliação de desempenho para área teste “A”	40
7.2.1.	Avaliação para simulação espaço-temporal (Amostragens I e II)	40
7.2.2.	Avaliação de desempenho para simulação espacial (Amostragem III).....	48
7.2.3.	Avaliação de desempenho para simulação temporal (Amostragem IV).....	51
7.3.	Avaliação de desempenho para áreas testes “B” e “C”	55
7.4.	Síntese dos resultados obtidos (áreas teste “A”, “B” e “C”)	58
8	CONCLUSÃO	65
	REFERÊNCIAS BIBLIOGRÁFICAS	69
	APÊNDICE A – Tabelas e Figuras referentes à amostragem “II”	77
	APÊNDICE B – Tabelas e Figuras referentes à amostragem “III”	79
	APÊNDICE C – Tabelas e Figuras referentes à amostragem “IV”	81
	APÊNDICE D – Resultado do Teste não paramétrico	83
	APÊNDICE E – Código R - Janela de Regressão.....	87

1 INTRODUÇÃO

As séries temporais de índice de vegetação têm sido utilizadas com sucesso para detecção de mudanças da cobertura terrestre, monitoramento e derivação de parâmetros biofísicos da vegetação (ZHANG et al., 2003; HUETE et al. 2002). Entretanto, as séries temporais de índice de vegetação podem conter distúrbios causados pelas diferentes condições da atmosfera (presença de nuvens, ozônio, poeira e outros aerossóis), pelos efeitos bidirecionais relacionados à geometria de aquisição de dados (fonte-alvo-sensor) e por falhas instrumentais ou de processamento dos dados. Assim, os ruídos presentes em série de dados de índice de vegetação devem ser minimizados antes de empregar os dados para investigações sobre fenologia da vegetação (HIRD; MCDERMID, 2009).

Vários métodos de interpolação de dados de séries temporais de índice de vegetação têm sido propostos para eliminação de ruído e reconstrução, com alta qualidade, da série de dados (ATZBERGER; EILERS, 2011a). De forma geral, os métodos disponíveis na literatura possuem vantagens e desvantagens, sendo que em alguns métodos a eficácia está condicionada à experiência do analista em definir tamanhos de janelas, graus de polinômios e limiares para eliminação dos ruídos.

Além disso, os métodos para filtragem de ruídos em séries temporais de índice de vegetação normalmente trabalham apenas num domínio - da frequência ou do tempo, ou seja, analisam o comportamento do pixel em apenas uma dimensão da informação.

Neste contexto, o objetivo desta pesquisa é desenvolver uma metodologia capaz de identificar os ruídos em séries temporais de índice de vegetação, e substituir os seus valores por meio de estimativas que envolvam análises dos dados no espaço e no tempo.

1.1. Justificativa

O Índice de Vegetação por Diferença Normalizada (NDVI) é amplamente utilizado para diversas aplicações relacionadas ao mapeamento e classificação da cobertura terrestre, detecção de mudanças, derivação de propriedades biofísicas da vegetação (por exemplo, biomassa, índice de área foliar, fração da radiação fotossinteticamente ativa, estádios fenológicos) e estimativa de produtividade (GU et al., 2009).

Verbesselt et al. (2010) destacam que as mudanças em ecossistemas podem ser divididas em três classes: (1) mudança sazonal – oriundas da variação de temperatura e precipitação ao longo do ano; (2) mudança gradual – alterações interanuais do clima, do regime pluviométrico, gestão territorial ou degradação da superfície terrestre; (3) mudança abrupta – causada por distúrbios como desflorestamentos, queimadas, catástrofes (enchente, deslizamento), etc. Os autores enfatizam que estimativa de mudança a partir de dados de sensoriamento remoto não é um processo simples, uma vez que as séries temporais de dados podem conter informações provenientes da combinação entre as mudanças sazonais, graduais e bruscas, além da presença de ruídos originados por fatores como a presença de nuvens, geometria de aquisição de dados e espalhamento atmosférico.

O sensor MODIS (Moderate Resolution Imaging Spectroradiometer) fornece observações simultâneas das feições terrestres, oceânicas e atmosféricas. Os dados coletados são processados para fornecer uma gama de produtos como índice de vegetação, índice de área foliar, temperatura da superfície do mar, fração de nuvens, entre outros. Porém, as imagens são susceptíveis a grandes variações na geometria de iluminação e de visada, e esses efeitos podem gerar grandes variações nos valores da reflectância de superfície e consequentemente influenciar nos valores dos índices de vegetação.

As variações do NDVI em função da geometria de visada foram avaliadas por Breunig et al. (2011) e Epiphanyo e Huete (1995) para alvos agrícolas. Os resultados mostraram que, para um mesmo estágio fenológico da cultura, o

NDVI calculado para a direção do retroespalhamento tende a ser inferior ao derivado das imagens adquiridas na direção do espalhamento frontal.

A avaliação da qualidade (*Quality Assessment – QA*) e as atividades de validação em campo fazem parte de um protocolo dos produtos MODIS para demonstrar o desempenho dos dados distribuídos. Os valores de QA armazenados para cada pixel fornecem aos usuários um histórico sobre a condição atmosférica, geometria de visada e iluminação, e método de processamento do produto.

Roy et al. (2002) destacam que os valores de QA disponíveis nos produtos MODIS contêm informações que contribuem para uma correta interpretação dos valores presentes em uma série temporal de índice de vegetação.

Um exemplo da importância desses dados na interpretação e aplicação de um perfil temporal pode ser ilustrado na Figura 1.1, onde pixels vizinhos de um mesmo alvo possuem diferentes classificações de qualidade conforme o protocolo do produto Índice de Vegetação do MODIS – composição 16 dias – resolução 250m (MOD13Q1).

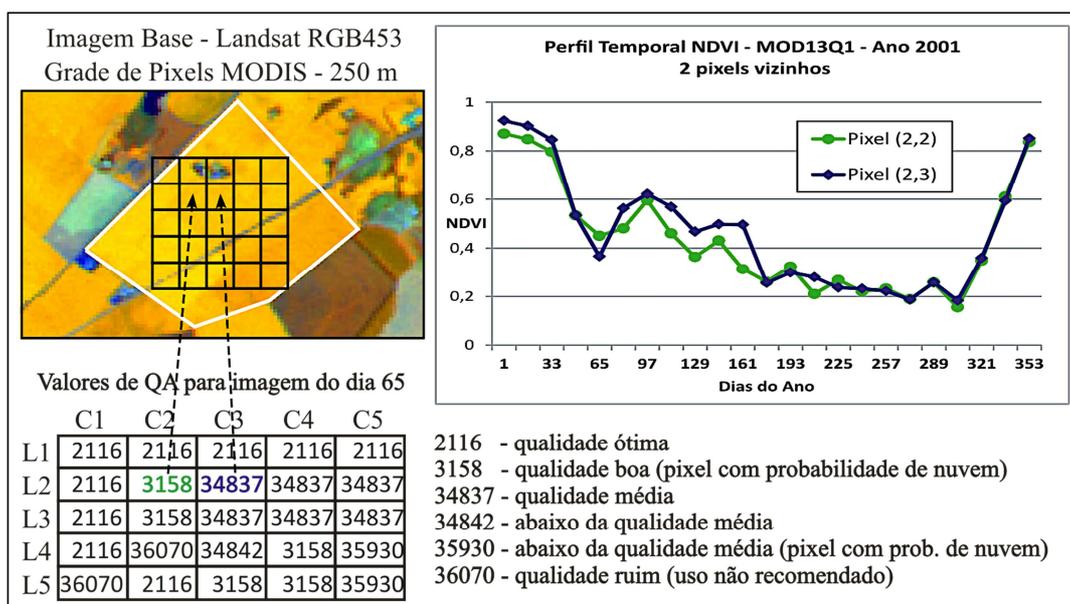


Figura 1.1 – Perfil Temporal e Valores de Qualidade para pixels vizinhos.

A discrepância de NDVI entre os pixels vizinhos apresentada na Figura 1.1 pode estar relacionada somente à resposta espectral da vegetação; porém, os valores de QA indicam que as condições para aquisição e processamento dos pixels não foram as mesmas. Partindo do princípio que os pixels analisados (Pixel 2,2 e 2,3) possuem o mesmo tipo de uso e cobertura do solo e estádios de desenvolvimento, seria esperado que as respostas espectrais entre os pixels fossem semelhantes. No entanto, esta situação não foi observada para algumas datas ao longo da série temporal de NDVI, como por exemplo, para a data 65, onde o pixel (2,2) teve um valor de NDVI 23% maior que o pixel (2,3).

As séries temporais de NDVI derivadas de sensores a bordo do NOAA/AVHRR, MODIS, Landsat e SPOT Vegetation têm proporcionado um aumento de informações sobre a cobertura vegetal em diversas escalas de mapeamento. Entretanto, é importante que as informações inconsistentes nas escalas espaço-tempo provenientes da presença de nuvens, variabilidade atmosférica, efeitos bidirecionais e problemas instrumentais sejam minimizadas ou eliminadas.

Parte dos ruídos presentes em séries temporais pode ser removida por métodos de processamento durante a composição da série temporal, como por exemplo, o método Maximum Value Composite – MVC (HOLBEN, 1986). Porém, a remoção de distúrbios atmosféricos numa área com grande incidência de nuvens não é efetiva. Por isso, vários grupos de pesquisas implementaram e continuam desenvolvendo métodos para filtragem de ruídos ou reconstrução de séries temporais para obtenção de dados com melhor qualidade (ATZBERGER; EILERS, 2011a).

Inúmeras metodologias têm sido utilizadas para redução de ruídos em séries temporais de índice de vegetação antes da utilização dos dados na estimativa de parâmetros fenológicos da vegetação. Entre as técnicas de filtragem presentes na literatura estão os filtros 4253H twice (VELLEMAN, 1980); Savitzky–Golay (CHEN et al., 2004; SAVITZKY; GOLAY, 1964); valor médio da iteração (MA; VEROUSTRAETE, 2006); *weighted least squares windowed regression* (SWETS et al., 1999); *best index slope extraction* (VIOVY et al.,

1992); *Whittaker smoother* (ATZBERGER; EILERS, 2011a, b); e *double logistic* (BECK et al., 2006; ZHANG et al., 2003). Outras técnicas de filtragem utilizam análises harmônicas (BRADLEY et al., 2007), transformada de Fourier (BROOKS et al., 2012; ROERINK et al., 2000) ou wavelet (CHEN et al., 2011; LU et al., 2007; SAKAMOTO et al., 2005, 2010) para redução de ruídos em séries temporais. A abordagem geoestatística também tem sido utilizada para o preenchimento de *gaps* ou ausência de dados ocasionados pela presença de nuvens (PRINGLE et al., 2009; ZHANG et al., 2009; ZHU et al., 2012), bem como o uso de técnicas de interpolação espaço-temporal na reconstrução do pixel (CHO; SUH, 2013; OLIVEIRA; EPIPHANIO, 2012; POGGIO et al., 2012).

Em geral, muitas metodologias de filtragem trabalham no domínio da frequência ou do tempo, ou seja, analisam o comportamento espectral do pixel em uma dimensão da informação. Estes métodos propõem a redução de ruídos e manutenção da integridade do índice de vegetação (IV) para uma série temporal de imagens; porém, não consideram a relação espacial e temporal dos valores dos pixels. Poggio et al. (2012) destacam que a informação espacial aliada à dimensão temporal de um conjunto de dados possibilita a reconstrução de uma série temporal com maior acurácia. Ao mesmo tempo, Roy et al. (2002) ressaltam que as informações de qualidade disponíveis nos produtos MODIS contribuem para uma adequada interpretação e aplicação dos dados, de forma que os produtos classificados como de baixa qualidade devem ser utilizados com cautela.

1.2. Hipótese

A hipótese desta pesquisa está baseada na pressuposição de que um grupo de pixels de um mesmo alvo possui relação espacial e temporal para o comportamento espectral da vegetação, e que uma análise em conjunto dos valores de IV ao longo do tempo e do espaço pode gerar informações que contribuam para a redução ou eliminação de ruídos presentes em composição temporal de imagens.

Acredita-se que a partir de uma análise espacial do entorno de um pixel, associada com o seu comportamento ao longo do tempo e com os parâmetros de qualidade, seja possível a reconstrução de uma série temporal de NDVI com menor nível de ruídos.

1.3. Objetivo

Em conformidade com a hipótese exposta acima, a presente pesquisa tem como objetivo geral desenvolver uma metodologia capaz de analisar a relação integrada do espaço e tempo dos valores de NDVI, e dos parâmetros de qualidade e aquisição dos dados disponibilizados pelo produto MOD13Q1, com a finalidade de reconstrução da série temporal de imagens de índice de vegetação com menor incidência de ruídos.

Como objetivos específicos, pretende-se:

- Identificar pixels que possam ser classificados como ruídos, ou de baixa qualidade, por meio de uma análise das informações contidas nos metadados do produto MODIS;
- Analisar a distribuição espacial dos pixels classificados com baixa qualidade ao longo de uma série temporal;
- Avaliar a variabilidade temporal das informações de qualidade do pixel ao longo de uma série temporal de imagens;
- Estimar valores de índice de vegetação por meio de análise de regressão envolvendo dados coletados no espaço-tempo e parâmetros de qualidade do pixel.
- Avaliar a eficiência da simulação de dados degradados no tempo e no espaço como suporte para aplicação das técnicas de redução de ruídos..
- Avaliar a acurácia das estimativas dos novos valores de NDVI por meio de medidas de dispersão.

2 PRODUTO ÍNDICE DE VEGETAÇÃO DO SENSOR MODIS

O programa internacional EOS (Earth Observing System), criado com o objetivo de estudar os processos de mudanças globais, pode ser considerado como um marco nos avanços tecnológico e científico para aquisição, monitoramento e análise nas escalas global, regional e local de fenômenos relacionados à atmosfera, oceanos e superfície terrestre (SALOMONSON et al., 2011). Dos conjuntos de sensores desenvolvidos para atender ao programa espacial EOS, o sensor MODIS destaca-se em virtude das características de alta resolução temporal, moderada resolução espacial, bandas espectrais que abrangem as faixas desde o visível até o infravermelho termal e da disponibilidade de produtos que têm contribuído para uma ampla aplicação dos dados para mapeamento e monitoramento da superfície terrestre (PITTMAN et al., 2010).

O Land Processes Distributed Active Archive Center (LP DAAC) é responsável pela distribuição dos produtos MODIS, entre eles os índices de vegetação Normalized Difference Vegetation Index (NDVI) proposto por Rouse et al. (1973), e o Enhanced Vegetation Index (EVI) apresentado por Huete et al. (2002).

Os produtos de índice de vegetação derivados das plataformas Terra e Aqua possuem variações conforme o período de aquisição dos dados e resolução espacial. Esta gama de produtos está sintetizada na Tabela 2.1.

Tabela 2.1 – Nome, período e resolução espacial dos produtos de índice de vegetação das plataformas Terra e Aqua.

Terra	Aqua	Período	Resolução Espacial
MOD13Q1	MYD13Q1	16 dias	250 m
MOD13A1	MYD13A1	16 dias	500 m
MOD13A2	MYD13A2	16 dias	1 km
MOD13A3	MYD13A3	mensal	1 km
MOD13C1	MYD13C1	16 dias	0,05°
MOD13C2	MYD13C2	mensal	0,05°

Fonte: Huete et al. (2011).

A validação dos produtos obtidos a partir de dados do sensor MODIS é realizada com base em métodos que vão desde a comparação de informações coletadas em áreas testes, até a comparação dos dados obtidos com informações adquiridas por outros sensores. O MODIS Land (MODLAND) Science Team (ST) tem desenvolvido e coordenado atividades para avaliar e validar o desempenho dos produtos MODLAND, incorporando aos metadados gerados as informações de qualidade de cada pixel em cada produto (JUSTICE et al., 2002). Essas informações de qualidade (Quality Assessment - QA) permitem ao usuário dos dados MODIS verificar a confiabilidade do pixel em relação às condições da atmosfera (presença de nuvens, ozônio, poeira e outros aerossóis), aos efeitos bidirecionais relacionados à geometria de aquisição de dados (fonte-alvo-sensor) e sobre possíveis falhas instrumentais ou métodos de processamento dos dados (ROY et al., 2002).

Para cada produto de índice de vegetação disponibilizado no MOD13 é gerado um conjunto de parâmetros de saída (Science Datasets - SDS), que contém as seguintes informações para cada pixel da imagem: 1) composição NDVI e EVI; 2) qualidade da informação disponibilizada (QA); 3) reflectância nas bandas 1, 2 e 7 (620–670 nm, 841–876 nm e 2.105–2.155 nm, respectivamente); 4) ângulo zenital de visada, ângulo zenital solar e azimute relativo; 5) dia do ano que foi obtido o pixel; 6) confiabilidade do pixel (HUETE et al., 2011).

As imagens de QA possuem valores em bits que documentam as condições sob as quais cada pixel foi adquirido e processado. A confiabilidade do pixel (do termo em inglês *reliability*) resume o estado da qualidade do produto por meio de cinco categorias (Sem dados – Ótimo – Marginal – Neve/Gelo – Nuvem) com valores variando de -1 a 3, e por isso Didan e Huete (2006) ressaltam que os usuários podem consultar esta informação ao invés do produto de QA.

Maiores detalhes sobre os instrumentos ou produtos disponibilizados a partir dos dados obtidos pelo sensor MODIS podem ser obtidos em Ramachandran et al. (2011).

3 REDUÇÃO DE RUÍDOS EM ANÁLISE DE SÉRIES TEMPORAIS

Morettin e Toloí (2006) definem uma série temporal como um conjunto de observações de uma variável observada ao longo do tempo $(x_1, x_2, \dots, x_t, \dots)$. Esta sequência de dados pode ser observada de forma contínua (observações obtidas em qualquer intervalo de tempo) ou discreta (observações obtidas em intervalos de tempo discreto e equidistantes).

Uma análise de série temporal essencialmente busca determinar a periodicidade, tendência, pontos influentes e a não-linearidades presente na sequência de dados. Basicamente, os tipos de análise de séries temporais possuem dois enfoques. No primeiro enfoque, a análise é realizada no domínio temporal por meio de modelos paramétricos (como por exemplo, o *Autoregressive Integrated Moving Average (ARIMA)*). No segundo, a análise é conduzida no domínio da frequência, onde os modelos propostos são modelos não paramétricos (como por exemplo, a transformada de Fourier) (MORETTIN e TOLOI, 2006).

Trabalhos como o de Bradley et al. (2007), Lasaponara (2006) e Sakamoto et al. (2005) demonstram que o comportamento fenológico da vegetação pode ser observado, analisado e mapeado por meio de um perfil temporal de NDVI. Entretanto, Hird e Mcdermid (2009) e Ma e Veroustraete (2006) destacam a necessidade de reduzir os ruídos presentes no perfil temporal de NDVI antes de sua utilização.

Na literatura estão disponíveis inúmeras metodologias com a finalidade de reduzir a presença de ruídos em séries temporais e promover a reconstrução do conjunto de imagens com uma melhor qualidade. Basicamente, os filtros empregados para a redução de ruídos em séries temporais podem ser agrupados em duas categorias:

1. Redução de ruído no domínio da frequência por meio de modelos não paramétricos (como por exemplo, a transformada de Fourier);

2. Redução do ruído no domínio do tempo, como por exemplo o algoritmo *Best Index Slope Extraction* (BISE); Savitzky–Golay; e o filtro do Valor Médio proposto por Ma e Veroustraete (2006).

3.1. Composição de Máximo Valor (MVC – Maximum Value Composite)

A metodologia MVC proposta por Holben (1986) consiste em avaliar uma sequência de imagens de NDVI e designar como representante de todo o conjunto o máximo valor observado. Esta técnica tem como propósito minimizar a contaminação de nuvem, dos efeitos direcionais da visada off-nadir, do ângulo solar e dos efeitos do sombreamento na resposta espectral dos alvos. O autor destaca que os benefícios que a técnica oferece dependem do estado da atmosfera, do tipo de cobertura do solo e da geometria de visada.

Huete et al. (2011) destacaram que o valor máximo de NDVI selecionado na série de imagens reduz a influência de nuvens e aerossóis, mas o valor selecionado não corresponde necessariamente aos menores ângulos de visadas. Além disso, ângulos de visadas na posição de espalhamento frontal aliados às propriedades anisotrópicas da maioria das superfícies vegetadas produzem maiores valores de NDVI, mesmo quando as imagens estão contaminadas por nuvens e aerossóis.

Uma das metodologias empregadas na geração do produto MOD13 é uma adaptação da técnica MVC, pois consiste na seleção de pixels baseados em critérios que envolvem a integridade do dado e a proximidade com o nadir (HUETE et al., 2002). Essa adaptação recebeu o nome de Constrained View Angle - Maximum Value Compositing (CV-MVC).

3.2. Filtro 4253H twice

O filtro 4253H twice é um filtro não linear proposto por Velleman (1980) que consiste em aplicar sobre a série temporal uma sequência de filtros de mediana com diferentes tamanhos de janelas. A parte numérica no nome do filtro representa os tamanhos das janelas utilizadas para o cálculo da mediana. A primeira fase de análise do filtro consiste em calcular os valores de mediana a

partir de uma janela equivalente a 4 datas. As etapas seguintes são com tamanhos de janelas equivalentes a 2, 5, e 3 datas, respectivamente.

O termo “H” disposto no nome do filtro refere-se à função de Hann, que é uma função simétrica utilizada para ponderar os elementos presentes em uma janela, onde o maior peso é aplicado ao elemento central.

Após a filtragem com os diferentes tamanhos de janela, a próxima etapa consiste em aplicar a função de Hann, com janela de tamanho igual a 3 e pesos de 1/4, 1/2 e 1/4, ao longo da série temporal. Posteriormente, são calculados os resíduos entre os dados originais e os filtrados.

O termo *twice* presente no nome do filtro está associado à aplicação dos procedimentos de filtragem tanto nos dados como nos resíduos. O resultado da filtragem sobre os resíduos é adicionado à série temporal filtrada para obter o produto final do filtro 4253H *twice*.

Dijk et al. (1987) ressaltam que o filtro 4253H *twice* apresentou resultados mais eficientes do que as técnicas de redução de ruídos baseados na média móvel ponderada, e maiores informações sobre esta técnica de filtragem podem ser obtidas em Adami (2010), Sargent e Bedford (2010), Hird e Mcdermid (2009) e Ma e Veroustraete (2006).

3.3. Valor Médio da Iteração (MVC - Mean Value Iteration)

A metodologia intitulada de Valor Médio da Iteração (Mean Value Iteration – MVI) foi desenvolvida por Ma e Veroustraete (2006) com o objetivo de reduzir os ruídos presentes em séries temporais de NDVI a partir de uma análise iterativa entre datas que compõem o perfil temporal. Basicamente, é efetuada uma comparação do valor do NDVI de uma data *i* com a média dos valores de NDVI obtidos anteriormente e posteriormente a esta data. O valor do NDVI da data em análise será substituído pelo valor da média caso a divergência entre o ponto analisado e a média dos vizinhos ultrapasse um limiar pré-estabelecido (Figura 3.1).

Antes de iniciar todo o procedimento de filtragem é realizada uma análise sobre a série temporal original com o objetivo de substituir os pixels sem valores pelo valor médio de NDVI calculado para um determinado intervalo da série de dados.

Ma e Veroustraete (2006) utilizaram um limiar de 10% para aplicar esta metodologia em uma série temporal de NDVI derivada de dados AVHRR. O desempenho do algoritmo foi superior quando comparado com metodologias baseadas em *Fast Fourier Transform* (FFT) (VERHOEF et al., 1996) ou com o algoritmo BISE modificado proposto por Lovell e Graetz (2001). Porém, os autores destacam a importância da escolha do limiar e sua influência direta no resultado da filtragem a ser obtido.

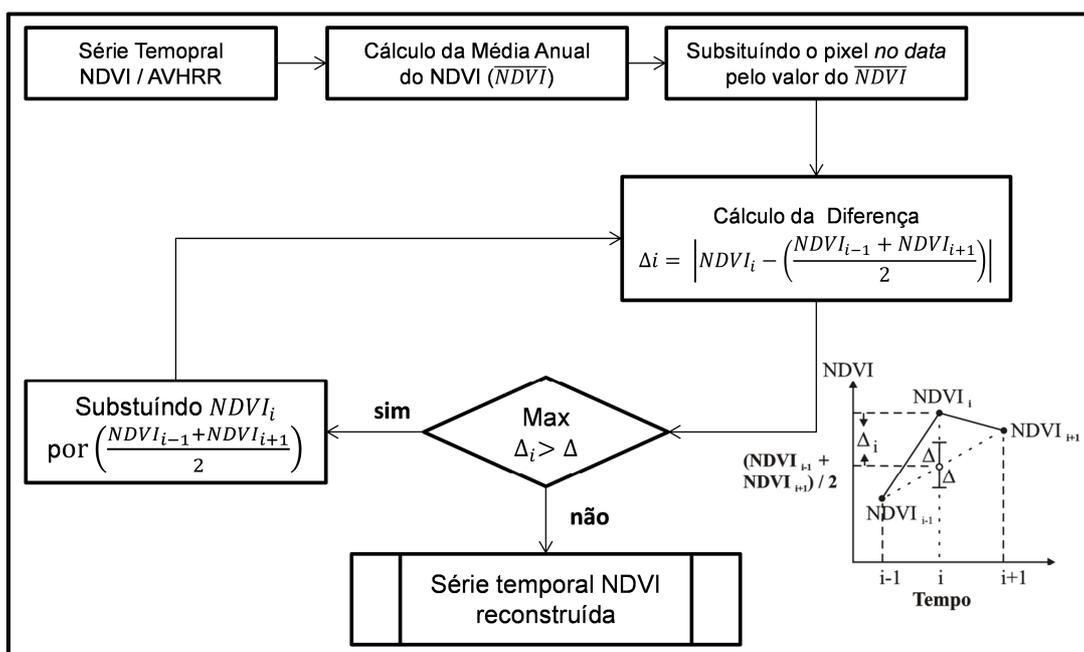


Figura 3.1 – Fluxograma para reconstrução de série temporal com base no método do valor médio da iteração.
Fonte: adaptado de Ma e Veroustraete (2006).

Ma e Veroustraete (2006) demonstraram que o filtro MVI é uma técnica promissora na redução de ruídos em séries temporais, visto que seu desempenho foi superior quando comparado a outras técnicas de filtragem (p.ex. BISE). Além da fácil implementação, a natureza iterativa do filtro MVI

permite sua aplicação independentemente da variabilidade das séries temporais de NDVI (HIRD; MCDERMID, 2009).

3.4. Filtro Savitzky–Golay

Filtro desenvolvido por Savitzky e Golay (1964), e reestruturado posteriormente por Chen et al. (2004), emprega um método de mínimos quadrados para suavizar um conjunto de dados consecutivos. Especificamente, este filtro pode ser entendido como um filtro de média móvel ponderada cujos pesos são determinados por um ajuste polinomial. Chen et al. (2004) destacam que este polinômio é desenvolvido com o objetivo de preservar os pontos máximos ao longo da série de dados e minimizar os desvios introduzidos pelo filtro. O processo de filtragem de uma série temporal de NDVI pelo filtro Savitzky–Golay pode ser geralmente descrito por:

$$Z_t = \frac{\sum_{i=-m}^{i=m} C_i Y_{i+t}}{N} \quad (3.1)$$

em que Z_t é o novo valor de NDVI para a data t ; C_i é o coeficiente para o ponto i -ésimo valor da janela de análise do filtro; Y_{i+t} é o valor original de NDVI; m é a metade do tamanho da janela do filtro; e N é o número de pontos analisados ao longo da série temporal. O tamanho da janela de análise do filtro é equivalente a $(2m+1)$, e os coeficientes do filtro podem ser obtidos no trabalho de Savitzky e Golay (1964) ou calculados a partir das equações descritas em Madden (1978).

De acordo com Cerqueira et al. (2000), o processo de análise do filtro Savitzky–Golay pode ser descrito em cinco etapas: (i) determinar a largura da janela de análise; (ii) definir o ponto central da janela de análise e removê-lo do conjunto de dados; (iii) ajustar, por meio do método de mínimos quadrados, um polinômio de grau variável com as informações presentes no conjunto de dados; (iv) utilizar o polinômio para estimar e substituir o valor do ponto removido; (v) deslocar a janela de análise ao longo do tempo e repetir todo o procedimento.

CHEN et al. (2004) avaliaram e compararam o desempenho deste filtro com métodos baseados em análise de Fourier e algoritmo BISE, e concluíram que as três técnicas de filtragem foram eficientes na reconstrução da série temporal de NDVI. Porém, a facilidade e simplicidade de uso do filtro Savitzky–Golay foi destacada como vantagem entre os filtros analisados.

4 ÁREA DE ESTUDO

O município de Ipuã está localizado no norte do Estado de São Paulo (Brasil) entre as coordenadas geográficas 20° 17' e 20° 32' latitude sul e 47° 54' a 48° 13' longitude oeste e foi utilizado como área de estudo para o desenvolvimento e avaliação da metodologia (Figura 4.1). O município de Ipuã possui uma área equivalente a 46.559 ha.

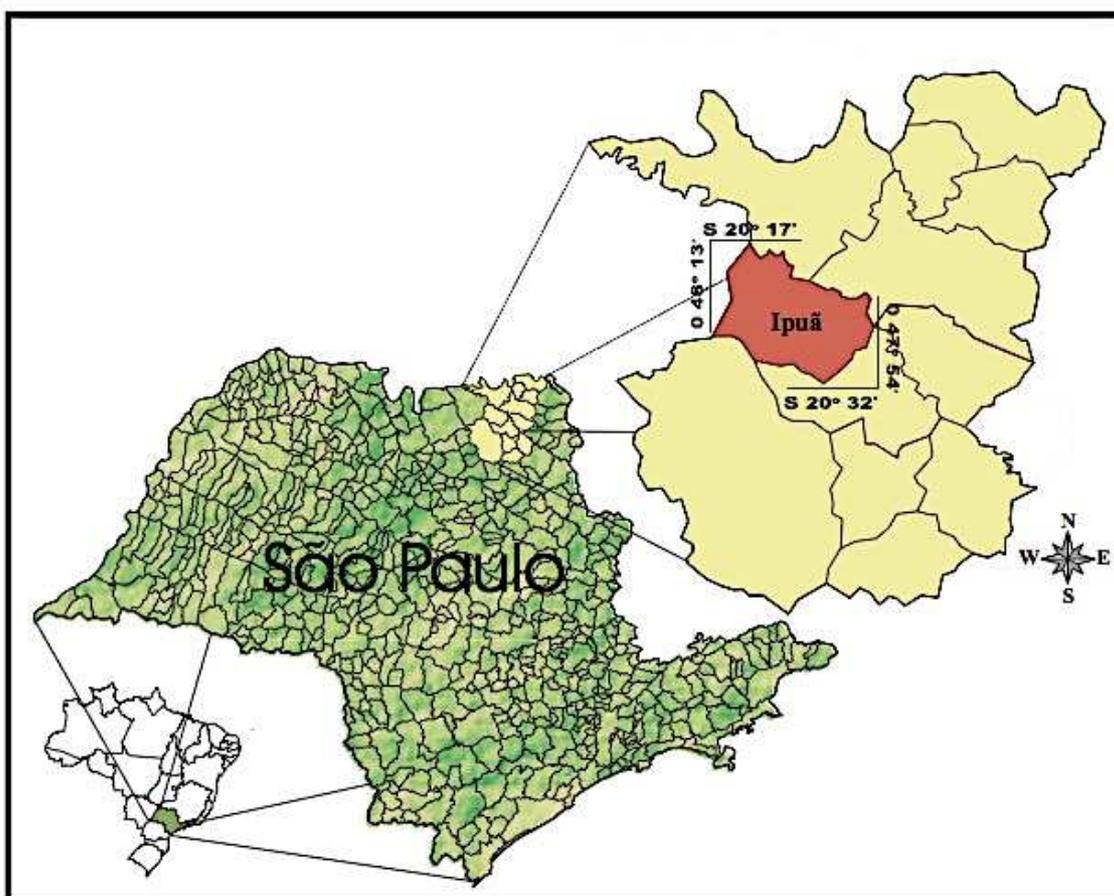


Figura 4.1 – Localização geográfica da área de estudo.

Para o desenvolvimento da metodologia de análise e redução de ruídos na série temporal de NDVI, foi definido como dado-base o produto MOD13Q1 (MODIS / Índice de Vegetação – 16 dias – 250m – tile h13v11), especificamente os subprodutos relacionados ao índice de vegetação e aos arquivos de confiabilidade do pixel disponibilizados pelo LP DAAC. A série temporal de dados analisada no presente estudo foi obtida no catálogo de

metadados da EOS/NASA, por meio da ferramenta Reverb (<http://reverb.echo.nasa.gov/reverb/>), para o período de 2003 a 2010.

Wardlow et al. (2007) investigaram a aplicabilidade das imagens MODIS no mapeamento de culturas agrícolas, e concluíram que as séries temporais de IV do produto MOD13 possuem resoluções temporal e radiométrica suficientes para discriminar os principais tipos de culturas e práticas agrícolas. Esses autores também observaram que as imagens MODIS podem fornecer informações sobre a fenologia da vegetação e características climáticas regionais, pois as variações intraclases decorrentes das variações climáticas e de datas de plantio foram detectadas na análise das séries temporais.

A justificativa para a delimitação geográfica e temporal para o desenvolvimento da pesquisa apoia-se no fato de o município de Ipuã apresentar ampla variabilidade de classes e dimensões dos fragmentos presentes na cobertura do solo para o período proposto. Esta variabilidade de uso e cobertura do solo foi mapeada e divulgada por Sanches (2004), que utilizou de trabalhos de campo e imagens dos sensores TM e ETM+ (Landsat 5 e 7 respectivamente) obtidas entre 08/01/00 e 24/12/03 para analisar o comportamento espectro-temporal e estimar a área agrícola das principais culturas no município de Ipuã. O mapeamento realizado por Sanches (2004) para o período de agosto de 2003 está apresentado na Figura 4.2.

As variações sazonais e interanuais dos alvos presentes na área de estudo foram consideradas adequadas para avaliar o potencial da metodologia proposta na redução de ruídos em séries temporais provenientes do sensor MODIS.

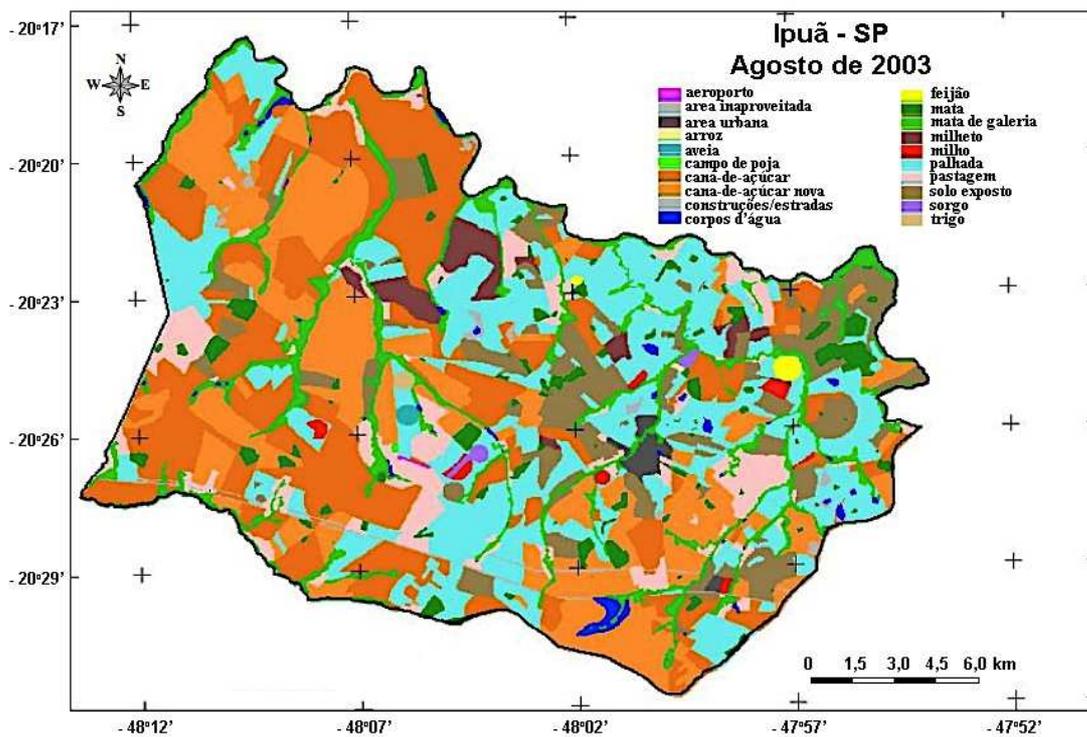


Figura 4.2 – Mapa de referência de campo do município de Ipuã para o mês de Agosto de 2003.

Fonte: Sanches (2004).

5 METODOLOGIA

Normalmente, um perfil temporal de NDVI deve apresentar um comportamento contínuo e suave, pois as mudanças que ocorrem no dossel da vegetação são pequenas em relação ao tempo. Porém, há a ocorrência de amplas flutuações de valores ou falhas ao longo da composição de imagens de NDVI por causa da presença de nuvens, erro na transmissão de dados, correção atmosférica com procedimentos incompletos ou inconsistentes e efeitos bidirecionais (MA; VEROUSTRAETE, 2006).

Druck et al. (2004) destacam que a ênfase da análise espacial é medir propriedades e relacionamentos levando em consideração a localização espacial do fenômeno. Sendo assim, a presente metodologia para redução de ruídos em série temporal de NDVI propõe uma avaliação mútua dos valores do IV ao longo do tempo e do espaço, aliada às informações de qualidade geradas durante a aquisição do dado e processamento da série temporal. O princípio desta análise contextual parte da primeira lei da geografia, em que *“todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes”* (TOBLER, 1970). A dependência espacial é considerada a chave na compreensão e análise dos fenômenos espaciais, e Cressie (1993) enfatiza que ela está presente em todas as direções e fica mais fraca à medida que aumenta a dispersão na localização dos dados.

Para atingir os objetivos desta pesquisa foram definidos quatro segmentos metodológicos, conforme ilustrado na Figura 5.1. Os dois primeiros segmentos estão relacionados à descrição da metodologia proposta para a estimativa dos valores de NDVI em pixels classificados como de baixa qualidade. Esta metodologia será intitulada ao longo deste trabalho como Janela de Regressão (no inglês – Window Regression (WR)).

O segmento (1) consiste em uma análise descritiva da série temporal a partir das informações de qualidade do produto MOD13Q1. Informações sobre a confiabilidade do pixel, qualidade do índice de vegetação, data de aquisição e ângulos da geometria de visada serão úteis para a ordenação e triagem de

dados de forma que seja possível a construção de uma máscara de qualidade dos pixels.

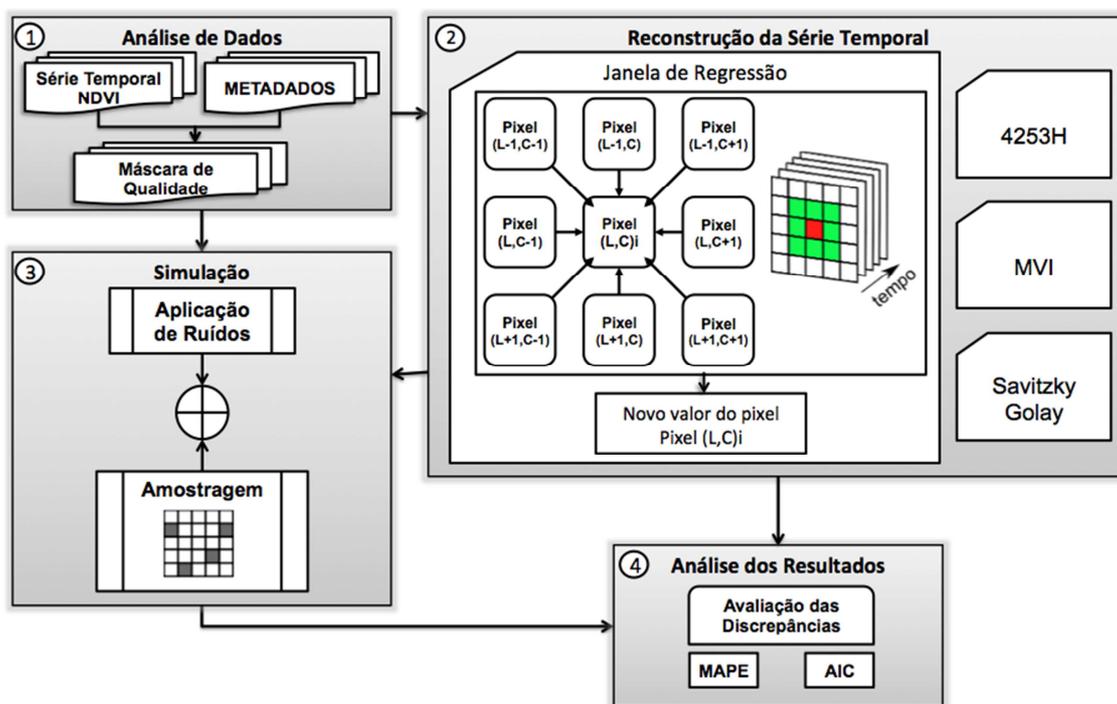


Figura 5.1 – Esquema metodológico das principais etapas realizadas na presente pesquisa.

No segmento (2) serão descritos os procedimentos que compõem o método Janela de Regressão. Em um contexto geral, o método emprega uma análise de regressão linear a partir de dados selecionados por uma janela espaço-temporal, para recalculer os valores de NDVI em pixels identificados com baixa qualidade. As técnicas 4253H twice, MVI e Savitzky–Golay foram selecionadas para comparação do desempenho na estimativa de novos valores de NDVI entre as técnicas de filtragem.

O segmento (3) desenvolvido neste trabalho descreve os métodos elaborados para a simulação de ruídos sobre um conjunto de imagens temporais. Por fim, o segmento (4) analisa as discrepâncias com o objetivo de apresentar o desempenho de cada técnica de filtragem na estimativa de novos valores de NDVI.

5.1. Análise de Dados e Obtenção da Máscara de Qualidade

A primeira etapa metodológica consiste em identificar o pixel $(L,C)_i$ (L-linha, C-coluna, para a data i) onde o produto MODIS foi classificado como de baixa qualidade. No presente estudo, a seleção de pixels com baixa qualidade foi realizada a partir de uma consulta ao produto de confiabilidade do MODIS, de forma que os pixels classificados como gelo/neve (*reliability* igual 2) ou nuvem (*reliability* igual a 3) nas imagens de confiabilidade foram classificados como de baixa qualidade na máscara.

Ao final da consulta à base de dados, a máscara de qualidade será composta apenas pelos pixels e datas que terão os respectivos valores de NDVI recalculados na segunda etapa metodológica desse estudo.

5.2. Janela de Regressão - Método para estimar novos valores

Os pixels presentes na máscara de qualidade que foram classificados como de qualidade insatisfatória para o objetivo proposto terão os respectivos valores de NDVI estimados pela técnica JR.

Uma série temporal de doze imagens com dimensão de 5x5 (linhas x colunas) foi simulada com o intuito de ilustrar os procedimentos da metodologia desenvolvida na JR. O pixel $(3,3)_6$ (linha 3, coluna 3, data 6) da Figura 5.2 está presente na máscara de qualidade por ter sido classificado como de baixa qualidade na série simulada, e ele será utilizado como exemplo para estimativa de um novo valor a partir da abordagem espaço temporal sugerida no presente estudo.

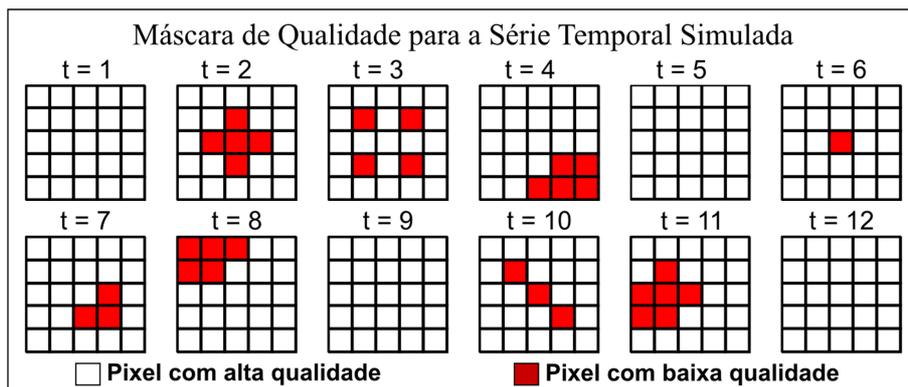


Figura 5.2 – Máscara de qualidade simulada ao longo de uma série temporal.

Sobre o ponto de interesse (pixel $(3,3)_6$) é definida uma janela de análise espacial equivalente aos 8 pixels vizinhos, de forma que esta janela percorra o conjunto de dados ao longo do tempo para adquirir os valores de NDVI dos pixels e compor as análises de regressão.

A qualidade dos pixels ao longo de uma série temporal de NDVI pode variar em virtude das mudanças nas condições atmosféricas e também nas geometrias de iluminação e visada no momento de aquisição da imagem. Sendo assim, para cada data posterior ou anterior à data de interesse (ou data “i”), a janela espaço-temporal seleciona somente os pixels classificados a partir dos metadados do produto NDVI/MODIS como de alta qualidade (Figura 5.3). Sendo assim, a constituição da janela espaço-temporal inicia-se com a verificação dos parâmetros de qualidade dos 8 pixels vizinhos ao pixel de interesse, de forma que somente os pixels considerados aptos, ou que não pertencem à máscara de qualidade, irão compor os dados para a análise de regressão.

A janela de análise ao longo do tempo tem como objetivo utilizar a informação do perfil temporal contida em datas anteriores e posteriores à data de análise. Desta forma, esta análise temporal analisa um conjunto de valores de NDVI para um intervalo de tempo equivalente a 11 datas, ou seja, para o pixel de interesse (pixel $(3,3)_i$), a janela de análise temporal será aplicada entre as datas “i-5” e “i+5”.

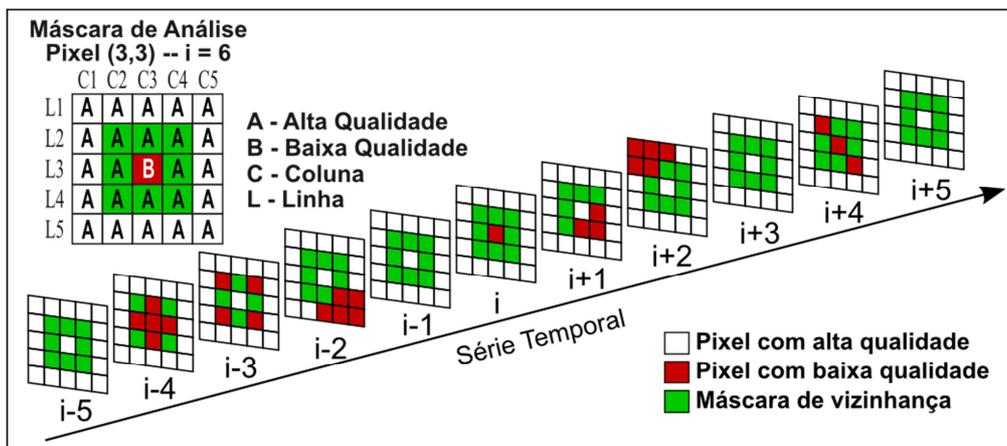


Figura 5.3 – Janela espaço-temporal para o pixel (3,3) na data 6 ($i = 6$).

Colditz et al. (2008) ressaltam que tanto o número como o intervalo de datas com baixa qualidade presentes em uma análise temporal influenciam na estimativa de novos dados. A quantidade de dados inaptos é representada pelo somatório das datas em que um pixel foi considerado como de baixa qualidade ao longo da série temporal, o que permite estimar a porcentagem de datas consideradas inadequadas ao longo do perfil temporal. Já o intervalo de datas (ou tamanho do *gap*) considera a maior sequência ao longo do tempo em que os dados são considerados inválidos. A combinação da janela espaço-temporal com a máscara de qualidade para o pixel de interesse permite estimar o número de datas consideradas inaptas e o tamanho do *gap* para cada pixel vizinho, e este resultado está ilustrado na Figura 5.4.

		Pixels com Baixa Qualidade			
		Perfil Temporal	Σ	%	Máximo Valor do GAP
Pixel	(2,2)	<input checked="" type="checkbox"/>	3	27	1
	(2,3)	<input checked="" type="checkbox"/>	1	9	1
	(2,4)	<input checked="" type="checkbox"/>	1	9	1
	(3,2)	<input checked="" type="checkbox"/>	1	9	1
	(3,3)	<input checked="" type="checkbox"/>	3	27	1
	(3,4)	<input checked="" type="checkbox"/>	1	9	1
	(4,2)	<input checked="" type="checkbox"/>	1	9	1
	(4,3)	<input checked="" type="checkbox"/>	2	18	1
	(4,4)	<input checked="" type="checkbox"/>	4	36	2
		i-5 i i+5	Tempo		

Figura 5.4 – Análise do número e intervalo de pixels classificados como de baixa qualidade ao longo do perfil temporal.

O novo valor do pixel (3,3)_i será obtido com base em uma análise de regressão entre os dados selecionados pela janela espaço-temporal do pixel de interesse. A Figura 5.5 ilustra um perfil temporal do pixel de interesse com um dos seus vizinhos, destacando os dados considerados inválidos pela máscara definida anteriormente. Entre os critérios metodológicos pré-definidos para compor a análise de regressão está a seleção de pares de dados considerados aptos, ou seja, os dados que em ambas as datas devem ser classificados como de alta qualidade. Desta forma, o exemplo ilustrado na Figura 5.5 demonstra que em três datas os pares de dados não serão utilizados na análise de regressão por serem considerados inválidos em pelo menos uma data.

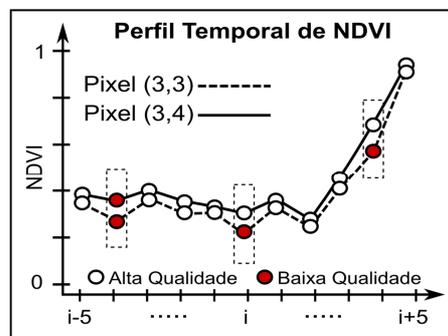


Figura 5.5 – Perfil temporal dos valores de NDVI para os pixels (3,3) e (3,4).

5.2.1. Regressão linear e estimativa de um novo valor

A função de relação entre duas variáveis pode ser expressa pela fórmula $Y = f(X)$, onde “Y” é a variável dependente e “X” a variável independente. Considerando uma variável preditora e uma função de relação linear, o modelo pode ser descrito como:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (5.1)$$

onde y_i é o valor da variável resposta na observação “i”; β_0 e β_1 são parâmetros; x_i é o valor da variável preditora na observação “i”; e ε_i é o termo erro aleatório, sendo que o método dos mínimos quadrados pode ser utilizado para obter os estimadores da regressão (NETER et al., 1996).

No presente estudo, a variável dependente é atribuída ao conjunto de dados selecionado para o pixel de interesse (pixel $(L,C)_i$). A variável independente é constituída pelos dados de um dos pixels vizinhos do ponto a ser estimado. Ressalta-se que o valor de um pixel considerado inapto pelos critérios de qualidade não é utilizado na regressão.

O valor do pixel de interesse é excluído da análise de regressão por ter sido classificado como inválido pelos critérios de qualidade. Porém, a estimativa de um novo valor para este pixel (pixel $(L,C)_i$) é considerada como uma nova observação (Y_{novo}) do conjunto de dados dependente. Da mesma forma, o valor da variável independente utilizada para estimar Y_{novo} é representado pelo valor do pixel vizinho ao pixel $(L,C)_i$ intitulado de X_{novo} .

Neter et al. (1996) destacam que um estimador não enviesado da variância do conjunto de dados com a inclusão de X_{novo} é expresso por:

$$s_{\text{novo}}^2 = \text{MSE} * \left[1 + \frac{1}{n} + \frac{(X_{\text{novo}} - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \quad (5.2)$$

onde S_{novo}^2 é a variância do conjunto de dados após a inclusão de um novo elemento; MSE é o erro quadrático médio; n equivale ao total de observações; X_{novo} é o novo elemento da variável independente; X é a variável independente e \bar{X} representa o valor médio de X .

Com base no que foi descrito anteriormente, todas as análises de regressão propostas ao longo da metodologia são realizadas com base nos seguintes critérios:

- (I) Utilizar somente pares de dados considerados válidos ou classificados como de boa qualidade a partir da consulta das informações de qualidade do produto MODIS;
- (II) Um número mínimo de quatro pares de dados válidos para a realização das análises de regressão, sendo que pelo menos dois

pares de dados válidos devem estar presentes tanto no período anterior como no posterior à data de interesse;

- (III) A análise de regressão que fornecer a menor variância para um novo dado é classificada como a de melhor ajuste para o conjunto de dados analisado, e será utilizada para recalculá-lo para o pixel de interesse.

A estimativa de um novo valor para o pixel $(L,C)_i$ é um processo iterativo, pois caso o critério II não seja atendido, o pixel é mantido na lista de dados inválidos e será analisado posteriormente. A Figura 5.6 exemplifica as análises de regressão possíveis para todos os vizinhos do Pixel $(3,3)_i$, e conforme os critérios de seleção pré-estabelecidos pela janela espaço-temporal, bem como pela análise de variância do conjunto de dados, a regressão linear obtida entre os pixels $(3,3)$ e $(3,4)$ será utilizada ao fim do processo para estimar o dado de interesse.

5.2.2. Estimativa dos valores de NDVI dos pixels com baixa qualidade

O objetivo deste tópico consiste em sintetizar os procedimentos que a técnica JR realiza para estimar um novo valor de NDVI para os pixels classificados como de baixa qualidade. A abordagem proposta para redução de ruídos está baseada em análise de regressão de dados selecionados por meio de uma janela espaço-temporal, com as seguintes etapas de execução:

- (A) Elaboração de uma máscara com os pixels e datas classificados como de baixa qualidade, ou seja, pontos que devem ter os valores de NDVI recalculados.
- (B) Entre todos os pixels identificados como de baixa qualidade é efetuado um processo de seleção aleatória para iniciar a análise de regressão. Este procedimento tem como propósito reduzir a influência de uma análise sequencial de pixels vizinhos, ou datas consecutivas, na estimativa dos novos valores de IV.

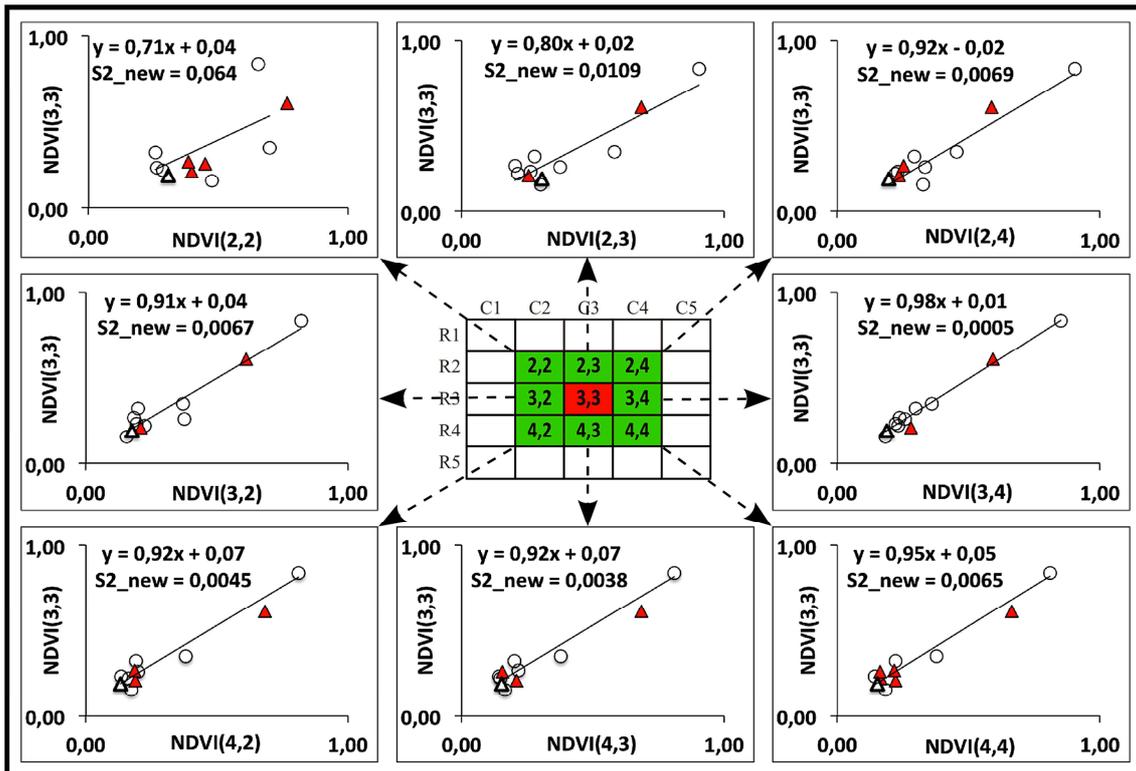


Figura 5.6 – Análise de regressão para o pixel (3,3); com uma janela temporal de (i-5 a i+5). Os círculos representam dados com qualidade satisfatória para as datas analisadas. Os triângulos significam que o dado foi considerado inválido para a regressão por possuir baixa qualidade em pelo menos uma data. O triângulo branco é o ponto de interesse.

- (C) Após a seleção do pixel $(L,C)_i$, a metodologia de reconstrução de um novo valor utiliza uma combinação de janelas no espaço e no tempo para selecionar o conjunto de dados a serem analisados na análise de regressão.
- (D) A janela espacial é representada por uma máscara de análise composta pelos oito pixels vizinhos ao pixel a ser analisado (pixel $(L,C)_i$); porém, somente serão utilizados os valores dos pixels que não tenham sido classificados como de baixa qualidade.
- (E) A janela temporal analisa juntamente com a data de interesse (i) um conjunto de até cinco datas posteriores (i+1, i+2, ..., i+5) e anteriores (i-1, i-2, ..., i-5) à data analisada. Este intervalo temporal é estratificado

em quatro intervalos correspondentes a duas datas (de $i-2$ a $i+2$), três datas (de $i-3$ a $i+3$), quatro datas ($i-4$ a $i+4$) e cinco datas ($i-5$ a $i+5$).

- (F) Para cada intervalo temporal são realizadas análises de regressão entre os pixels vizinhos ao ponto de interesse, e é selecionada a regressão que fornecer a menor variância para um novo valor a ser estimado. A partir da reta de regressão escolhida para cada intervalo temporal é estimado um novo valor de IV para o pixel de interesse (pixel $(L,C)_i$). O valor final que será atribuído ao ponto de interesse corresponde à mediana dos valores estimados para cada intervalo temporal analisado
- (G) O valor estimado para o pixel $(L,C)_i$ está apto para ser utilizado em outra análise de regressão, pois o processamento retorna à etapa “B” até que todos os pontos selecionados como de baixa qualidade sejam analisados.

Os procedimentos descritos nesta fase da metodologia foram implementados no software R (R CORE TEAM, 2013) conforme o código descrito no Apêndice E.

6 SIMULAÇÃO DE DADOS PARA ANÁLISE

Neste estudo foi elaborado uma simulação de dados de forma que seja possível comparar os resultados obtidos com uma referência. Esta comparação possibilitará uma análise objetiva da eficiência e do desempenho da metodologia na estimativa de valores de NDVI em séries temporais.

A simulação de dados envolveu quatro etapas: seleção de dados classificados com qualidade satisfatória no produto MOD13Q1; degradação dos dados; reconstrução da série temporal; e por fim a comparação com os dados de referência (Figura 6.1).

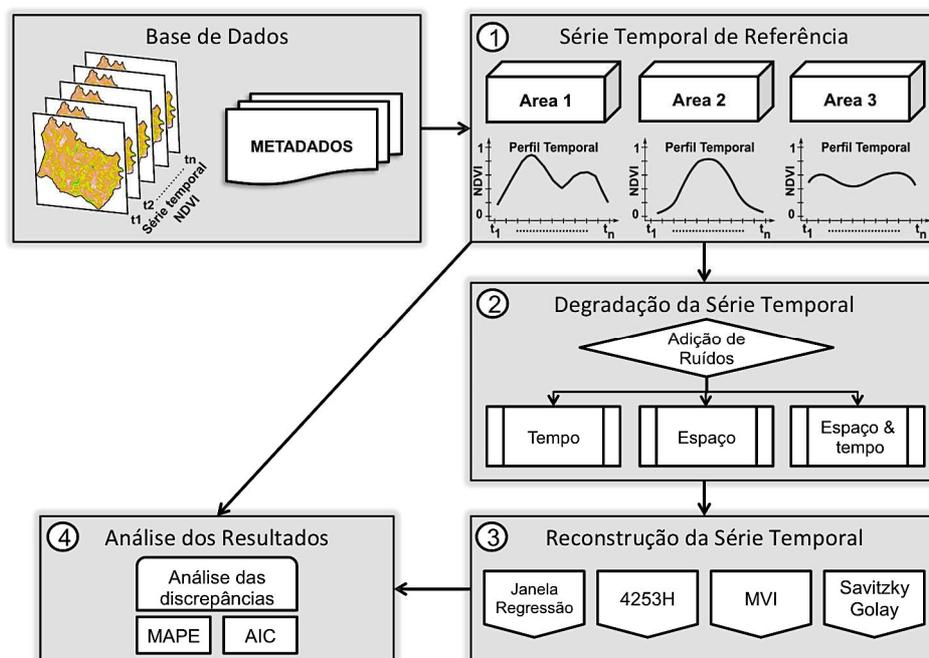


Figura 6.1 – Etapas para simulação de dados e avaliação do desempenho das técnicas de filtragem na estimativa de novos valores.

6.1. Série temporal de referência

Apesar da variabilidade das classes de cobertura do solo presente no município de Ipuã, pode-se observar com base nos dados do Canasat (RUDORFF et al., 2010) um aumento das áreas destinadas à cana-de-açúcar ao longo dos últimos anos. O mapeamento da cobertura do solo do município

demonstra que as áreas destinadas à cultura de cana-de-açúcar passaram de 45% para 70% entre os anos de 2003 e 2010 (Figura 6.2).

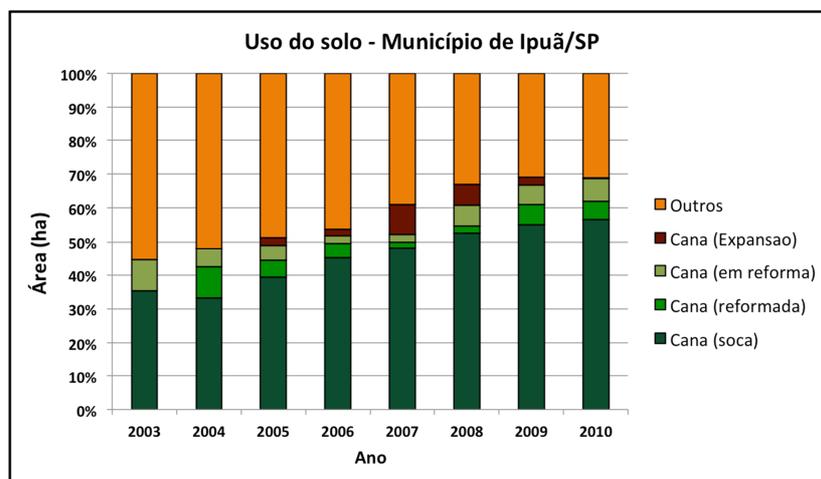


Figura 6.2 – Cobertura do solo do município de Ipuã entre os anos 2003 – 2010.

Fonte: Rudorff et al. (2010).

A base de dados analisada no presente estudo é constituída por uma série temporal do produto MOD13Q1 entre os anos de 2003 e 2010. Em virtude da variabilidade de classes e dimensões dos alvos agrícolas presentes no município de Ipuã, é possível que mais de uma cultura agrícola seja representada por um mesmo pixel MODIS com uma resolução espacial de 250 m. Assim, a distribuição de áreas testes sobre a área de estudo visa avaliar o desempenho da metodologia em relação à variabilidade espacial das classes de cobertura do solo, além do comportamento temporal distinto entre alvos perenes e semi-perenes.

Por isso, foram definidas três áreas testes que representam as classes de uso e cobertura do solo na área de estudo. A localização geográfica de cada área teste está relacionada à seleção de locais que apresentam áreas homogêneas ou heterogêneas. Para a delimitação espacial das áreas foram observadas as informações de qualidade dos pixels disponíveis no produto MOD13Q1, de forma que para todo o conjunto de dados da área teste os valores de IV estivessem classificados como de excelente qualidade.

A área teste “A” foi localizada sobre um alvo de cana-de-açúcar, pois esta classe é a mais representativa em toda a área de estudo. As demais áreas testes foram selecionadas com o objetivo de representar a diversidade das classes de cobertura do solo no município, de forma que a Área “B” está localizada sobre uma região onde a cobertura do solo é mais fragmentada e diversificada, e a área “C” foi localizada sobre uma região que combina área agrícola e floresta (Figura 6.3).

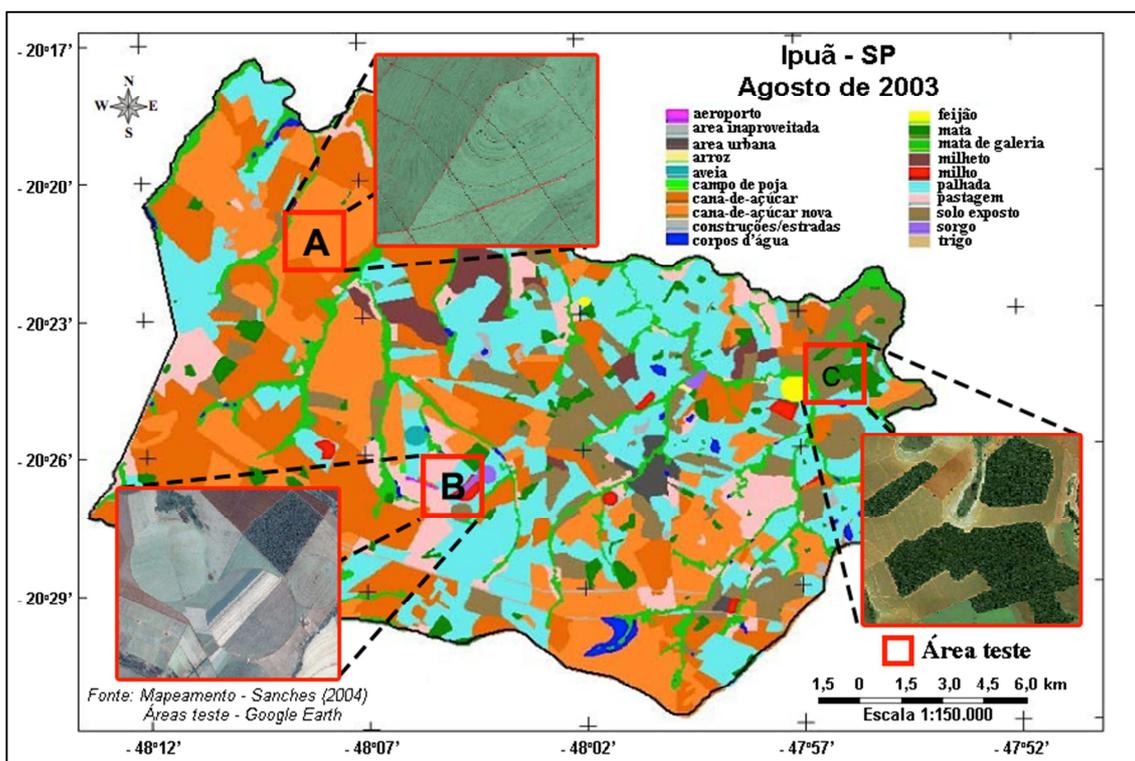


Figura 6.3 – Localização das áreas de teste.

Fonte: adaptado de Sanches (2004).

6.2. Degradação da série temporal

As séries temporais de NDVI referente aos pixels presentes nas áreas testes apresentam excelente confiabilidade conforme o produto MOD13Q1, e por isso foram consideradas como referência no presente estudo.

A partir dos dados de referência, foram simuladas série temporais com baixa qualidade por meio da introdução de ruídos, conforme proposto por Adami (2010) e Hird e Mcdermid (2009). Esta simulação de dados tem como objetivo

permitir uma avaliação da acurácia e desempenho do processo de redução de ruídos nos pontos simulados com baixa qualidade.

A primeira etapa do processo de simulação de ruídos consiste na elaboração de métodos de amostragem de pixels e datas a serem analisados dentro de uma série temporal de NDVI. Esta amostragem tem como objetivo verificar nos resultados das técnicas de filtragem a influência espaço-temporal dos ruídos presentes na série temporal, principalmente no método de redução de ruídos que utiliza a informação espacial em sua análise. Desta forma, o processo de amostragem dos ruídos foi subdividido em 4 tipos: (I) amostragem de datas ao longo da série temporal (amostragem temporal); (II) amostragem de pixels e datas (amostragem espaço-temporal); (III) amostragem de um grupo de pixels para uma mesma data (amostragem espacial); (IV) amostragem de um grupo de datas em um perfil temporal (amostragem temporal direcionada). A Figura 6.4 ilustra os procedimentos de amostragem.

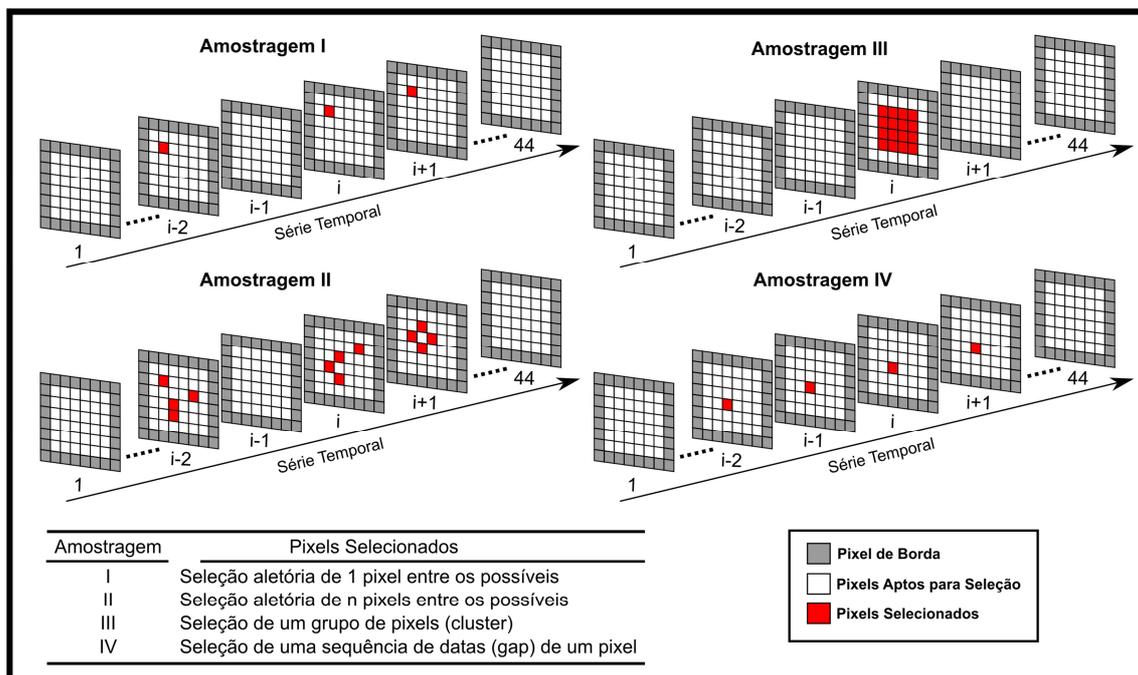


Figura 6.4 – Ilustração dos tipos de amostragem simuladas para avaliar as técnicas de filtragem utilizadas neste estudo.

O método de amostragem “I” consiste na aplicação de ruído em uma porcentagem das datas que compõem a série temporal de um pixel da área teste, ou seja, consiste em uma degradação temporal nos valores de NDVI para um determinado pixel. Destaca-se que os pixels presentes nas bordas da área teste não foram utilizados nos procedimentos de amostragem. Sendo assim, dos 64 pixels (matriz de 8x8 pixels) presentes em cada área teste, somente 36 foram considerados aptos para os métodos de amostragem. Além disso, para reduzir a influência da ausência de dado anterior ou posterior à data a ser analisada, foi estipulado que tanto as três primeiras como as três últimas datas que compõem a série temporal não serão utilizadas para a amostragem.

O método de amostragem “II” é uma combinação entre a degradação temporal e espacial. Neste método, é utilizado o mesmo intervalo de datas e pixels considerados aptos para amostragem como descrito anteriormente; porém, para cada data selecionada, é aplicado ruído para mais de um pixel da imagem. A degradação espacial deste método consiste em selecionar aleatoriamente uma porcentagem de pixels para aplicação de ruídos, que no presente estudo equivale a 5%, 10% e 20% dos pixels aptos para seleção na área teste.

O método de amostragem “III” visa avaliar o desempenho da metodologia na estimativa de um novo valor de NDVI em relação à degradação de um grupo de pixels (ou cluster). O objetivo é simular a situação em que uma determinada região da imagem foi classificada com baixa qualidade em virtude da presença de nuvem ou de sua sombra. A partir de uma data selecionada aleatoriamente entre o intervalo disponível para análise, são introduzidos ruídos em um grupo de pixels com as dimensões de 2x2, 3x3, 4x4, 5x5 e 6x6. A posição dos clusters de pixels dentro da área teste é definida aleatoriamente entre o conjunto de pixels considerados aptos para análise.

O método de amostragem “IV” simula uma sequência de datas com baixa qualidade ao longo de um perfil temporal, e por isso este método seleciona aleatoriamente um pixel entre os possíveis e avalia o tamanho do *gap* na estimativa de novos valores de NDVI. No presente estudo, foram simulados

gaps variando de 2 a 6, ou seja, uma sequência de 2 até 6 datas consecutivas que foram classificadas com baixa qualidade ao longo de um perfil temporal.

A segunda etapa do processo de simulação de ruídos consiste na introdução de ruídos aos valores originais de NDVI amostrados pelos métodos descritos anteriormente. No presente estudo foram definidos 3 níveis de ruídos (10%, 30% e 50%) que representam a porcentagem que será adicionada ou subtraída aleatoriamente ao valor original de NDVI.

A Figura 6.5 ilustra um exemplo de uma degradação do perfil temporal de NDVI composto por 44 imagens do produto MOD13Q1, onde foram aplicados 3 níveis de ruídos em 12 datas da série temporal de um pixel.

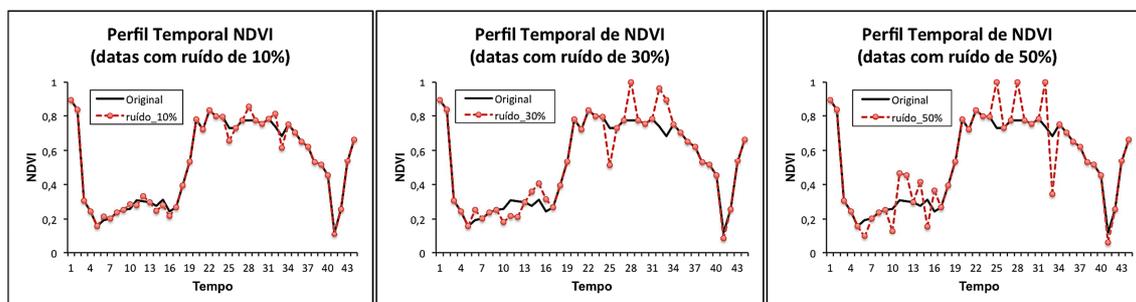


Figura 6.5 – Perfil Temporal de NDVI original e degradado com 3 níveis de ruídos para o pixel (3,3).

6.3. Reconstrução da série temporal

Os tipos de amostragens definidos anteriormente combinam níveis de degradações do dado original tanto no tempo como no espaço, e essas amostragens têm como objetivo simular as diversas situações em que os dados presentes no produto MOD13Q1 podem apresentar baixa qualidade. Além disso, o monitoramento entre o dado de referência e as condições que os ruídos foram simulados possibilita uma avaliação objetiva da acurácia e do desempenho do método de redução de ruído.

Todos os pixels simulados com ruídos terão os seus respectivos valores de NDVI re-estimados pela metodologia proposta neste estudo. Além da técnica JR, também serão utilizados os filtros 4253H twice; Savitzky–Golay; e Valor

Médio da Iteração para redução dos ruídos simulados em séries temporais de NDVI. O objetivo é comparar o desempenho da metodologia proposta nesta pesquisa com o de outras técnicas de filtragem descritas na literatura.

6.4. Análise dos resultados

O objetivo desta etapa da pesquisa consiste em avaliar quantitativamente o desempenho dos filtros escolhidos para a redução dos ruídos simulados. Para isto, foram avaliadas as discrepâncias obtidas entre o dado original e o filtrado.

Musial et al. (2011) enfatizam a importância da escolha das medidas de dispersão que definem o melhor ajuste entre o valor original presente na série temporal e o resultado do processo de filtragem. Entre as técnicas tradicionalmente utilizadas para medir o ajuste dos resultados obtidos destaca-se a raiz do erro médio quadrático (*Root Mean Square Error* - RMSE). Porém, Willmott e Matsuura (2005) sugerem o uso da média do erro absoluto (*Mean Absolute Error* - MAE) na avaliação e comparação da magnitude das discrepâncias, pois esta medida de dispersão proporciona um melhor indicador da qualidade do ajuste ou do desempenho do filtro na reconstrução de uma série temporal, sendo que o MAE é menos influenciado por valores anômalos do que o RMSE.

As técnicas descritas acima para estimar a qualidade do ajuste medem as diferenças entre os correspondentes dados originais e filtrados, de forma que quanto maior o valor da medida de dispersão pior o resultado do ajuste. No entanto, estas técnicas diferem na ênfase que dão à avaliação, pois o MAE enfatiza o desempenho geral da média, enquanto que o RMSE penaliza grandes diferenças individuais.

Porém, as diferentes escalas presentes entre séries temporais podem induzir a comparações inexpressivas. Por isso, no presente estudo foi utilizado o erro percentual médio absoluto (*Mean Absolute Percentage Error* - MAPE) para comparar o desempenho das técnicas. Willmott e Matsuura (2005) classificam o erro médio absoluto como uma medida de dispersão não ambígua da

magnitude do erro médio. A formulação da medida de dispersão MAPE pode ser expressa por:

$$\text{MAPE} = \frac{1}{n} * \sum_{i=1}^n \left(\left| \frac{IV_{i_fit} - IV_{i_obs}}{IV_{i_obs}} \right| * 100 \right) \quad (6.1)$$

onde o n é o número de pontos avaliados, IV_{i_fit} é valor do índice de vegetação filtrado e IV_{i_obs} é o valor original do índice de vegetação presente na série temporal.

Outra medida estatística utilizada para avaliar o desempenho das técnicas de filtragem foi o *Akaike Information Criterion* (AIC) (AKAIKE, 1973) que ajusta um modelo entre os dados originais e filtrados. Segundo Atkinson et al. (2012) o AIC pode ser empregado para avaliar o desempenho das técnicas de filtragem utilizadas, e calculado pela seguinte expressão:

$$\text{AIC} = 2 * k - 2 * \log(L) \quad (6.2)$$

onde k é o número de parâmetros do modelo, $\log(L)$ é o logaritmo da máxima verossimilhança estimada para o modelo. Pequenos valores de AIC indicam um melhor desempenho do modelo.

7 RESULTADOS E DISCUSSÃO

7.1. Análise da base de dados

As informações sobre o número e intervalo de datas classificadas com baixa qualidade são essenciais para realizar uma simulação de ruídos adequada com a qualidade da informação presente na base de dados. Roy et al. (2002) ressaltam que as informações de qualidade disponíveis nos produtos MODIS contribuem para uma adequada interpretação e aplicação dos dados, de forma que os produtos classificados como de baixa qualidade devem ser utilizados com cautela. Sendo assim, a primeira análise realizada sobre a base de dados consistiu em verificar a incidência de dados classificados com baixa qualidade ao longo do tempo.

A Figura 7.1 apresenta a distribuição espacial da qualidade dos pixels após uma reclassificação dos metadados do produto MOD13 (JUSTICE et al., 2002). Foram considerados pixels de baixa qualidade todos aqueles identificados com valores de confiabilidade (produto *reliability*) iguais a 2 ou 3 (relativo à presença de gelo/neve ou nuvem, respectivamente). Após a reclassificação, os dados foram integrados para compor mapas de qualidade para cada ano de estudo. A integração de dados consistiu em computar o percentual de dados considerados como de baixa qualidade ao longo de 23 imagens que compõem um ano de dados do produto MOD13Q1. As classes do mapa de qualidade foram subdivididas em intervalos de 0 a 5%, 5 a 10%, 10 a 15% e de 15 a 20% de datas consideradas com baixa qualidade ao longo de um ano.

A série temporal de dados referente à área de estudo não demonstra um padrão espaço-temporal em relação à qualidade dos pixels. Porém, os anos de 2004 e 2010 apresentam os menores percentuais de datas classificadas com baixa qualidade para toda a área de estudo.

A Figura 7.2 ilustra o percentual de pixels da área de estudo para cada classe do mapa de qualidade, e pode-se observar que pelo menos 50% da área de estudo foi classificada no menor intervalo de qualidade ao longo do período de

estudo. Este intervalo indica que no máximo duas datas foram classificadas com baixa qualidade entre as 23 imagens que compõem a série de dados. Entretanto, a base de dados relativa ao ano de 2009 foi a que apresentou a maior incidência de datas classificadas com baixa qualidade ao longo da área de estudo.

É possível constatar que a área de estudo apresenta em alguns períodos imagens com uma maior porcentagem de pixels de baixa qualidade, como demonstrado para o ano de 2009, onde cerca de 25% dos pixels presentes na área de estudo possui de 10% a 20% da série temporal classificada como de baixa qualidade. O melhor período pode ser verificado no ano 2004, onde aproximadamente 99% dos pixels presentes na área de estudo apresentaram uma taxa de datas classificadas como inadequadas inferior a 5%. Como comentado por Atzberger e Eilers (2011a), em alguns casos a região de estudo é significativamente afetada por grandes períodos com imagens de baixa qualidade, principalmente pela presença de nuvens, as quais são encontradas com grande frequência nos estádios de desenvolvimento da cultura, fazendo com que seja necessário o emprego de dados filtrados.

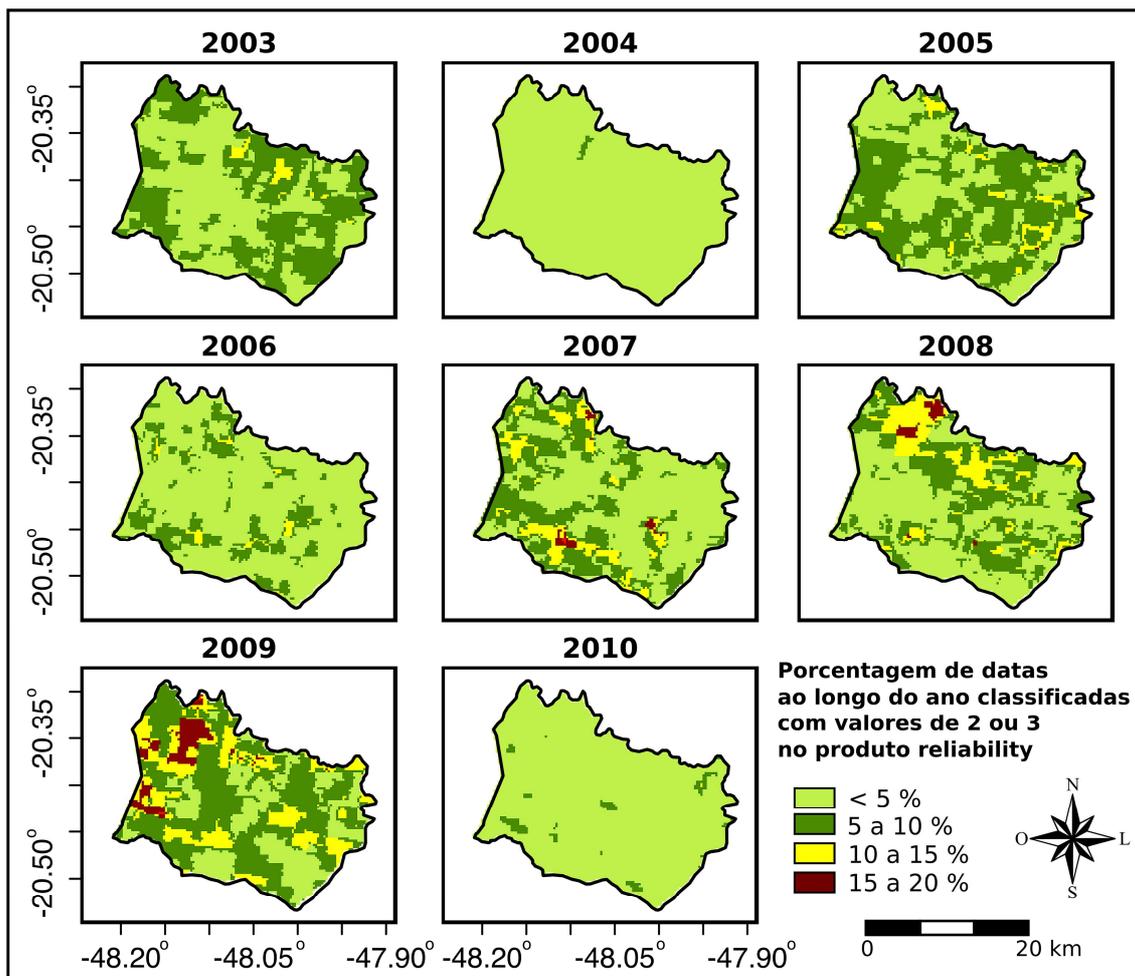


Figura 7.1 – Espacialização da porcentagem das datas obtidas ao longo de um ano (23 imagens) em que os pixels foram classificados com valores 2 ou 3 no produto *reliability*.

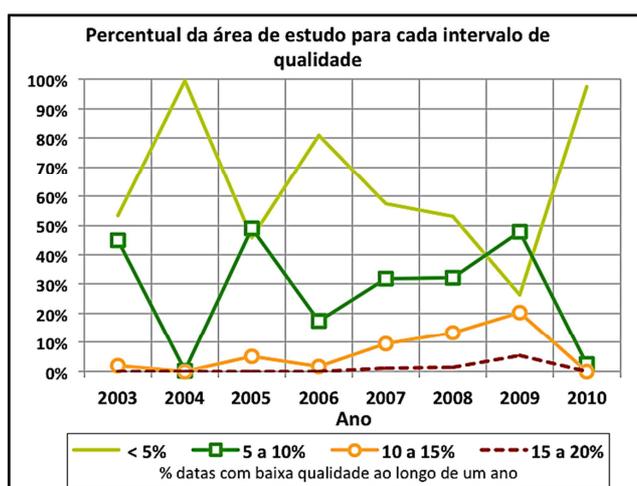


Figura 7.2 – Percentagem da área de estudo distribuída nos intervalos de qualidades (< 5%, 5 a 10%, 10 a 15% e 15 a 20% de datas com baixa qualidade) da série temporal para cada ano safra.

Além de verificar a intensidade de datas inadequadas para cada perfil temporal, é interessante avaliar como a qualidade das datas está distribuída ao longo da série. Colditz et al. (2008) destacam que existe uma relação inversa entre o tamanho do intervalo de datas classificadas como de baixa qualidade (ou tamanho do *gap*) e a confiança no processo de interpolação dos dados.

A Figura 7.3 demonstra o percentual de pixels da área de estudo em máximo *gap* identificado ao longo de um ano. O maior tamanho de *gap* encontrado na área de estudo foi equivalente a 4 datas na base de dados dos anos de 2005 a 2009. Novamente, o ano safra 2004 apresentou o melhor cenário em relação à qualidade da série temporal, pois além de conter o menor número de datas inválidas, também possui o menor *gap* para cerca de 99% dos pixels que compõem a área de estudo.

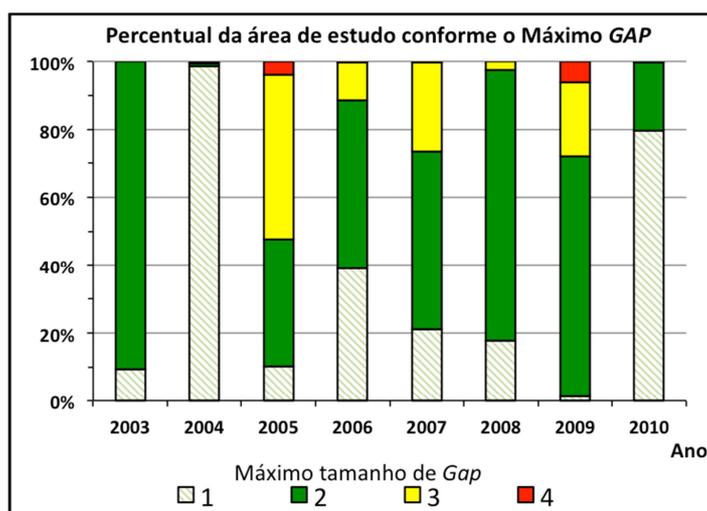


Figura 7.3 – Percentagem de pixels da área de estudo em relação ao Máximo tamanho de *gap* obtido ao longo de um ano.

7.2. Avaliação de desempenho para a área teste “A”

7.2.1. Avaliação para simulação espaço-temporal (Amostragens I e II)

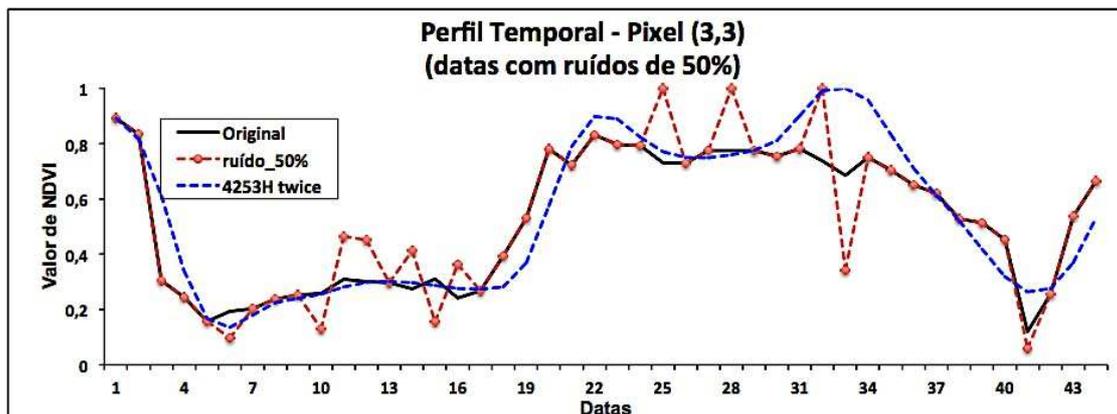
A análise da base de dados referente ao produto MOD13Q1 demonstrou que a maior porcentagem de datas classificadas com baixa qualidade ao longo de um ano não ultrapassou 20%, ou seja, de um perfil temporal de 23 imagens um

número inferior a 5 datas entre elas possui baixa qualidade. Além disso, não foram classificadas como de baixa qualidade mais do que 4 datas consecutivas sobre os perfis temporais analisados. Esses valores serviram de base para as simulações das séries temporais e para a avaliação do desempenho da metodologia proposta nesta pesquisa, na qual foi estipulado um percentual de 30% para a degradação temporal simulada nos tipos de amostragem “I” e “II”.

Em um primeiro momento, a área teste “A” foi selecionada entre as 3 áreas pré-definidas para avaliar o desempenho entre as técnicas de redução de ruídos escolhidas neste estudo. A escolha desta área deve-se principalmente à disponibilidade de um maior conjunto de imagens em que os pixels foram classificados como de boa confiabilidade, além de todos os pixels delimitados pela área “A” estarem localizados sobre uma cobertura do solo homogênea.

A série temporal da área “A” é composta por 44 datas em que não consta a presença de gelo, neve ou nuvem, conforme as informações disponíveis no produto *reliability* (Figura 7.4). Este conjunto de dados representa aproximadamente dois anos de imagens do produto MOD13Q1, e esta maior disponibilidade de dados, quando comparado às áreas “B” e “C”, que possuem 25 e 22 datas, respectivamente, proporciona simulações e/ou amostragem mais representativas da dinâmica da classe de cobertura do solo de maior ocorrência na área de estudo.

A amostragem “I” propõe uma simulação de ruídos ao longo do tempo equivalente a 30% das datas disponíveis de uma curva temporal de NDVI, ou seja, foram aplicados ruídos em 12 datas entre as 44 que compõem a série temporal de um pixel da área “A”. Para cada nível de ruído simulado foi analisado o desempenho das técnicas de filtragem utilizadas no presente estudo. A Figura 7.5 ilustra o procedimento para a avaliação de desempenho, por meio da medida de dispersão MAPE, do filtro 4253H twice sobre uma curva degradada com o nível de ruído equivalente a 50%.



Valor de NDVI para cada data

Datas	1	2	3	4	5	6*	7	8	9	10*	11*	12*	13	14*	15*	16*	17	18	19	20	21	22
Original	0,89	0,84	0,3	0,24	0,16	0,19	0,2	0,24	0,25	0,26	0,31	0,3	0,3	0,28	0,31	0,24	0,27	0,39	0,53	0,78	0,72	0,83
ruído_50%	0,89	0,84	0,3	0,24	0,16	0,1	0,2	0,24	0,25	0,13	0,46	0,45	0,3	0,41	0,16	0,36	0,27	0,39	0,53	0,78	0,72	0,83
4253H twice	0,89	0,82	0,61	0,34	0,17	0,13	0,18	0,22	0,24	0,25	0,28	0,3	0,3	0,3	0,29	0,28	0,27	0,28	0,37	0,58	0,79	0,9
EPA						31,6%				3,8%	9,7%	0,0%		7,1%	6,5%	16,7%						
Datas	23	24	25*	26	27	28*	29	30	31	32*	33*	34	35	36	37	38	39	40	41*	42	43	44
Original	0,8	0,79	0,73	0,73	0,77	0,77	0,78	0,76	0,78	0,74	0,68	0,75	0,71	0,65	0,62	0,53	0,51	0,45	0,12	0,26	0,54	0,66
ruído_50%	0,8	0,79	1,1	0,73	0,77	1,16	0,78	0,76	0,78	1,11	0,34	0,75	0,71	0,65	0,62	0,53	0,51	0,45	0,06	0,26	0,54	0,66
4253H twice	0,89	0,82	0,77	0,75	0,75	0,76	0,77	0,81	0,9	0,99	1,02	0,96	0,83	0,71	0,61	0,52	0,42	0,32	0,26	0,28	0,37	0,53
EPA			5,5%			1,3%				33,8%	50,0%									116,7%		

* data que foi simulado ruído de 50%

** Erro percentual Absoluto (EPA) = | (original - 4253H twice) / original | * 100

MAPE = 23,55%

Figura 7.5 – Exemplo da avaliação de desempenho do filtro 4253H twice sobre um perfil temporal simulado com ruído de 50%.

A Tabela 7.1 apresenta uma análise descritiva da qualidade do ajuste dos filtros (4253H twice, Mean Value Iteration (MVI), Savitzky–Golay (SG) e Janela de Regressão (JR)) para cada nível de ruído simulado. A Figura 7.6 ilustra a variação do MAPE, por meio de um gráfico violino, para cada técnica de filtragem e nível de ruído simulado na amostragem “I”. De acordo com Hintze e Nelson (1998), o gráfico violino combina em um único gráfico as informações do boxplot e da densidade kernel referentes ao conjunto de dados analisados.

Willmott e Matsuura (2005) enfatizam que um menor valor de MAPE representa um melhor desempenho para o conjunto de dados analisados, e a análise das discrepâncias obtidas na amostragem “I” indica que a técnica JR apresentou o melhor desempenho entre as técnicas avaliadas. A média dos valores de MAPE para o filtro JR foi inferior a 7% para os três níveis de ruído aplicados.

Tabela 7.1 – Análise descritiva dos valores de MAPE para as 1000 iterações realizadas na amostragem “I”.

Nível de Ruídos	Medidas	Técnicas Avaliadas (valores em %)			
		4253H twice	MVI	SG	JR
10%	Min.	2,6	4,5	3,4	1,2
	Mediana	11,4	15,2	8,3	6,1
	Média	14,2	16,8	9,2	6,9
	Max.	44,1	49,8	22,3	28,4
30%	Min.	4,3	3,9	9,5	1,2
	Mediana	13,7	16,9	16,4	6,1
	Média	16,2	18,6	16,8	6,9
	Max.	49,7	53,7	32,3	28,4
50%	Min.	3,1	6,5	15,8	1,2
	Mediana	15,2	18,6	25,4	6,1
	Média	17,3	20,2	25,7	6,9
	Max.	49,0	55,3	40,6	28,4

O valor médio do índice AIC obtido após as 1000 iterações demonstra um desempenho constante da técnica JR para os 3 níveis de ruído (Figura 7.6). Os filtros 4253H twice, MVI e SG apresentaram queda no desempenho com o aumento da intensidade de ruído aplicado.

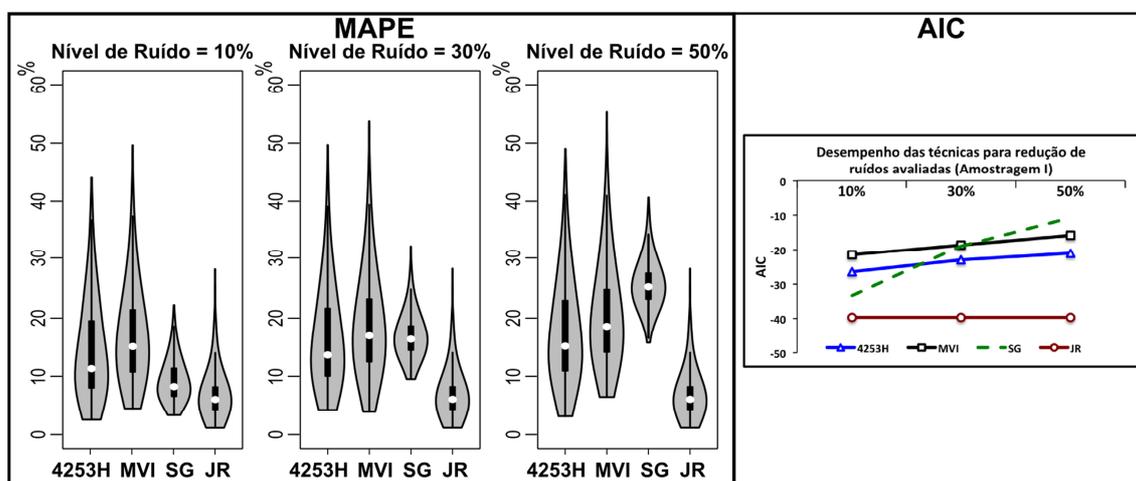


Figura 7.6 – Acurácia das técnicas de redução de ruídos (4253H twice – MVI – SG – JR) para a amostragem “A” com base no MAPE (erro percentual médio absoluto) e valores médios de AIC obtido entre 1000 iterações.

No procedimento de amostragem “II” foi realizada uma simulação de ruídos no tempo e no espaço, de forma que além das 12 datas selecionadas ao longo da

série temporal, também foi amostrado um número de pixels equivalente a 5%, 10% e 20% da cena. Este procedimento foi repetido 1000 vezes para cada simulação da amostragem “B” e foram calculadas as respectivas medidas de dispersão, de forma que os menores valores de MAPE e AIC representam o melhor ajuste entre os valores de NDVI original e filtrado.

A Figura 7.7 apresenta a variação do MAPE obtido por cada técnica de filtragem, nível de ruído e degradação espacial simulada. Pode-se observar que as técnicas apresentaram desempenhos semelhantes entre as degradações espaciais simuladas, com exceção do filtro SG, que apresentou queda no desempenho para a degradação espacial igual a 20%.

A técnica JR apresentou o melhor desempenho entre as técnicas avaliadas independentemente do nível de ruído ou degradação espacial simulada. Os valores das discrepâncias obtidos pela técnica JR na amostragem “II” foram semelhantes aos obtidos na amostragem “I”, onde o valor médio de MAPE foi inferior a 7% em todas as simulações realizadas.

Os valores da estatística AIC são influenciados pelo número de amostras contidas em cada modelo analisado, de forma que quanto maior o número de amostras, menor será o valor do AIC. Sendo assim, as curvas presentes na Figura 7.8 devem ser comparadas levando em consideração o tamanho da base de dados analisada em cada degradação espaço-temporal simulada. Uma amostragem de 20% dos pixels em uma cena resulta em um conjunto de dados maior do que os percentuais de 10% ou 5% e, por isso, um aumento na degradação espacial resultou em queda dos valores de AIC.

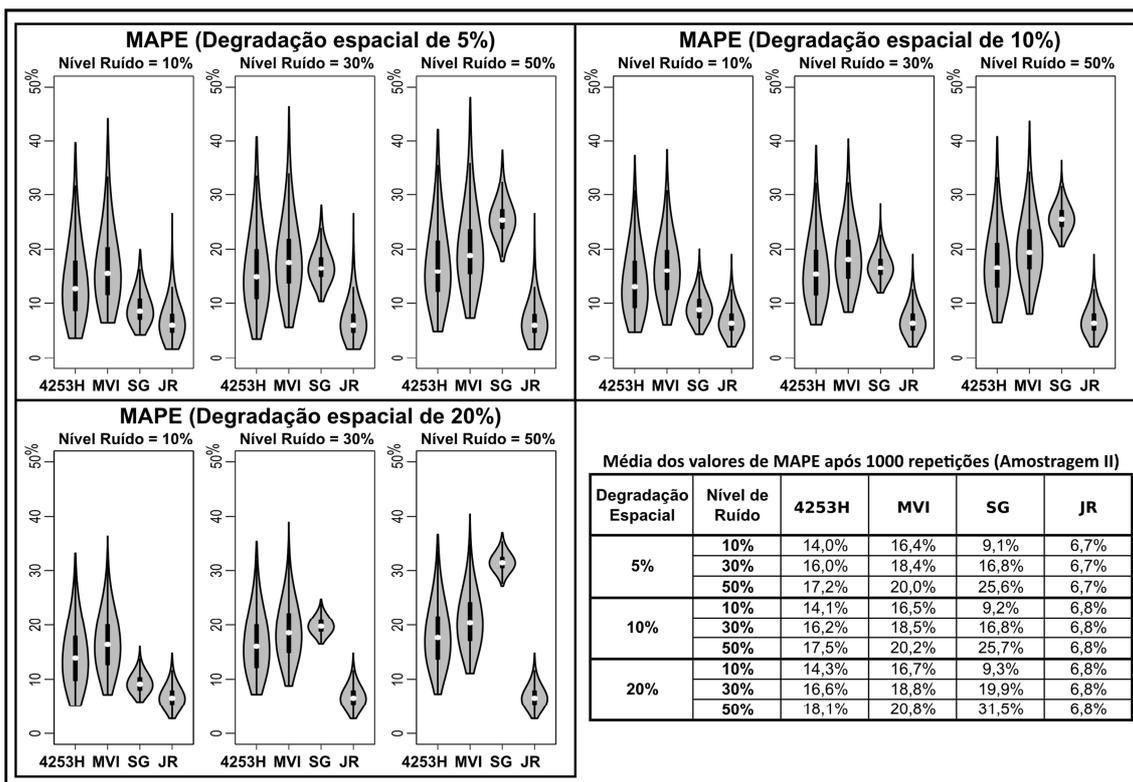


Figura 7.7 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações realizadas na amostragem “II”. Entretanto, o comportamento das curvas da Figura 7.8 demonstra que a hierarquia da acurácia obtida para cada técnica analisada foi praticamente a mesma, independentemente da degradação espacial ou intensidade de ruídos simulados.

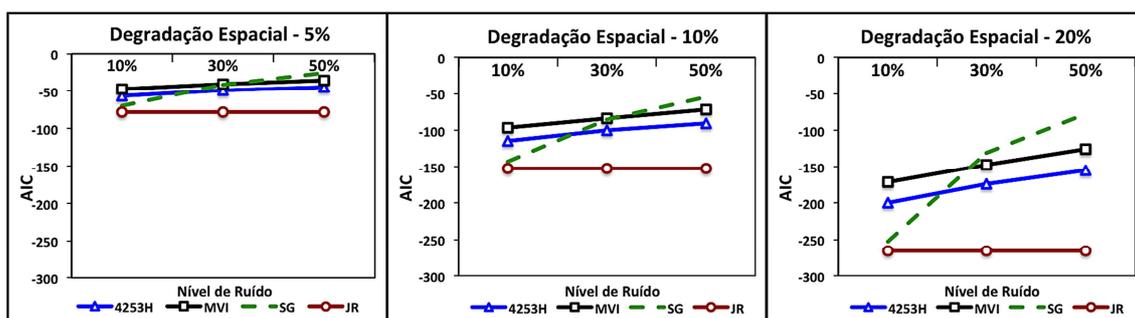


Figura 7.8 – Acurácia das técnicas de redução de ruídos (4253H twice – MVI – SG – JR) para a amostragem “II” com base no valor médio do AIC (Akaike Information Criterion) obtido entre 1000 iterações.

Os valores de AIC demonstram que a técnica JR forneceu o melhor ajuste nas estimativas de novos valores de NDVI para todas as simulações das

amostragens “I” (Figura 7.8), e o comportamento desta técnica apresentou um ajuste constante independentemente do número de pixels classificados como de baixa qualidade em uma mesma cena. Por outro lado, as demais técnicas para redução de ruído analisadas apresentaram queda da acurácia com o aumento do nível de ruído aplicado, com destaque para o filtro SG que apresentou a maior queda de desempenho.

Ambas as medidas de dispersão (MAPE e AIC) que avaliaram o desempenho das técnicas de filtragem apresentaram o mesmo comportamento para as metodologias de amostragem “I” e “II”. Isto se justifica pelo fato de que os filtros 4253H twice, MVI e SG utilizam somente a informação temporal da curva, de forma que a presença de um ou mais pixels degradados para a mesma data não influencia o comportamento desses filtros. Embora a técnica JR utilize a informação espacial para estimar o valor de um determinado ponto, a degradação de 5%, 10% ou 20% dos pixels para uma mesma data não influenciou a qualidade do ajuste desta técnica para a área teste “A”.

Pode-se observar que quanto maior a intensidade do ruído aplicado, pior é o desempenho das técnicas de filtragem 4253H twice, MVI e SG, onde este comportamento pode ser verificado em todas as medidas estatísticas que avaliaram as discrepâncias obtidas. A técnica JR não foi influenciada pelos níveis de ruído simulados, e isto se justifica pelo fato de o ponto classificado com baixa qualidade (ou com ruído) não ser utilizado na estimativa do novo valor (OLIVEIRA et al., 2012).

O teste não paramétrico de Wilcoxon pareado foi utilizado para comparar o desempenho das técnicas de redução de ruídos analisadas no presente estudo. Para isto, foram avaliadas as medidas de posição para as 1000 iterações realizadas, de forma que os valores de MAPE de uma técnica foram comparados e ordenados em relação aos valores obtidos pelos outros filtros. A Tabela 7.2 apresenta uma ordenação do desempenho das técnicas de filtragem com base no teste de Wilcoxon pareado. O filtro JR foi considerado estatisticamente o mais eficiente nas três intensidades de ruídos simulados no estudo, independentemente da técnica de amostragem empregada.

Tabela 7.2 – Desempenho das técnicas de redução de ruídos com base no Teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações.

Nível de Ruído	Desempenho (decrecente)	Amostragem I	Amostragem II (Degradação Espacial)		
			5%	10%	20%
10%	1	JR	JR	JR	JR
	2	SG	SG	SG	SG
	3	4253H	4253H	4253H	4253H
	4	MVI	MVI	MVI	MVI
30%	1	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H
	3	SG	SG	SG	MVI
	4	MVI	MVI	MVI	SG
50%	1	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG

Obs.: Todos os desempenhos são significativamente diferentes de acordo com o teste de Wilcoxon pareado e nível de significância de 5%.

Para o menor nível de ruído simulado, o filtro SG apresentou um desempenho mais próximo à técnica JR (Figuras 7.6 e 7.8); porém, o filtro SG não manteve o mesmo desempenho com o aumento dos níveis de ruídos simulados, pois da segunda posição passou para a última no ordenamento dos desempenhos entre os níveis de ruídos.

7.2.2. Avaliação de desempenho para simulação espacial (Amostragem III)

Trabalhos como o de Asner (2001), Asner e Alencar (2010) e Sano et al. (2007) destacam que a possibilidade de adquirir uma cena com extensas áreas cobertas por nuvem aumenta nos períodos como maior regime pluviométrico. Desta forma, a amostragem “III” simulou baixa qualidade em grupos de pixels (cluster) com dimensões de 2x2, 3x3, 4x4, 5x5 e 6x6, com o objetivo de avaliar a influência da dimensão espacial dos clusters na estimativa de novos valores. O processo de amostragem consistiu em selecionar aleatoriamente uma data na base de dados, degradar com três níveis de ruídos (10%, 30% e 50%) os

valores de NDVI referentes aos pixels de cada cluster, e por fim avaliar o desempenho das técnicas de filtragens na estimativa de novos valores para os pixels simulados. A Figura 7.9 ilustra os resultados de MAPE, obtidos após 1000 iterações para a amostragem “III”.

O número de dados amostrados em cada cluster são diferentes, de forma que a amostragem de um cluster 2x2 simula ruídos em 4 pixels, enquanto que um cluster 6x6 simula ruído em 36 pixels. Esta diferença no tamanho do conjunto de dados analisado para cada cluster refletiu em queda da variância dos valores de MAPE após 1000 iterações. Esta redução da variância pode ser observada nas amplitudes dos gráficos violinos que ilustram os valores de MAPE obtidos pelos filtros 4253H twice, MVI e SG nos três níveis de ruído simulado (Figura 7.9). Particularmente, essas técnicas de filtragem analisam cada pixel do cluster independentemente do ruído aplicado no pixel vizinho, pois não utilizam da informação espacial para a redução dos ruídos simulados.

Novamente, um teste não paramétrico de Wilcoxon pareado foi utilizado para comparar o desempenho das técnicas de redução de ruídos analisadas em relação à degradação espacial simulada. Considerando os resultados do teste de Wilcoxon, e um nível de significância de 5%, pode-se afirmar que a técnica JR apresentou o melhor desempenho para todos os clusters simulados entre as 4 técnicas avaliadas (Tabela 7.3). Além disso, o nível de ruído simulado não influenciou no desempenho da técnica JR. Já a técnica SG foi negativamente influenciada com o aumento do nível de ruído.

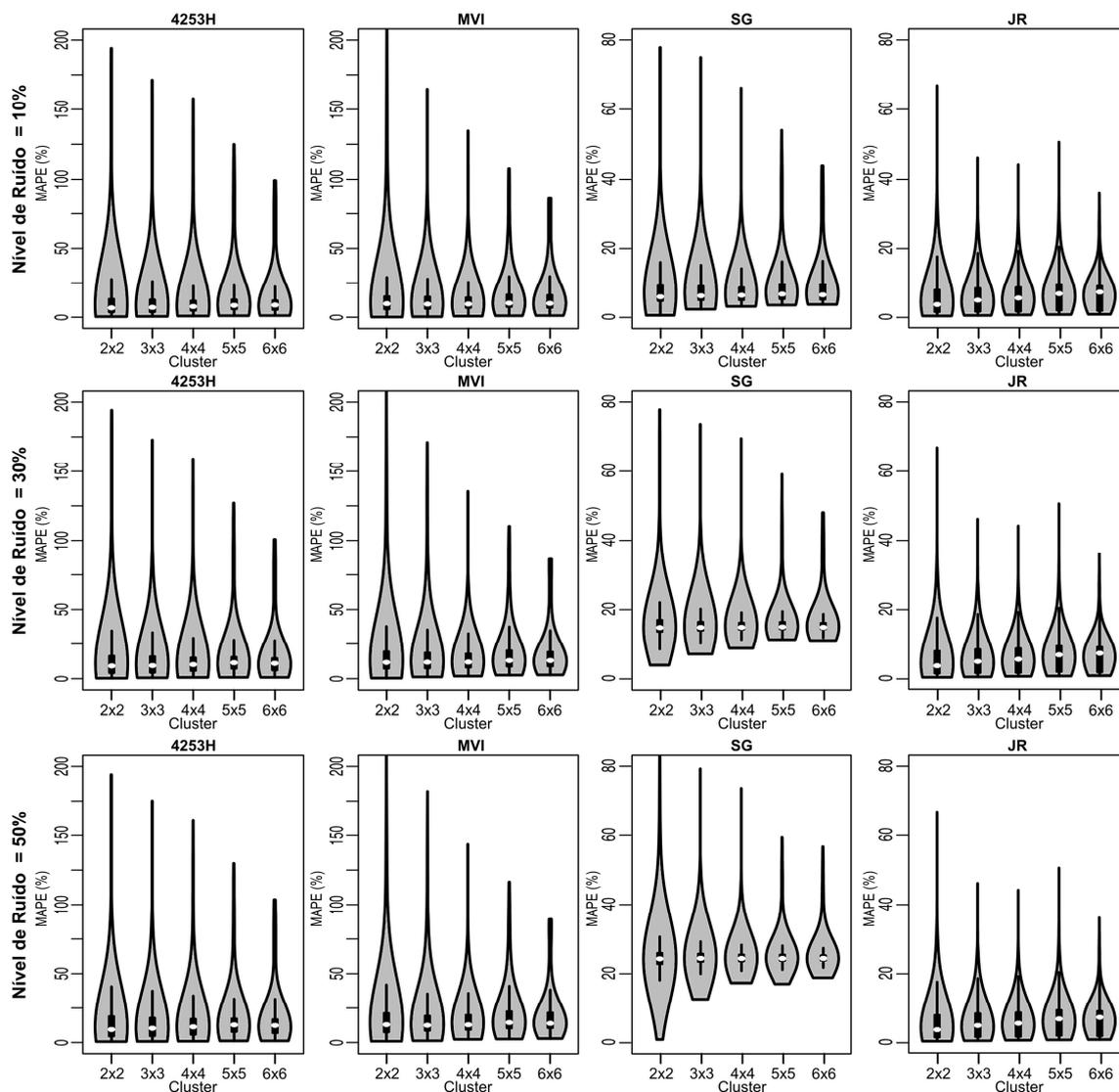


Figura 7.9 – Boxplot (no formato de violino) do erro percentual médio absoluto para as 1000 iterações realizadas na amostragem “III”.

A influência do aumento da dimensão do cluster no desempenho de cada técnica foi avaliada com base em um teste de Mann-Whitney. O teste demonstrou que o desempenho das técnicas entre os diferentes clusters foram significativamente diferentes a nível de significância de 5% (Tabela D.1 - Apêndice D). Os filtros 4253H twice e MVI apresentaram comportamento semelhante em relação à acurácia obtida para cada nível de ruído e dimensão do cluster simulado. Para o filtro SG, o teste de Mann-Whitney demonstrou que somente para o nível de ruído igual a 50% não foram observadas diferenças significativas entre os valores de MAPE dos clusters simulados.

Tabela 7.3 – Desempenho das técnicas de redução de ruídos com base no Teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações - Amostragem “III”.

Nível Ruído	Desempenho (decrecente)	Dimensão do cluster simulado				
		2x2	3x3	4x4	5x5	6x6
10%	1	JR	JR	JR	JR	JR
	2	SG	SG	SG	SG	SG
	3	4253H	4253H	4253H	4253H	4253H
	4	MVI	MVI	MVI	MVI	MVI
30%	1	JR	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG
50%	1	JR	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG

Obs.: Todas os desempenhos apresentaram diferenças significativas pelo teste de Wilcoxon pareado com nível de significância de 5%.

Particularmente para a técnica JR, que utiliza da informação espacial em sua análise, o aumento na dimensão do cluster resultou em queda de desempenho em todas as dimensões de clusters e níveis de ruídos simulados. Pela análise do desempenho obtido entre as diferentes dimensões de clusters (de 2x2 a 6x6 pixels) é possível perceber uma pequena queda no desempenho da técnica JR, pois os valores médios de MAPE após 1000 iterações passaram de 5,5% para 6,7% (Figura 7.9). Mesmo assim, este valor de 6,7% foi inferior aos valores de MAPE obtidos para as demais técnicas independentemente da intensidade de ruído ou dimensão de cluster simulado.

7.2.3. Avaliação de desempenho para simulação temporal (Amostragem IV)

Sano et al. (2007) analisaram imagens Landsat do cerrado brasileiro cobrindo mais de 30 anos, e concluíram que a probabilidade para obter uma cena com cobertura de nuvem inferior a 10% entre os meses de dezembro a fevereiro não ultrapassa a 18%. Também foi identificado por Asner (2001) uma baixa

probabilidade de obter imagens sem cobertura de nuvem na Amazônia brasileira, onde foram mapeadas áreas com presença de nuvem por mais de 30 dias (ASNER; ALENCAR, 2010).

Sendo assim, a amostragem “IV” foi desenvolvida para simular baixa qualidade em datas consecutivas (ou *gap*) ao longo do perfil temporal de um pixel. Foram simulados intervalos de 2 a 6 datas consecutivas e a introdução de ruídos em cada análise. Os níveis de ruídos nas intensidades de 10%, 30% e 50% foram aleatoriamente adicionados ou subtraídos ao valor original de NDVI do ponto simulado.

A Figura 7.10 apresenta o desempenho das técnicas de filtragem após 1000 iterações realizadas com a amostragem “IV” na área de teste “A”. A análise do comportamento dos valores de MAPE para as 1000 iterações demonstra que a técnica JR apresentou o melhor desempenho entre as demais técnicas (4253H twice, MVI e SG), independentemente do nível de ruído ou tamanho de *gap* simulado. Os valores médios de MAPE fornecidos pela técnica JR foram inferiores a 7% em todas as simulações, e novamente a intensidade de ruído não influenciou na acurácia em virtude de o ponto simulado (ou com ruído) não ser utilizado na estimativa de um novo dado.

As técnicas 4253H twice, MVI e SG apresentaram queda na amplitude dos valores de MAPE com o aumento do tamanho de *gap*. Esta redução na amplitude pode ser atribuída ao maior número de pontos analisados com o aumento do *gap* simulado, pois para cada interação cujo *gap* simulado é equivalente a 6, a estatística MAPE é obtida com a análise de 6 pontos. Em contrapartida, um *gap* igual a 2 representa uma análise de 2 datas para compor a mesma análise estatística.

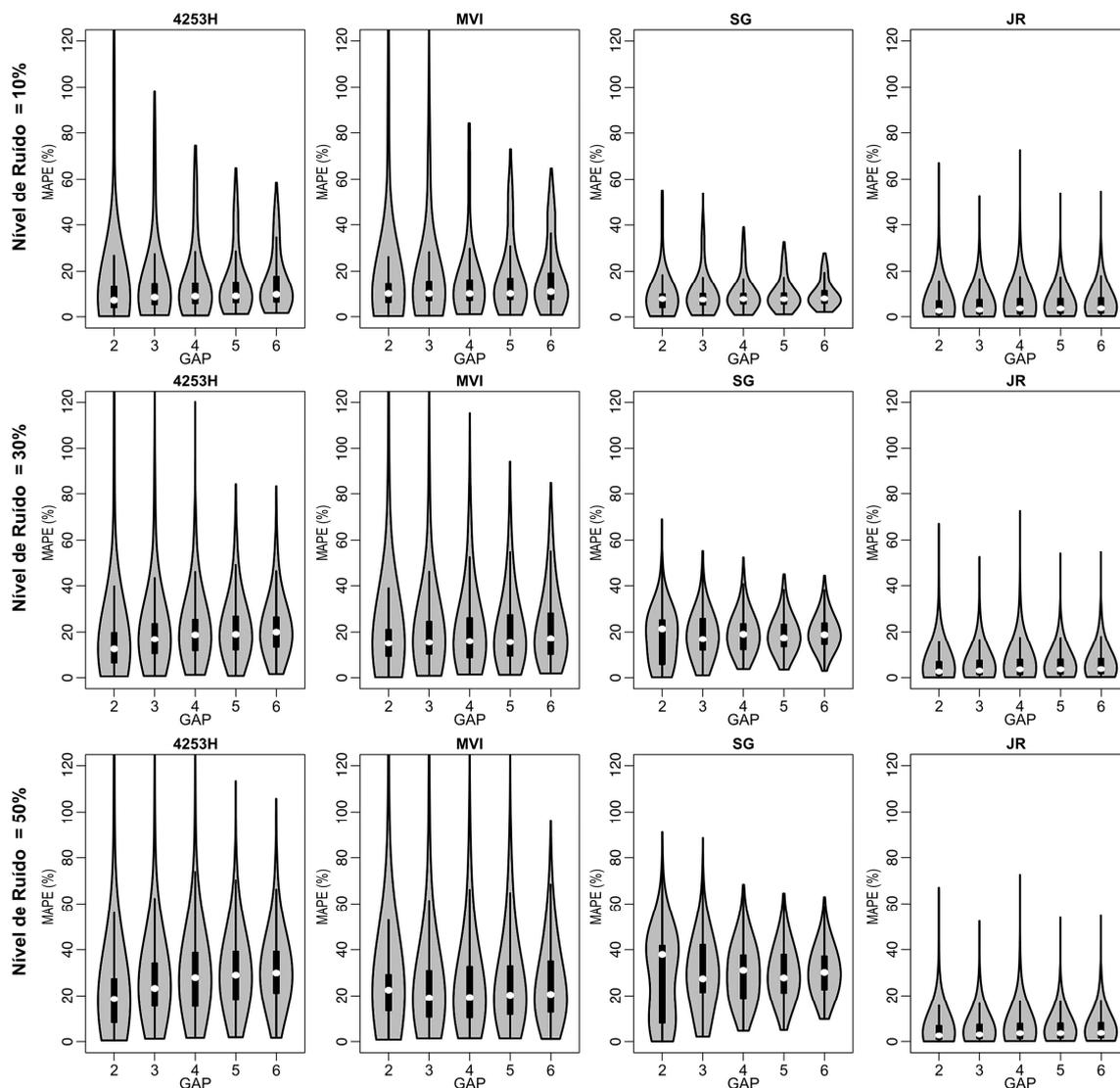


Figura 7.10 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações realizadas na amostragem “IV”.

As técnicas (4253H twice, MVI e SG) que usam somente a informação temporal para estimativas de novos valores apresentaram variabilidade na acurácia entre os diferentes níveis de ruídos simulados, de forma que um aumento no nível de ruído refletiu em diferentes desempenhos da acurácia obtida.

Novamente, o teste não paramétrico de Wilcoxon pareado foi utilizado para comparar o desempenho das técnicas de redução de ruídos analisadas em relação à simulação de *gaps*. Para isto, foram avaliadas as medidas de posição

entre as 1000 iterações, de forma que os valores de MAPE de uma técnica foram comparados e ordenados em relação aos valores obtidos nos demais filtros. A Tabela 7.4 apresenta uma ordenação do desempenho das técnicas de filtragem com base no teste de Wilcoxon pareado realizado para cada *gap* simulado. O filtro JR foi considerado estatisticamente o mais eficiente nas três intensidades de ruídos simulados no estudo, independentemente do tamanho de *gap* ou nível de ruído simulado. Para o menor nível de ruído simulado o filtro SG apresentou um desempenho próximo à técnica JR, porém à medida que a intensidade de ruído aumentou os filtros 4253H twice, MVI e SG não mantiveram o mesmo desempenho.

Tabela 7.4 – Desempenho das técnicas de redução de ruídos com base no teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações - Amostragem “IV.

Nível Ruído	Desempenho (decréscante)	Tamanho de <i>gap</i> simulado				
		2	3	4	5	6
10%	1	JR	JR	JR	JR	JR
	2	SG	SG	SG	SG	SG
	3	4253H	4253H	4253H	4253H	4253H
	4	MVI	MVI	MVI	MVI	MVI
30%	1	JR	JR	JR	JR	JR
	2	4253H	MVI ^b	MVI ^c	MVI ^d	SG
	3	MVI ^a	SG ^b	SG ^c	SG ^d	MVI
	4	SG ^a	4253H	4253H	4253H	4253H
50%	1	JR	JR	JR	JR	JR
	2	4253H	MVI	MVI	MVI	MVI
	3	MVI	4253H	4253H	SG ^e	4253H ^f
	4	SG	SG	SG	4253H ^e	SG ^f

Obs.: Filtros com letras iguais não expressam diferenças significativas pelo teste de Wilcoxon pareado com nível de significância de 5%.

Foi utilizado o teste não paramétrico de Mann-Whitney para comparar as tendências centrais dos valores de MAPE obtidos para os *gaps* simulados. Este teste foi realizado para cada tipo de filtro e intensidade de ruído aplicada, e os resultados (Tabela D.2 – Apêndice D) retratam que os valores de MAPE obtidos após 1000 iterações apresentam diferenças significativas (a nível de significância de 5%) com o aumento do *gap*.

As técnicas que utilizam somente a informação temporal para a redução de ruídos apresentaram desempenhos constantes em relação ao tamanho do *gap*. A exceção foi a técnica 4253H twice, que forneceu piores resultados para os maiores níveis de ruídos e tamanho de *gap* superior a 2.

7.3. Avaliação de desempenho para áreas testes “B” e “C”

As simulações de ruídos realizadas na área “A” demonstraram que a técnica JR possui um desempenho satisfatório na estimativa de novos valores de NDVI para áreas homogêneas. Entretanto, a variabilidade de uso e cobertura do solo mapeada por Sanches (2004) demonstra que o município de Ipuã possui áreas com maior fragmentação e heterogeneidade entre alvos.

As simulações e medidas de dispersão utilizadas para avaliação de desempenho das técnicas de filtragem na área de teste “A” foram replicadas sobre as áreas “B” e “C”. A área “B” está localizada em uma região agrícola, porém com maior diversidade das classes de cobertura do solo e dimensões dos talhões menores quando comparado aos alvos agrícolas presentes na área “A”. A área “C” representa duas classes de uso e cobertura de solo (agricultura e floresta) com diferentes dimensões dos fragmentos.

As técnicas de redução de ruído avaliadas no presente estudo apresentaram acurácias semelhantes em ambas as áreas (Figura 7.11). Os resultados para a amostragem do tipo “II”, que degrada o dado original tanto no tempo como no espaço, revelam que os filtros 4253H twice, MVI e SG foram influenciados negativamente com o aumento da intensidade de ruído simulada. A técnica JR, por não utilizar o dado com ruído na estimativa de novos valores, apresentou uma acurácia constante independentemente da intensidade de ruído. Os valores de MAPE obtidos após 1000 iterações da amostragem “II” sobre as Áreas “B” e “C” estão ilustrados nas Figuras A.1 e A.2 (Apêndice A), respectivamente.

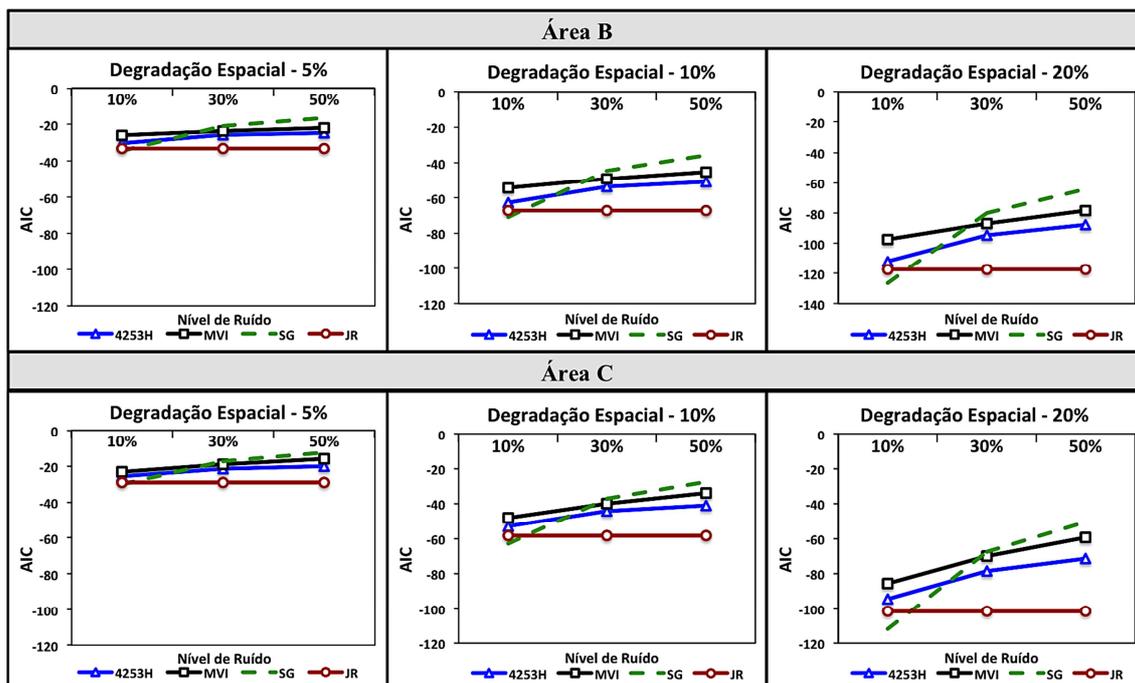


Figura 7.11 – Acurácia das técnicas de redução de ruídos (4253H twice – MVI – SG – JR) para a amostragem “II” com base no valor médio do AIC obtido entre 1000 iterações sobre as áreas teste “B” e “C”.

Os desempenhos das técnicas foram comparadas por meio do teste de Wilcoxon, que revelou acurácias equivalentes a um nível de significância de 5% entre as técnicas SG e JR (Tabela 7.5). Entretanto, esta equivalência na acurácia foi identificada somente para o menor nível de ruído simulado (ou nível de 10%) sobre a área “B”. Para o mesmo nível de ruído sobre a área “C”, a técnica SG apresentou desempenho superior à técnica JR para a maior degradação espacial simulada; porém, esta superioridade não foi identificada com degradação espacial de 10%, e foi até inferior para uma degradação espacial de 5%.

Para os níveis de ruídos equivalentes a 30% e 50% os desempenhos entre as técnicas não foram alterados em ambas as áreas teste; porém a técnica SG apresentou uma queda de desempenho entre as técnicas avaliadas com o aumento do ruído.

Tabela 7.5 – Desempenho das técnicas de redução de ruídos com base no Teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações sobre as áreas “B” e “C”.

Nível de Ruído	Desempenho (decrecente)	Área B			Área C		
		Degradação Espacial			Degradação Espacial		
		5%	10%	20%	5%	10%	20%
10%	1	SG ^a	SG ^b	SG ^c	JR	SG ^d	SG
	2	JR ^a	JR ^b	JR ^c	SG	JR ^d	JR
	3	4253H	4253H	4253H	4253H	4253H	4253H
	4	MVI	MVI	MVI	MVI	MVI	MVI
30%	1	JR	JR	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG	SG
50%	1	JR	JR	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG	SG

Obs.: Filtros com letras iguais não expressam diferenças significativas pelo teste de Wilcoxon pareado com nível de significância de 5%.

A Tabela 7.6 apresenta o desempenho das técnicas avaliadas nesta pesquisa em relação à simulação de grupos de pixels com ruídos (amostragem III) para as áreas “B” e “C”. Novamente, as técnicas apresentaram desempenhos semelhantes em ambas as áreas teste. Entretanto, merece destaque o desempenho superior da técnica SG para a maioria das dimensões de clusters simulados com ruído de 10%. Porém, esta técnica sofreu a maior queda de desempenho com o aumento do nível de ruído, pois apresentou o pior desempenho com níveis de ruídos equivalentes a 30% e 50%.

Os valores de MAPE obtidos após 1000 iterações da amostragem “III” sobre as Áreas “B” e “C” foram agrupados e estão ilustrados nas Figuras B.1 e B.2 (Apêndice B), respectivamente.

A influência do nível de ruído ao longo de uma sequência de datas (amostragem IV) também foi avaliada nas áreas teste “B” e “C”, e o

desempenho obtido por cada técnica em relação ao nível de ruído e tamanho de *gap* simulado está agrupado na Tabela 7.7. Novamente, as técnicas JR e SG apresentaram os melhores desempenhos na estimativa de novos valores após a simulação do nível de ruído equivalente a 10%. Porém, para a área “B”, a técnica JR apresentou desempenho superior para a maioria dos *gaps* simulados. E na área “C” a técnica SG foi a que forneceu as melhores estimativas após as 1000 iterações.

Para os níveis de ruídos equivalentes a 30% e 50% os desempenhos das técnicas apresentaram o mesmo comportamento entre os *gaps* simulados, com destaque para os filtros MVI e SG, que apresentaram desempenhos opostos com o aumento do nível de ruído. O filtro MVI apresentou uma melhora no ordenamento dos desempenhos em níveis de ruídos maiores, ao contrário do filtro SG, que demonstrou ter desempenho negativamente influenciado. Os valores de MAPE referentes à amostragem “IV” para as Áreas “B” e “C” estão ilustrados nas Figuras C.1 e C.2 (Apêndice C), respectivamente.

7.4. Síntese dos resultados obtidos (áreas teste “A”, “B” e “C”)

A comparação dos resultados obtidos pelas técnicas de filtragem mostrou que a intensidade de ruído simulada interfere diretamente nos resultados. Além disso, a forma com que os ruídos estão presentes no perfil temporal pode alterar a eficiência das técnicas de filtrações.

Entre as técnicas analisadas no presente estudo a metodologia proposta (JR) apresentou habilidade superior para minimizar os ruídos simulados sobre as séries temporais de NDVI. Esse desempenho superior está relacionado a I) os valores dos pixels/ou datas simuladas com ruídos não são utilizados no processo de filtragem; II) por ser um método de filtragem local que utiliza uma janela espaço-temporal adaptativa à informação de qualidade do pixel; e III) por não ter a necessidade de definir limiares ou parâmetros para sua aplicação.

Tabela 7.6 – Desempenho das técnicas de redução de ruídos com base no teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações sobre as áreas “B” e “C”.

Nível de ruído	Desempenho (decrecente)	Área B					Área C				
		Dimensão de cluster simulado					Dimensão de cluster simulado				
		2x2	3x3	4x4	5x5	6x6	2x2	3x3	4x4	5x5	6x6
10%	1	JR	SG	SG	SG	SG	JR ^b	SG	SG	SG	SG
	2	SG	JR	JR	JR	JR ^a	SG ^b	JR	JR	JR	JR
	3	4253H	4253H	4253H	4253H	4253H ^a	4253H	4253H	4253H	4253H	4253H
	4	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI
30%	1	JR	JR	JR	JR	JR	JR	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG	SG	SG	SG	SG	SG
50%	1	JR	JR	JR	JR	JR	JR	JR	JR	JR	JR
	2	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H
	3	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG	SG	SG	SG	SG	SG

Obs.: Filtros com letras iguais não expressam diferenças significativas pelo teste de Wilcoxon pareado com nível de significância de 5%.

Tabela 7.7 – Desempenho das técnicas de redução de ruídos com base no teste de Wilcoxon pareado nos valores de MAPE das 1000 iterações sobre as áreas “B” e “C”.

Nível de ruído	Desempenho (decrecente)	Área B					Área C				
		Tamanho do <i>gap</i> simulado					Tamanho do <i>gap</i> simulado				
		2	3	4	5	6	2	3	4	5	6
10%	1	JR	JR	JR	JR	JR ^a	JR	SG	SG	SG	SG
	2	SG	SG	SG	SG	SG ^a	SG	JR	JR	JR	JR ^b
	3	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H	4253H ^b
	4	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI	MVI
30%	1	JR	JR	JR	JR	JR	JR	JR	JR	JR	JR
	2	4253H	MVI	MVI	MVI	MVI	4253H	MVI	MVI	MVI	MVI
	3	MVI	4253H	4253H	4253H	4253H	MVI	4253H	4253H	SG ^c	4253H ^d
	4	SG	SG	SG	SG	SG	SG	SG	SG	4253H ^c	SG ^d
50%	1	JR	JR	JR	JR	JR	JR	JR	JR	JR	JR
	2	4253H	MVI	MVI	MVI	MVI	4253H	4253H	4253H	4253H	4253H
	3	MVI	4253H	4253H	4253H	4253H	MVI	MVI	MVI	MVI	MVI
	4	SG	SG	SG	SG	SG	SG	SG	SG	SG	SG

Obs.: Filtros com letras iguais não expressam diferenças significativas pelo teste de Wilcoxon pareado com nível de significância de 5%.

Jönsson e Eklundh (2004) destacam que o filtro Savitzky–Golay é um filtro local e que apresenta bons resultados na filtragem de séries temporais que são consideradas quasi-periódicas. Além disso, os autores evidenciam o bom desempenho desse filtro quando os valores de NDVI não são afetados por ruído. Isso corrobora os resultados obtidos por esta técnica para as simulações com a menor intensidade de ruído. Porém, o filtro Savitzky–Golay demonstrou ser uma técnica desfavorável para redução de altos níveis de ruídos. Uma justificativa para esta queda de desempenho pode estar relacionada à configuração do tamanho da janela e grau de polinômio definidos para aplicação do Savitzky–Golay, pois Chen et al. (2004) mostram que o melhor ajuste depende de uma combinação desses parâmetros. No presente estudo, o filtro Savitzky–Golay foi configurado com uma janela de tamanho 5 e grau de polinômio igual a 3, e acredita-se que uma análise sobre a série temporal com a finalidade de definir os valores dos parâmetros deste filtro pode proporcionar uma melhoria no desempenho para a redução dos altos níveis de ruídos. Para tamanhos de janelas maiores, o filtro pode preservar as tendências temporais presentes na série temporal de NDVI, limitando assim a sua capacidade de identificar rápidas, mas importantes, mudanças no valor do índice de vegetação (como, por exemplo, uma resposta relacionada ao aumento do vigor da vegetação). Por outro lado, as janelas estreitas podem gerar um ajuste acentuado da série temporal, capturando as variações ao longo do conjunto de dados e retendo mais ruídos (CHEN et al., 2004; JÖNSSON; EKLUNDH, 2004).

Para os maiores níveis de ruídos simulados, o filtro 4253H twice apresentou bom desempenho em todos os tipos de cobertura do solo avaliados. Dijk et al. (1987) e Velleman (1980) destacaram a habilidade do filtro 4253H twice em manter a integridade do sinal original enquanto elimina as oscilações espúrias. Portanto, o bom desempenho obtido no presente estudo está relacionado aos procedimentos de obtenção consecutiva dos valores de medianas sobre os dados, de forma que tais procedimentos foram capazes de remover as descontinuidades abruptas simuladas pelos altos níveis de ruídos sobre as séries temporais.

O filtro MVI apresentou um comportamento estável independentemente do nível de ruído simulado, tipos de amostragens ou classe de cobertura do solo considerados nesta pesquisa. Entretanto, Ma e Veroustraete (2006) mencionam que a escolha do valor limiar é essencial para o desempenho deste filtro. Para elevados valores de limiar, o filtro tende a manter as oscilações abruptas presentes ao longo da série temporal, enquanto que para um baixo valor de limiar, o filtro MVI tende a ajustar os valores próximos aos de um filtro de média móvel.

Efeitos de borda ou restrições de análise para as datas que estão nas extremidades de um perfil temporal estão presentes em todos os filtros analisados nesta pesquisa. O efeito de borda para os filtros 4253H twice, SG e JR está associado ao limite de no mínimo duas datas em cada extremidade da série temporal para completar a janela de análise de cada filtro. O filtro MVI tem limitação para analisar a primeira e última data. Entretanto, Chen et al. (2004) ressaltam que este problema pode ser resolvido quando se assume que as séries temporais de NDVI são cíclicas, ou seja, o valor vizinho à última data da série temporal é representando pelo valor da primeira data.

As técnicas 4253H twice, Valor Médio da Iteração e Savitzky–Golay não dependem da informação espacial para realizar os procedimentos para redução de ruídos em séries temporais. Porém, os diferentes tamanhos de clusters simulados em cada degradação espacial resultaram em pequenas alterações nos valores médios de MAPE. Além disso, essas técnicas apresentaram queda no desempenho conforme o aumento do nível de ruído simulado, com destaque para o filtro SG, que apresentou a maior queda de desempenho.

A técnica JR apresentou bom desempenho na redução de ruídos simulados espacialmente sobre uma cena, porém uma maior sensibilidade desta técnica foi identificada sobre áreas heterogêneas. A maior variabilidade espacial dos valores de NDVI em áreas com cobertura do solo heterogênea proporcionou uma ligeira queda do desempenho desta técnica para a redução do menor nível de ruído simulado. Esta sensibilidade fica mais evidente com o aumento

da dimensão espacial dos dados a serem analisados sobre áreas cuja cobertura do solo seja mais fragmentada.

O máximo intervalo de datas classificadas consecutivamente com baixa qualidade, ou máximo *gap*, foi o parâmetro utilizado para avaliar o desempenho dos filtros em relação à degradação temporal das séries temporais de NDVI. Entretanto, Zhou et al. (2012) destacam que a posição que o *gap* ocorre ao longo do perfil temporal tem influência significativa no desempenho dos filtros. Em geral, os *gaps* que ocorrem próximos aos valores de máximo, ou próximos às extremidades de uma série temporal tendem a introduzir maiores incertezas no resultado final.

As mesmas condições de *gap* aplicadas sobre séries temporais com padrões diferentes podem resultar em alterações nos desempenhos dos filtros, como ilustrado pela amostragem “IV”. De acordo com Zhou et al. (2012), a variabilidade dos valores de NDVI ao longo da série temporal pode ser considerada como indicador do desempenho do filtro, pois a reconstrução de uma série de dados que apresente maior variação sazonal é influenciada negativamente pelo *gap*.

Sendo assim, a queda de desempenho da técnica JR em relação aos tamanhos de *gaps* simulados para a área teste “C” pode estar relacionada à maior variabilidade sazonal e espacial dos valores de NDVI para os alvos analisados.

8 CONCLUSÃO

Os ruídos presentes em séries temporais de NDVI causados pela contaminação por presença de nuvens e variabilidade atmosférica são considerados um problema nas análises de mudanças ambientais em que as séries temporais são utilizadas como dados de entrada nos modelos.

Atzberger e Eilers (2011a) enfatizam a necessidade de reduzir os ruídos presentes em séries temporais de NDVI e mencionam que este procedimento envolve as etapas de preenchimento de *gaps* e subsequente suavização dos dados. O método de Janela de Regressão, proposto no presente estudo, foi desenvolvido para classificar como *gap* todo dado avaliado como de baixa qualidade (ou ruído) ao longo da série temporal. Posteriormente, o método proposto utiliza de procedimentos que envolvam análises dos dados em relação ao espaço e o tempo para a recomposição da série temporal.

Poucas técnicas de filtragem utilizam a informação da qualidade do pixel no procedimento de redução de ruídos, ou mesmo ponderam a influência de um dado como baixa qualidade no resultado final do processo. Roy et al. (2002) recomendam aos usuários de produtos MODIS inspecionarem os metadados relacionados à qualidade das imagens e aconselham que os produtos com baixa garantia de qualidade devem ser utilizados com certa prudência.

Uma análise da distribuição espaço-temporal dos ruídos presentes nos dados permite uma melhor compreensão da qualidade da base de dados e, conseqüentemente, uma definição de parâmetros e/ou filtros mais adequados para a reconstrução da série com qualidade. O desempenho de um filtro está diretamente relacionado aos valores de limiares e tamanhos das janelas de análise definidos pelo usuário.

A intensidade do ruído, aliada à decisão de extração de variáveis fenológicas da base de dados, são considerados fatores influentes na escolha de uma abordagem para a redução de ruídos em séries temporais de NDVI (HIRD; MCDERMID, 2009). Musial et al. (2011) evidenciam que a escolha de uma

abordagem para preencher a falta de dados ou estimar novos valores em séries temporais depende do entendimento do sinal, da distribuição das datas com *gaps* e da expectativa do analista em obter dados próximos aos originais ou uma representação das características gerais da série histórica.

Neste trabalho, foi proposta uma nova abordagem para estimar valores de pixels classificados como de baixa qualidade, e o desempenho desta metodologia foi comparada com outras três técnicas de filtragem de ruídos em séries temporais de NDVI.

Tanto os métodos de amostragem quanto as medidas de dispersão empregadas nesta pesquisa permitiram avaliar o desempenho dos filtros na redução de ruídos presentes em uma base de dados temporal.

Os ruídos simulados sobre as séries temporais foram subdivididos em três intensidades, de forma que as técnicas de filtragem Janela de Regressão e Savitzky–Golay forneceram os melhores ajustes na redução dos ruídos de baixa intensidade (ruídos de 10%). Porém, para os ruídos simulados com os níveis de 30% e 50%, a técnica de filtragem JR apresentou desempenho superior ao obtido pelas demais técnicas analisadas.

Os resultados demonstraram que a intensidade de um ruído no valor do índice de vegetação influencia nos resultados obtidos pelas técnicas de filtragem 4253H twice, MVI e SG. Porém, isso não ocorre para a técnica de filtragem JR, pois o valor de NDVI do pixel a ser analisado, bem como os valores dos pixels considerados como de baixa qualidade não são utilizados no procedimento para a estimativa de um novo valor do pixel de interesse. A técnica JR apresentou desempenho satisfatório com até seis datas degradadas consecutivamente (*gap*) e degradação espacial para um grupo de 6x6 pixels. No entanto, os critérios de seleção da janela espaço-temporal fazem com que esta técnica seja mais afetada negativamente com *gaps* superiores a 7 ou por um maior número de pixels classificados como de baixa qualidade na mesma data.

O método proposto utiliza a análise espaço-temporal para estimar um novo valor de NDVI. Mas uma análise entre pixels vizinhos deverá ser aplicada com cautela sobre áreas heterogêneas, ou em áreas cuja a cobertura do solo seja mais fragmentada. Diferentes classes de uso e cobertura do solo podem apresentar maior variabilidade espacial nos valores de NDVI, e essa variabilidade pode fornecer baixa correlação entre pixels vizinhos e aumento na incerteza dos valores estimados. Entretanto, uma análise conjunta dos valores de NDVI no tempo-espaço, juntamente com a informação de qualidade dos pixels, provou ser uma alternativa viável e promissora para estimar os pixels classificados como de baixa qualidade em séries temporais NDVI.

A presente metodologia foi desenvolvida com uma finalidade generalista, ou seja, poderá ser aplicada em outras séries temporais além do produto MOD13Q1 analisado no presente estudo. Dessa forma o método Janela de Regressão poderá ser empregado e avaliado, em futuros trabalhos, sobre séries temporais com características distintas, como por exemplo, as séries temporais provenientes de outros produtos MODIS (MOD09 – Reflectância de Superfície e MOD11 – Temperatura de Superfície) e derivadas de sensores como o NOAA/AVHRR, TM e ETM+ do Landsat. Além disso, a presente metodologia poderá ser aplicada em conjunto de dados com alta resolução espacial ou temporal, bem como em séries temporais cujas as imagens foram irregularmente adquiridas ao longo do tempo.

Os resultados obtidos nesta pesquisa indicam bom potencial da metodologia para reduzir o ruído em séries temporais NDVI/MODIS. No entanto, mais investigações são necessárias para redução de ruído por meio de análise espaço-temporal. A técnica JR considera que pixels vizinhos oriundos da mesma cobertura do solo são mais correlacionados do que pixels com diferentes coberturas. Portanto, futuras pesquisas devem incluir informação sobre as classes de cobertura do solo no processo de seleção dos dados pela janela espacial, de modo que os pixels com diferentes usos sejam excluídos, ou ponderados, na estimativa de um novo valor para o pixel de interesse. Além disso, o desempenho da técnica JR deve ser testado em regiões com

diferentes coberturas e comparado com outras técnicas de redução de ruído disponíveis na literatura. Novos estudos devem ser direcionados para avaliar o desempenho do método proposto na estimativa de novos valores para os pixels localizados tanto nas bordas da imagem, bem como para os pixels localizados em regiões limítrofes entre classes de cobertura do solo.

Por último, destaca-se que são necessários esforços contínuos para desenvolver e melhorar a modelagem conjunta no tempo e espaço que permita representar a dinâmica de uso e cobertura do solo.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMI, M. **Estimativa da data de plantio da soja por meio de séries temporais de imagens MODIS**. 2010. 161 p. (sid.inpe.br/mtc-m19/2010/09.15.21.47-TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010. Disponível em: <<http://urlib.net/8JMKD3MGP7W/3897URE>>. Acesso em: 21 fev 2014.

AKAIKE, H. Information Theory and an Extension of the Maximum Likelihood Principle. In: SECOND INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 1973, Budapest, Hungary. **Proceedings...** Budapest: Akademia Kiado, p. 267–281.

ASNER, G. P. Cloud cover in Landsat observations of the Brazilian Amazon. **International Journal of Remote Sensing**, v. 22, n. 18, p. 3855–3862, jan. 2001.

ASNER, G. P; ALENCAR, A. Drought impacts on the Amazon forest: the remote sensing perspective. **The New phytologist**, v. 187, n. 3, p. 569–78, ago. 2010.

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.; EILERS, P. H. C. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. **International Journal of Remote Sensing**, v. 32, n. 13, p. 3689–3709, Jul. 2011a.

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, Set. 2011b.

BECK, P. S. A.; ATZBERGER, C.; HØGDA, K. A.; JOHANSEN, B.; SKIDMORE, A. K. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. **Remote Sensing of Environment**, v. 100, n. 3, p. 321–334, fev. 2006.

BRADLEY, B. A.; JACOB, R. W.; HERMANCE, J. F.; MUSTARD, J. F. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. **Remote Sensing of Environment**, v. 106, n. 2, p. 137–145, Jan. 2007.

BREUNIG, F. M.; GALVÃO, L. S.; FORMAGGIO, A. R.; EPIPHANIO, J. C. N. Directional effects on NDVI and LAI retrievals from MODIS: A case study in Brazil with soybean. **International Journal of Applied Earth Observation and Geoinformation**, v. 13, n. 1, p. 34-42, Feb. 2011.

BROOKS, E. B.; THOMAS, V. A.; WYNNE, R. H.; COULSTON, J. W. Fitting the Multitemporal Curve: A Fourier Series Approach to the Missing Data Problem in Remote Sensing Analysis. **IEEE Transactions on Geoscience and Remote Sensing**, v. 50, n. 9, p. 3340–3353, Set. 2012. CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T.; MELLO, C. Utilização de filtro de transformada de fourier para a minimização de ruídos em sinais analíticos. **Química Nova**, v. 23, n. 5, p. 690–698, Out. 2000.

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, Jun. 2004.

CHEN, C.; SON, N.; CHEN, C.; CHANG, L. Wavelet filtering of time-series moderate resolution imaging spectroradiometer data for rice crop mapping using support vector machines and maximum likelihood classifier. **Journal of Applied Remote Sensing**, v. 5, n. 1, p. 053525–1 – 053525–15, Maio 2011.

CHO, A.; SUH, M. Detection of contaminated pixels based on the short-term continuity of NDVI and correction using spatio-temporal continuity. **Asia-Pacific Journal of Atmospheric Sciences**, v. 49, n. 4, p. 511–525, Ago. 2013.

COLDITZ, R. R.; CONRAD, C.; WEHRMANN, T.; SCHMIDT, M.; DECH, S. TiSeG: A Flexible Software Tool for Time-Series Generation of MODIS Data Utilizing the Quality Assessment Science Data Set. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, n. 10, p. 3296–3308, Out. 2008.

CRESSIE, N. A. C. **Statistics for spatial data**. New York, NY: Wiley - interscience, 1993. 900 p. ISBN (978-04-7100-255-0).

DIDAN, K.; HUETE, A. **MODIS vegetation index product series collection 5 change summary**. Tucson: The University of Arizona, 2006. 17 p. Disponível em: <http://landweb.nascom.nasa.gov/QA_WWW/forPage/MOD13_VI_C5_Changes_Document_06_28_06.pdf>. Acesso em 20 nov. de 2013.

DIJK, A. V.; CALLIS, S. L.; SAKAMOTO, C. M.; DECKER, W. L. Smoothing vegetation index profiles: An alternative method for reducing radiometric disturbance in NOAA/AVHRR data. **Photogrammetric Engineering and Remote Sensing**, v. 53, n. 8, p. 1059–1067, Aug. 1987.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. **Análise espacial de dados geográficos**. Brasília, EMBRAPA, 2004. 208 p. ISBN (85-7383-260-6).

EPIPHANIO, J. C. N.; HUETE, A. R. Dependence of NDVI and SAVI on Sun/Sensor Geometry and its effect on fAPAR relationships in alfalfa. **Remote Sensing of Environment**, v. 51, n. 3, p. 351-360, Mar. 1995. (INPE-5926-PRE/2069).

GU, J.; LI, X.; HUANG, C.; OKIN, G. S. A simplified data assimilation method for reconstructing time-series MODIS NDVI data. **Advances in Space Research**, v. 44, n. 4, p. 501–509, Ago. 2009.

HINTZE, J. L.; NELSON, R. D. Violin Plots: A Box Plot-Density Trace Synergism. **The American Statistician**, v. 52, n. 2, p. 181–184, Maio 1998.

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, Jan. 2009.

HOLBEN, B. N. Characteristics of maximum-value composite images from temporal AVHRR data. **International Journal of Remote Sensing**, v. 7, n. 11, p. 1417–1434, 1986.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 195–213, 2002.

HUETE, A.; DIDAN, K.; VAN LEEUWEN, W.; MIURA, T.; GLENN, E. MODIS Vegetation Indices. In: RAMACHANDRAN, B.; JUSTICE, C. O.; ABRAMS, M. J. (Eds.). **Land remote sensing and global environmental change: NASA's Earth observing system and the science of ASTER and MODIS**. [S.l.]: Springer, 2011. 873 p. ISBN (978-1-4419-6749-7).

JÖNSSON, P.; EKLUNDH, L. TIMESAT - a program for analyzing time-series of satellite sensor data. **Computers & Geosciences**, v. 30, n. 8, p. 833–845, Out. 2004.

JUSTICE, C. O.; TOWNSHEND, J. R. G.; VERMOTE, E. F.; MASUOKA, E.; WOLFE, R. E.; SALEOUS, N.; ROY, D. P.; MORISETTE, J. T. An overview of MODIS Land data processing and product status. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 3–15, Nov. 2002.

LASAPONARA, R. Estimating Interannual Variations in Vegetated Areas of Sardinia Island Using SPOT/VEGETATION NDVI Temporal Series. **IEEE Geoscience and Remote Sensing Letters**, v. 3, n. 4, p. 481–483, Out. 2006.

LOVELL, J. L.; GRAETZ, R. D. Filtering Pathfinder AVHRR Land NDVI data for Australia. **International Journal of Remote Sensing**, v. 22, n. 13, p. 2649–2654, Jan. 2001.

LU, X.; LIU, R.; LIU, J.; LIANG, S. Removal of Noise by Wavelet Method to Generate High Quality Temporal Data of Terrestrial MODIS Products. **Photogrammetric Engineering & Remote Sensing**, v. 73, n. 10, p. 1129–1139, Out. 2007.

MA, M.; VEROUSTRAETE, F. Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. **Advances in Space Research**, v. 37, n. 4, p. 835–840, 2006.

MADDEN, H. H. Comments on the Savitzky–Golay convolution method for least-squares-fit smoothing and differentiation of digital data. **Analytical Chemistry**, v. 50, n. 9, p. 1383–1386, Ago. 1978.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo, SP : Editora Edgard Blucher, 2 ed., 2006. 564 p., ISBN (85-212-0389-6).

MUSIAL, J. P.; VERSTRAETE, M. M.; GOBRON, N. Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. **Atmospheric Chemistry and Physics Discussions**, v. 11, n. 5, p. 14259–14308, Maio 2011.

NETER, J.; KUTNER, M.H.; NACHTSHEIM, C.J.; WASSERMAN, W. **Applied linear statistical models**. New York: McGraw-Hill, v.1996. 1996. 1408 p.

OLIVEIRA, J. C.; EPIPHANIO, J. C. N. Noise reduction in MODIS NDVI time series data based on spatial-temporal analysis. In: Geoscience and Remote Sensing Symposium (IGARSS), 2012, Munique. **Proceedings...** Alemanha: IEEE International, 22-27 Jul. 2012. p. 2372–2375. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6350807>>. Acesso em: 21 out. 2013.

OLIVEIRA, J. C.; EPIPHANIO, J. C. N.; RENNO, C. D. The use of spatial-temporal analysis for noise reduction in MODIS NDVI time series data. In: Tenth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy), 2012, Florianópolis. **Proceedings ...** Santa Catarina: International Spatial Accuracy Research Association, 8-11 Jul. 2012. Disponível em: <<http://spatial-accuracy.org/system/files/OliveiraAccuracy2012.pdf>>. Acesso em: 05 nov. 2013.

PITTMAN, K.; HANSEN, M. C.; BECKER-RESHEF, I.; POTAPOV, P. V.; JUSTICE, C. O. Estimating Global Cropland Extent with Multi-year MODIS Data. **Remote Sensing**, v. 2, n. 7, p. 1844–1863, Jul. 2010.

POGGIO, L.; GIMONA, A.; BROWN, I. Spatio-temporal MODIS EVI gap filling under cloud cover: An example in Scotland. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 72, p. 56–72, Aug. 2012.

PRINGLE, M. J.; SCHMIDT, M.; MUIR, J. S. Geostatistical interpolation of SLC-off Landsat ETM+ images. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 64, n. 6, p. 654–664, Nov. 2009.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em 21 out. de 2013.

RAMACHANDRAN, B.; JUSTICE, C. O.; ABRAMS, M. J. (Eds.) **Land remote sensing and global environmental change: NASA's Earth observing system and the science of ASTER and MODIS**. Springer, 2011. 873 p. ISBN (978-1-4419-6749-7).

ROERINK, G. J.; MENENTI, M.; VERHOEF, W. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. **International Journal of Remote Sensing**, v. 21, n. 9, p. 1911–1917, 2000.

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

ROY, D. P.; BORAK, J.S.; DEVADIGA, S.; WOLFE, R.E.; ZHENG, M.; DESCLOITRES, J. The MODIS Land product quality assessment approach. **Remote Sensing of Environment**, v.83, n.1-2, p.62-76. 2002.

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, n. 4, p. 1057–1076, Abr. 2010.

SAKAMOTO, T.; YOKOZAWA, M.; TORITANI, H.; SHIBAYAMA, M.; ISHITSUKA, N.; OHNO, H. A crop phenology detection method using time-series MODIS data. **Remote Sensing of Environment**, v. 96, n. 3-4, p. 366–374, Jun. 2005.

SAKAMOTO, T.; WARDLOW, B. D.; GITELSON, A. A.; VERMA, S. B.; SUYKER, A. E.; ARKEBAUER, T. J. A Two-Step Filtering approach for detecting maize and soybean phenology with time-series MODIS data. **Remote Sensing of Environment**, v. 114, n. 10, p. 2146–2159, Jun. 2010.

SALOMONSON, V.; ABRAMS, M. J.; KAHLE, A.; BARNES, W.; XIONG, X.; YAMAGUCHI, Y. Evolution of NASA's Earth Observing System and Development of the Moderate-Resolution Imaging Spectroradiometer and the Advanced Spaceborne Thermal Emission and Reflection Radiometer Instruments. In: RAMACHANDRAN, B.; JUSTICE, C. O.; ABRAMS, M. J. (Eds.). **Land remote sensing and global environmental change: NASA's Earth Observing System and the Science of ASTER and MODIS**. Springer, 2011. 873 p. ISBN (978-1-4419-6749-7).

SANCHES, I. D. **Sensoriamento remoto para o levantamento espectro-temporal e estimativa de área de culturas agrícolas**. 2004. 172p. (INPE-10290-TDI/909) Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São Jose dos Campos, 2004. Disponível em: <<http://urlib.net/rep/sid.inpe.br/jeferson/2004/05.14.10.59?languagebutton=pt-BR>>. Acesso em: 21 set. 2013.

SANO, E. E.; FERREIRA, L. G.; ASNER, G. P.; STEINKE, E. T. Spatial and temporal probabilities of obtaining cloud-free Landsat images over the Brazilian tropical

savanna. **International Journal of Remote Sensing**, v. 28, n. 12, p. 2739–2752, Jun. 2007.

SARGENT, J.; BEDFORD, A. Improving Australian Football League player performance forecasts using optimized nonlinear smoothing. **International Journal of Forecasting**, v. 26, n. 3, p. 489–497, Jul. 2010.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, Jul. 1964.

SWETS, D. L.; REED, B. C.; ROWLAND, J. D.; MARKO, S. E. A Weighted least-squares approach to temporal NDVI smoothing. AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING, 1999, Oregon. **Proceedings...** Portland, Oregon, 1999. p. 526–536.

TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. **Economic Geography**, v. 46, p. 234–240, Jun. 1970.

VELLEMAN, P. F. Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms. **Journal of the American Statistical Association**, v. 75, n. 371, p. 609–615, Set. 1980.

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 1, p. 106–115, Jan. 2010.

VERHOEF, W.; MENENTI, M.; AZZALI, S. A colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981-1992). **International Journal of Remote Sensing**, v. 17, n. 2, p. 231–235, Jan. 1996.

VIOVY, N.; ARINO, O.; BELWARD, A. S. The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. **International Journal of Remote Sensing**, v. 13, n. 8, p. 1585–1590, 1992.

WARDLOW, B.; EGBERT, S.; KASTENS, J. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. **Remote Sensing of Environment**, v. 108, n. 3, p. 290–310, Jun. 2007.

WILLMOTT, C.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research**, v. 30, p. 79–82, Dez. 2005.

ZHANG, X.; FRIEDL, M. A.; SCHAAF, C. B.; STRAHLER, A. H.; HODGES, J. C. F.; GAO, F.; REED, B. C.; HUETE, A. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, v. 84, n. 3, p. 471–475, Mar. 2003.

ZHANG, C.; LI, W.; TRAVIS, D. J. Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging. **International Journal of Remote Sensing**, v. 30, n. 9, p. 2173–2195, Maio 2009.

ZHOU, J.; JIA, L.; HU, G.; MENENTI, M. Evaluation of Harmonic Analysis of Time Series (HANTS): impact of gaps on time series reconstruction. In: INTERNATIONAL WORKSHOP ON EARTH OBSERVATION AND REMOTE SENSING APPLICATIONS, 2., 2012, Shanghai, China.. **Proceedings...** Shanghai, China: Earth Observation and Remote Sensing Applications (EORSA), jun. 2012. p. 31–35.
Disponível em:
<<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6261129>>. Acesso em: 21 set. 2013.

ZHU, X.; LIU, D.; CHEN, J. A new geostatistical approach for filling gaps in Landsat ETM+ SLC-off images. **Remote Sensing of Environment**, v. 124, p. 49–60, Set. 2012.

APÊNDICE A – Tabelas e Figuras referentes à amostragem “II”

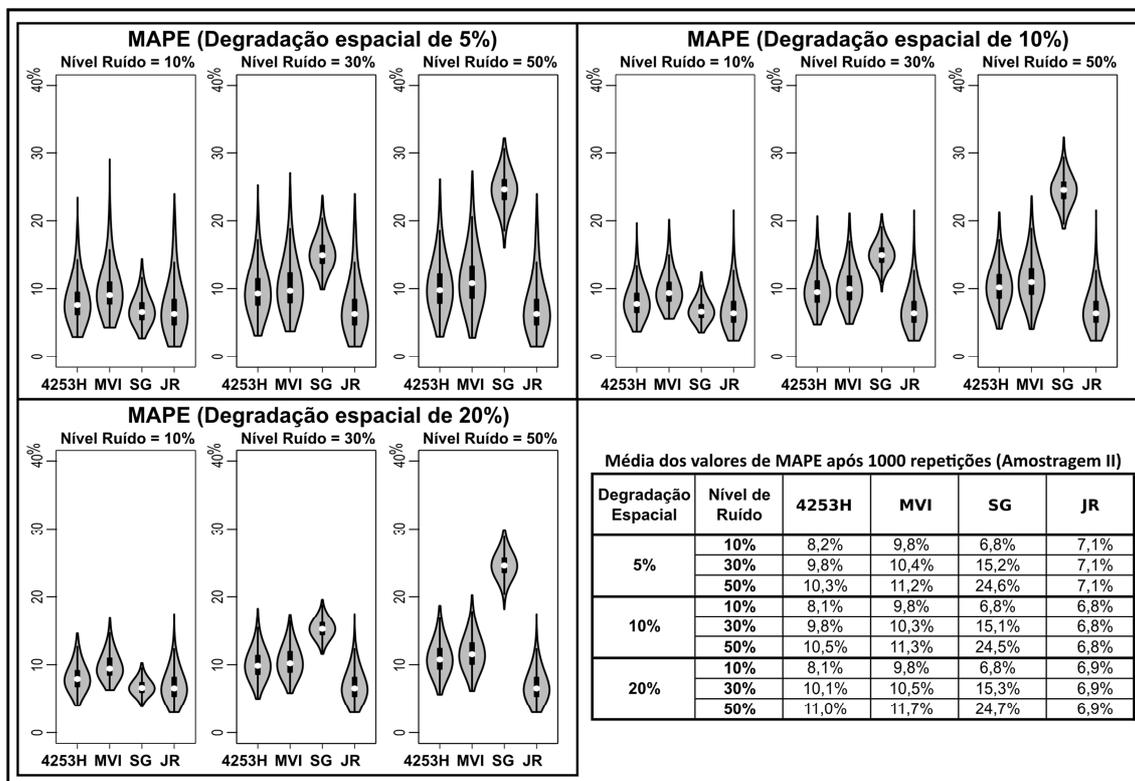


Figura A.1 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “II” realizadas na Área “B”.

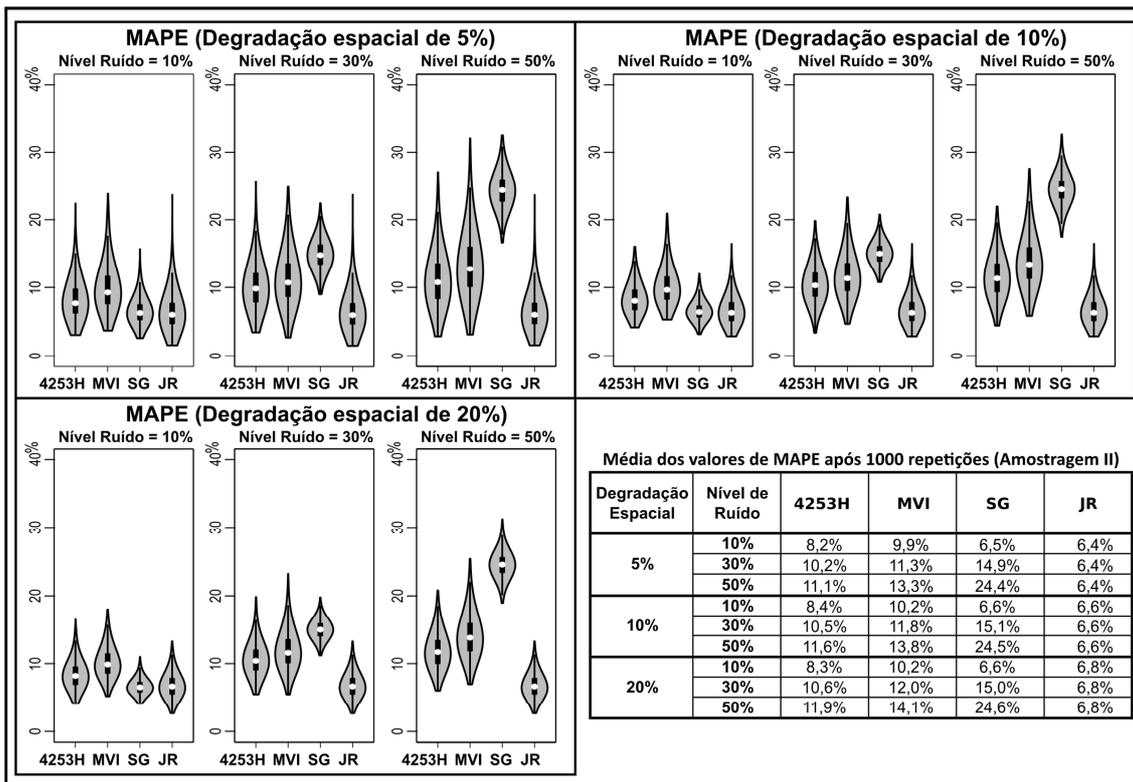


Figura A.2 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “II” realizadas na Área “C”.

APÊNDICE B – Tabelas e Figuras referentes à amostragem “III”

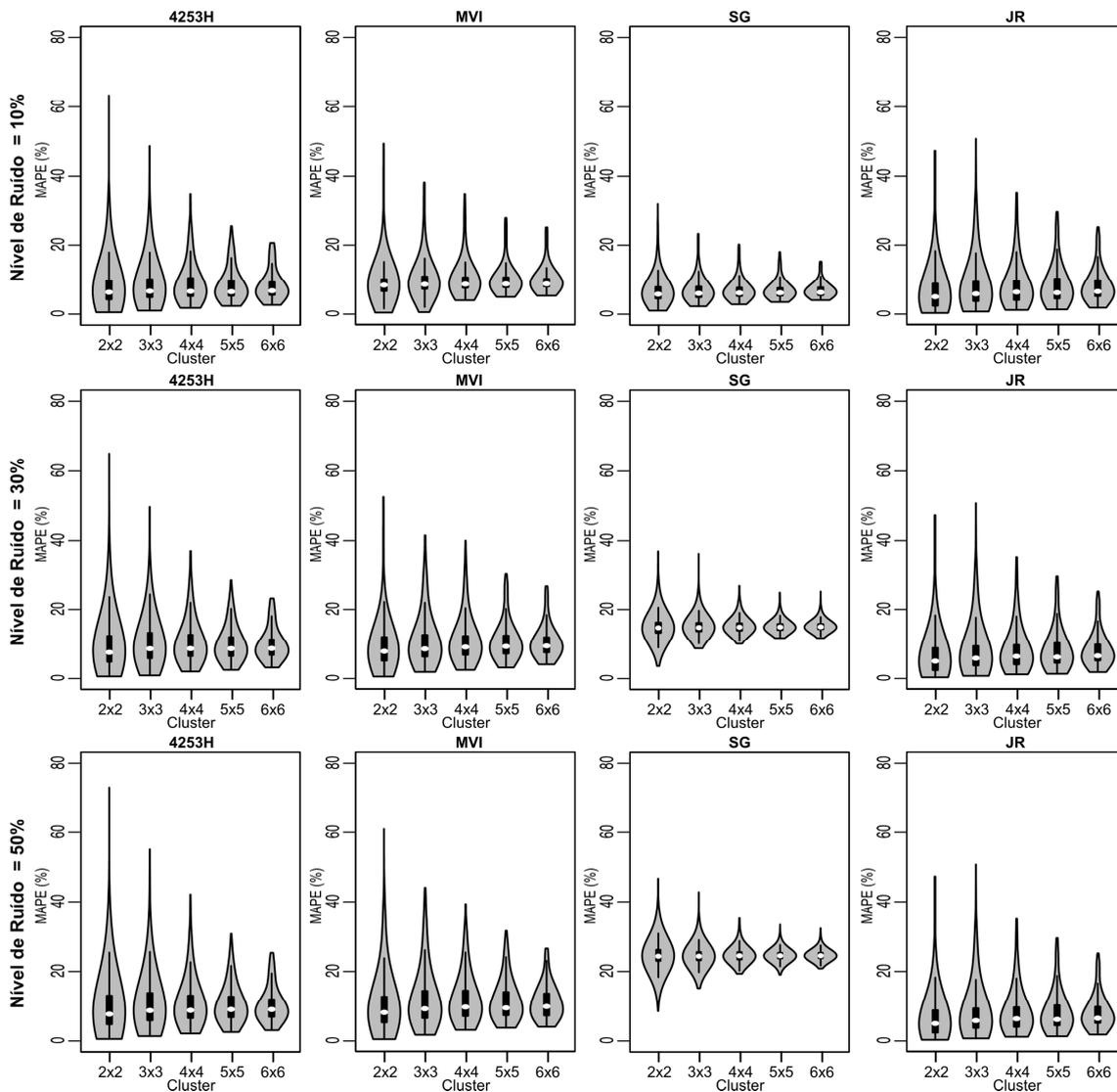


Figura B.1 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “III” realizadas na Área “B”.

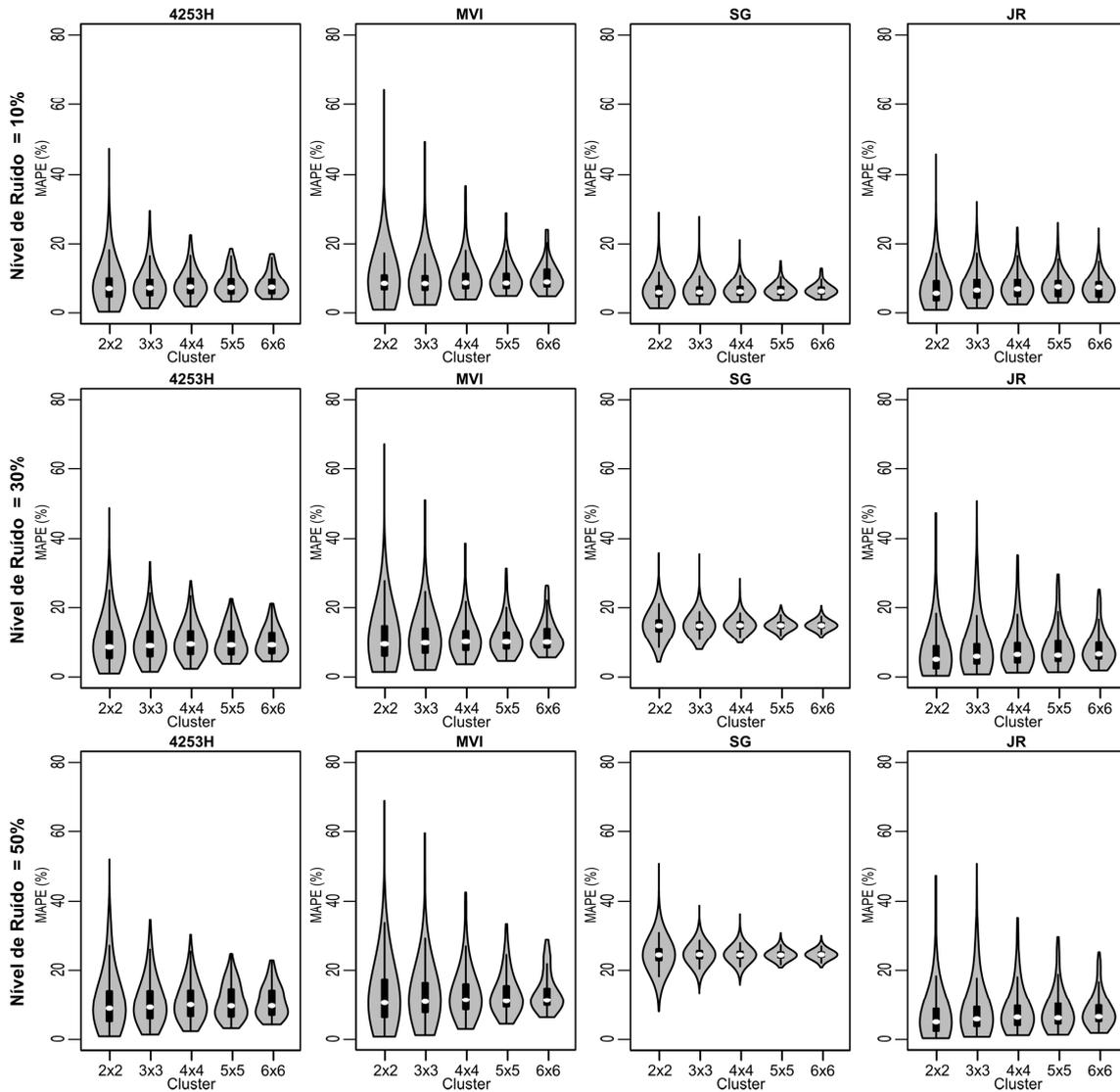


Figura B.2 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “III” realizadas na Área “C”.

APÊNDICE C – Tabelas e Figuras referentes à amostragem “IV”

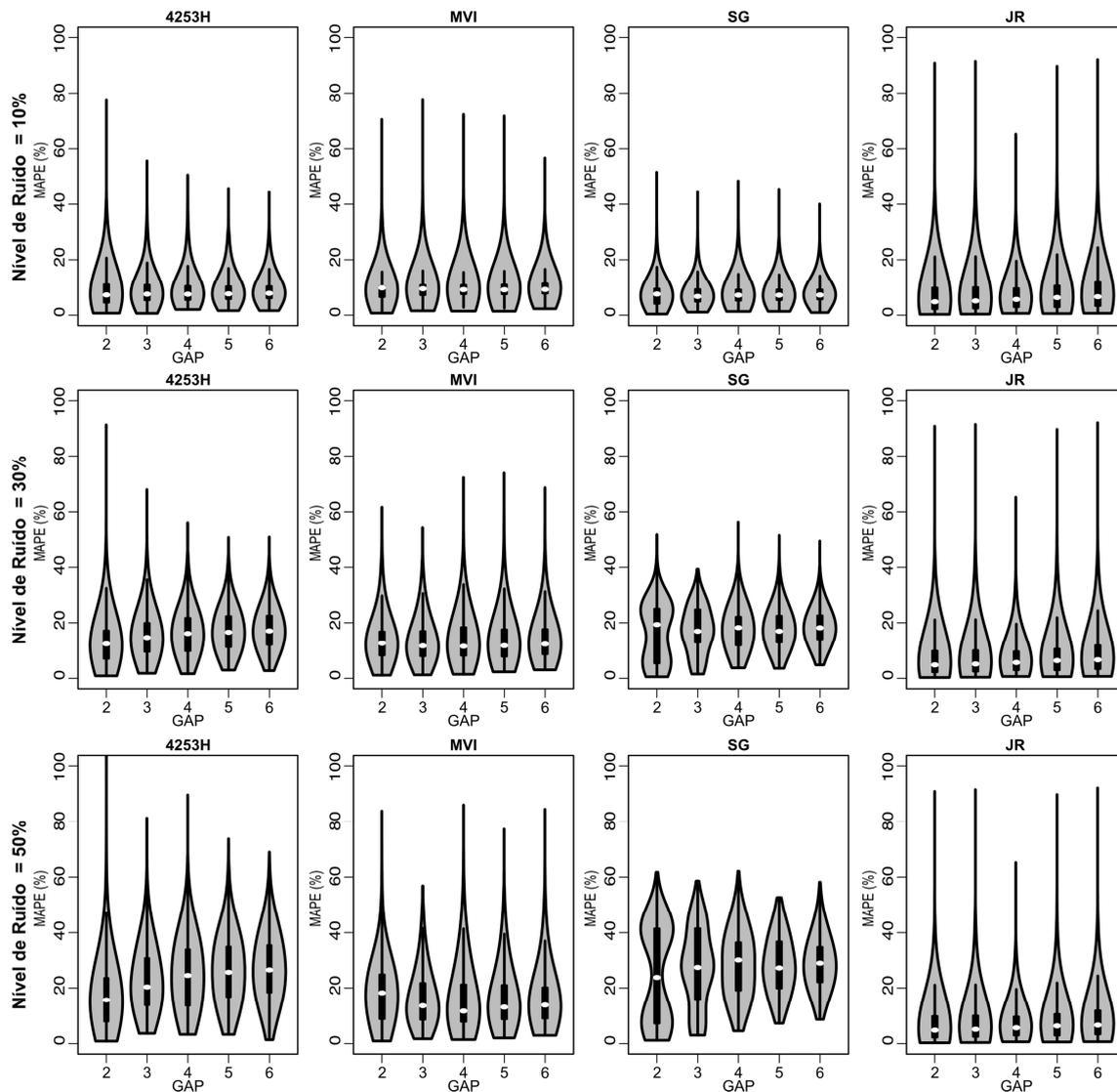


Figura C.1 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “IV” realizadas na Área “B”.

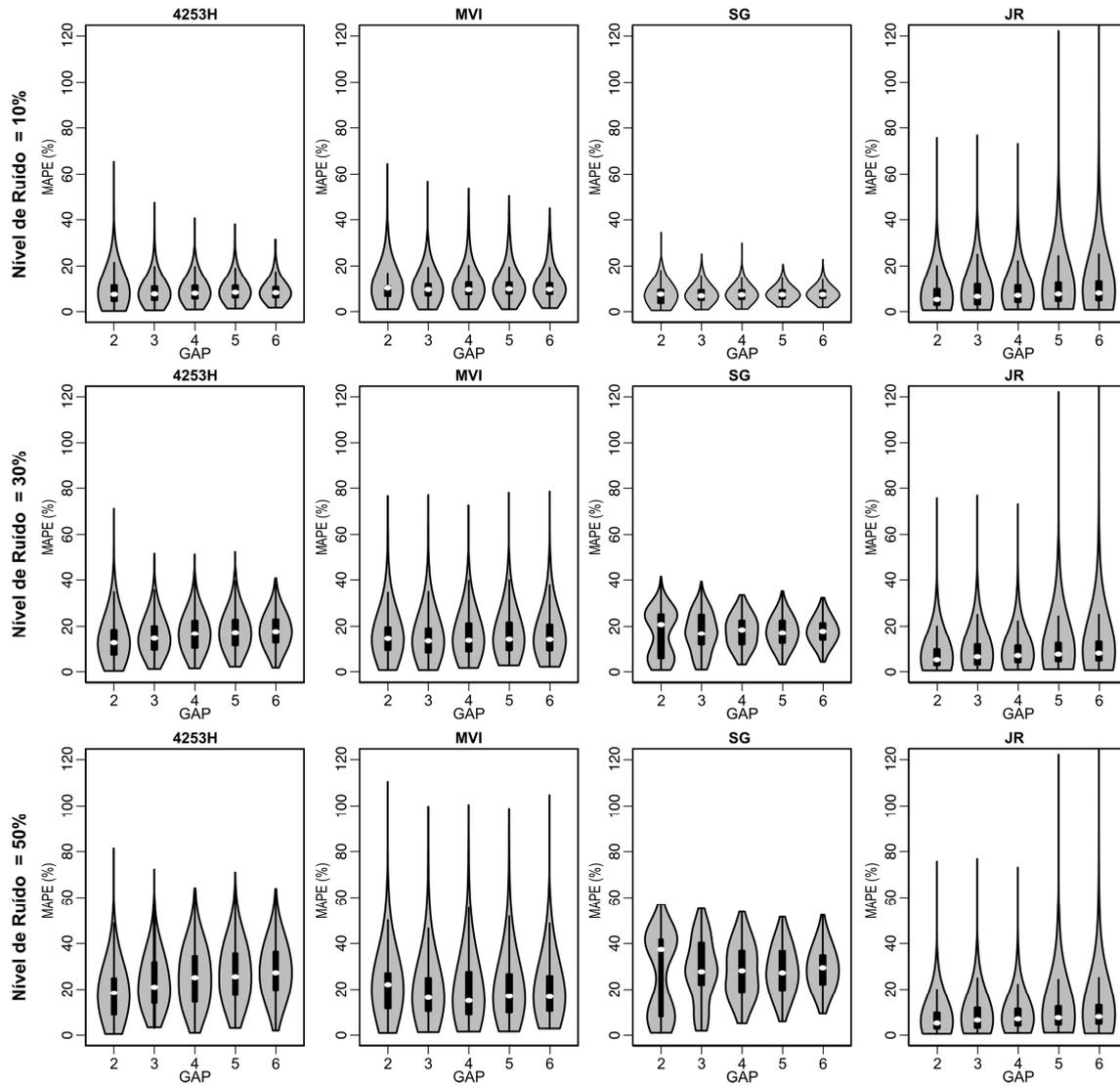


Figura C.2 – Boxplot (na forma de violino) do erro percentual médio absoluto para as 1000 iterações da amostragem “IV” realizadas na Área “C”.

APÊNDICE D – Resultado do Teste não paramétrico

Teste não paramétrico de Mann-Whitney utilizado para comparar as tendências centrais dos valores de MAPE obtidos para os *gaps* simulados conforme o tipo de filtro avaliado. O teste Mann-Whitney pode ser considerado a versão não paramétrica do teste t para amostras independentes. Ao contrário do teste t , que testa a igualdade das médias, o teste de Mann-Whitney testa a igualdade das medianas.

As hipóteses utilizadas no teste são:

H1: os grupos pertencem à mesma população, ou as medianas são iguais.

H0: os grupos avaliados são diferentes, ou as medianas são diferentes.

Tabela D.1 – Valores de p-valor para o teste de Mann-Whitney entre os clusters simulados para a área teste “A”.

Nível De Ruído	4253H twice					MVI					SG					JR					
	Cluster Simulado					Cluster Simulado					Cluster Simulado					Cluster Simulado					
	2x2	3x3	4x4	5x5	6x6	2x2	3x3	4x4	5x5	6x6	2x2	3x3	4x4	5x5	6x6	2x2	3x3	4x4	5x5	6x6	
10%	2x2	1	0,616	0,080	0,000	0,000	1	0,910	0,333	0,000	0,000	1	0,011	0,007	0,000	0,000	1	0,009	0,000	0,000	0,000
	3x3	---	1	0,240	0,000	0,001	---	1	0,365	0,000	0,000	---	1	0,982	0,000	0,005	---	1	0,079	0,000	0,000
	4x4	---	---	1	0,000	0,026	---	---	1	0,000	0,006	---	---	1	0,000	0,004	---	---	1	0,000	0,000
	5x5	---	---	---	1	0,790	---	---	---	1	0,250	---	---	---	1	0,147	---	---	---	1	0,761
	6x6	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1
30%	2x2	1	0,897	0,199	0,000	0,010	1	0,068	0,505	0,000	0,001	1	0,260	0,500	0,000	0,000	1	0,009	0,000	0,000	0,000
	3x3	---	1	0,294	0,000	0,002	---	1	0,812	0,001	0,004	---	1	0,448	0,003	0,005	---	1	0,079	0,000	0,000
	4x4	---	---	1	0,000	0,028	---	---	1	0,001	0,004	---	---	1	0,016	0,033	---	---	1	0,000	0,000
	5x5	---	---	---	1	0,077	---	---	---	1	0,658	---	---	---	1	0,765	---	---	---	1	0,761
	6x6	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1
50%	2x2	1	0,614	0,107	0,000	0,000	1	0,761	0,107	0,000	0,000	1	0,019	0,009	0,000	0,000	1	0,009	0,000	0,000	0,000
	3x3	---	1	0,299	0,000	0,001	---	1	0,141	0,000	0,000	---	1	0,899	0,319	0,201	---	1	0,079	0,000	0,000
	4x4	---	---	1	0,000	0,019	---	---	1	0,000	0,000	---	---	1	0,430	0,262	---	---	1	0,000	0,000
	5x5	---	---	---	1	0,110	---	---	---	1	0,810	---	---	---	1	0,674	---	---	---	1	0,761
	6x6	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1

Obs.: Os testes foram realizados por meio de uma comparação de conjuntos de dados na seguinte ordem: linha/coluna.

Os p-valores > 0,05 foram destacados na cor amarela.

Tabela D.2 – Valores de p-valor para o teste de Mann-Whitney entre os gaps simulados para a área teste “A”.

Nível De Ruído	4253H twice					MVI					SG					JR					
	Gap Simulado					Gap Simulado					Gap Simulado					Gap Simulado					
	Gap 2	Gap 3	Gap 4	Gap 5	Gap 6	Gap 2	Gap 3	Gap 4	Gap 5	Gap 6	Gap 2	Gap 3	Gap 4	Gap 5	Gap 6	Gap 2	Gap 3	Gap 4	Gap 5	Gap 6	
10%	Gap 2	1	0,000	0,000	0,000	0,000	1	0,083	0,016	0,000	0,000	1	0,125	0,001	0,000	0,000	1	0,001	0,000	0,000	0,000
	Gap 3	---	1	0,048	0,002	0,000	---	1	0,469	0,054	0,000	---	1	0,069	0,029	0,000	---	1	0,279	0,006	0,000
	Gap 4	---	---	1	0,277	0,001	---	---	1	0,243	0,003	---	---	1	0,760	0,005	---	---	1	0,107	0,012
	Gap 5	---	---	---	1	0,031	---	---	---	1	0,094	---	---	---	1	0,010	---	---	---	1	0,323
	Gap 6	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1
30%	Gap 2	1	0,000	0,000	0,000	0,000	1	0,021	0,051	0,007	0,000	1	0,078	0,006	0,072	0,000	1	0,001	0,000	0,000	0,000
	Gap 3	---	1	0,001	0,000	0,000	---	1	0,701	0,597	0,007	---	1	0,044	0,228	0,000	---	1	0,279	0,006	0,000
	Gap 4	---	---	1	0,339	0,007	---	---	1	0,340	0,003	---	---	1	0,395	0,018	---	---	1	0,107	0,012
	Gap 5	---	---	---	1	0,096	---	---	---	1	0,046	---	---	---	1	0,001	---	---	---	1	0,323
	Gap 6	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1
50%	Gap 2	1	0,000	0,000	0,000	0,000	1	0,003	0,018	0,501	0,516	1	0,163	0,053	0,235	0,073	1	0,001	0,000	0,000	0,000
	Gap 3	---	1	0,000	0,000	0,000	---	1	0,900	0,033	0,001	---	1	0,117	0,211	0,006	---	1	0,279	0,006	0,000
	Gap 4	---	---	1	0,063	0,000	---	---	1	0,044	0,001	---	---	1	0,860	0,033	---	---	1	0,107	0,012
	Gap 5	---	---	---	1	0,025	---	---	---	1	0,215	---	---	---	1	0,032	---	---	---	1	0,323
	Gap 6	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1	---	---	---	---	1

Obs.: Os testes foram realizados por meio de uma comparação de conjuntos de dados na seguinte ordem: linha / coluna.

Os p-valores > 0,05 foram destacados na cor amarela.

APÊNDICE E – Código R - Janela de Regressão

A seguir estão descritos os comandos implementados no software R (versão 3.0.1) para estimar novos valores de NDVI conforme proposto pela metodologia Janela de Regressão.

```
#####
#                                                                 #
# Janela de Regressão: uma análise espacial e temporal para estimar #
# valores de NDVI classificados com baixa qualidade em séries temporais #
# MODIS                                                                 #
#                                                                 #
# versão 1.0                                                                 #
# Script desenvolvido como parte das análises realizadas na Tese de Doutorado #
# do Curso de Pós-Graduação em Sensoriamento Remoto. #
# Autor: Julio C. Oliveira (oliveirajc@ufv.br) #
# Orientador: Dr. José Carlos Neves Epiphanyo #
#                                                                 #
#####

# =====
# imagens de entrada
# 1) série temporal com valores de índice de vegetação (ex.: imagens do
# produto MODIS MOD13Q1 - NDVI)
# 2) imagens de confiabilidade (ex.: produto reliability/MODIS)
# imagem de saída
# 1) NDVI_JR (com valores de NDVI estimados para os pixels classificados
# com baixa qualidade)
# =====

# Início do Script -----

# Arquivo para armazenar o resultado das análises de regressao
```

```

# dataframe com n dimensoes
(linha,coluna,numero_datas,pixelvizinho,janela_temporal,resultados[NDVI_estimado,r2_da_regressao,variancia_NETER
_novos_valores)
  result <- array(data = NA, dim = c(nrow(NDVI),ncol(NDVI),nlayers(NDVI),8,5,3))

# imagem de NDVI gerada após análise da metodologia Janela de Regressao
  NDVI_JR <- NDVI

# BEGIN - este arquivo apresenta o delta i e delta j para o centro de uma mascara 3x3
  pv <- data.frame(c(-1,-1,-1,0,0,1,1,1), c(-1,0,1,-1,1,-1,0,1)) # pv - pixel vizinho -- mascara 3x3
  colnames(pv) <- c("i","j")
# END

# Com base nos valores de qualidade da série temporal (imagens de reliability do produtos MODIS)
# perguntar ao usuario para qual o valor de reliability será estimado novos valores de NDVI

  cat(" What reliability will be recalculated of the NDVI value?", "\n" )
  cat(" 1 - Marginal data (Useful, but look at other QA information)", "\n")
  cat(" 2 - Snow/Ice (Target covered with snow/ice)", "\n" )
  cat(" 3 - Cloudy (Target not visible, covered with cloud)", "\n" )
  cat(" 4 - All options together", "\n" )
  opc <- readline ()

# Atribuir NA para pixels com baixa qualidade -----

# pwp - tabela com todos os pixels que foram classificados com baixa qualidade
if (opc == 4) pwp <- which((as.matrix(reliability))!=0, arr.ind=T)
if (opc != 4) pwp <- which((as.matrix(reliability))==opc, arr.ind=T)
pwp <- data.frame(pwp)
colnames(pwp) <- c("pixel", "date")

```

```

# Atribuir o valor de NA, na imagem NDVI_JR, para os pixels com baixa qualidade
if (opc == 4) NDVI_JR@data@values[reliability@data@values != 0] <- NA
if (opc != 4) NDVI_JR@data@values[reliability@data@values == opc] <- NA

# -----

ni <- 0 # numero de iteracoes
DNA <- sum(is.na(as.matrix(NDVI_JR))) # quantidade de pixels com valor NA na imagem NEDVI_rec2

pwp$check <- NA # atribuir NA para todos os campos da coluna CHECK na tabela pwp

# Rotina para estimar novos valores -----

while (DNA!=0) # realizar iterações enquanto houver pixels com valor NA na imagem NDVI_JR
{
  # atribuir uma ordem aleatoria para os pixels a serem avaliados
  pwp <- pwp[sample(nrow(pwp)),]
  ni <- ni + 1
  DNA_before <- sum(is.na(as.matrix(NDVI_JR)))

  for (p in 1:dim(pwp)[1]) # analisar os pontos da tabela pwp após aplicar atribuir ordem aleatória
  {
    i <- arrayInd(pwp$pixel[p], c(ncol(NDVI),nrow(NDVI)))[1,2] # valor da linha
    j <- arrayInd(pwp$pixel[p], c(ncol(NDVI),nrow(NDVI)))[1,1] # valor da coluna

    if (is.na(NDVI_JR[i,j][pwp$date[p]])) # se o pixel tem valor NA - aplicaar a Janela de regressao
    {

```

```

for (msc in 1:8) # analise para cada um dos 8 pixels vizinhos ao pixel ij -- com base em uma mascara 3x3
{
  if (is.na(NDVI_JR[i+pv$i[msc],j+pv$j[msc]][pwp$date[p]]) == FALSE) # efetuar a análise somente se o PV
possuir qualidade satisfatória (ou valor diferente de NA)
  {
    # Seleção dos dados com base no perfil temporal --- janelas variando de 2 a 5 datas
    for (win in 2:5) # janela temporal para as datas - perfil temporal
    {
      ini <- pwp$date[p]-win # data inicial para selecao dos dados
      if (ini <= 0) ini<-1 # data inicial começa com 1, pois nao pode ter valor 0 ou negativo
      end <- pwp$date[p]+win # data final para selecao dos dados
      if (end >= nlayers(NDVI))
      end <- nlayers(NDVI) # a maior data possivel tem que ser igual ao numero de bandas da
imagem NDVI

      # A Janela de Regressão estipula que pelos menos 2 pares de dados (tanto antes como depois da data de
interesse) devem estar aptos para análise

      if ((sum(!is.na(NDVI_JR[i,j][c((ini):(pwp$date[p])]) *
NDVI_JR[i+pv$i[msc],j+pv$j[msc]][c((ini):(pwp$date[p])])))) > 0 && (sum(!is.na(NDVI_JR[i,j][c((pwp$date[p]):(end))]) *
NDVI_JR[i+pv$i[msc],j+pv$j[msc]][c((pwp$date[p]):(end))])) > 0 && (sum(!is.na(NDVI_JR[i,j][c((ini):(end))]) *
NDVI_JR[i+pv$i[msc],j+pv$j[msc]][c((ini):(end))])) > 2 )
      {
        # Analise de regressao baseado nos dados selecionados pela janela temporal
        reg <- lm(NDVI_JR[i,j][c((ini):(end))]~NDVI_JR[i+pv$i[msc],j+pv$j[msc]][c((ini):(end))])
        result[i,j,pwp$date[p],msc,win,1] <-
          as.numeric(reg$coefficients[2]*NDVI_JR[i+pv$i[msc],j+pv$j[msc]][pwp$date[p]] +
reg$coefficients[1])
        result[i,j,pwp$date[p],msc,win,2] <- summary(reg)$r.squared
      }
    }
  }
}

```

```

# O valor da variável independente para a data de interesse eh considerado como novo valor (Xnovo)
novo <- NDVI_JR[i+pv$i[msc],j+pv$j[msc]][pwp$date[p]]

# calculando a variancia para novos valores com base no livro do NETER et al. (1996, pag. 59)

result[i,j,pwp$date[p],msc,win,3] <- as.numeric(summary(reg)$sigma^2 * (1 +
(1/length(reg$model[,2])) +
(((novo - mean(reg$model[,2]))^2) /
(sum((reg$model[,2] - mean(reg$model[,2]))^2))))))
  } # end if (par de dados)
} # end for (win)
} # end if (PV)
} # end for (msc)

# Substituido o valor estimado para o Pixel (i,j) na imagem NDVI_JR
if (sum(!is.na(result[i,j,pwp$date[p],,2]))!=0) # alterar o valor original de NDVI somente para pontos q possuem
regressao calculada
{
  # para cada janela temporal (msc 2,3,4 e 5) selecionar as menores variancias e os respectivos valores de NDVI
estimados
  # o valor do pixel de interesse será substituído pela mediana dos valores de NDVI selecionados em cada janela
temporal
  NDVI_JR[i,j][pwp$date[p]] <-
  median(c (result[i,j,pwp$date[p],,2,1][which.min(result[i,j,pwp$date[p],,2,3])],
  result[i,j,pwp$date[p],,3,1][which.min(result[i,j,pwp$date[p],,3,3])],
  result[i,j,pwp$date[p],,4,1][which.min(result[i,j,pwp$date[p],,4,3])],
  result[i,j,pwp$date[p],,5,1][which.min(result[i,j,pwp$date[p],,5,3])]), na.rm=T)
  cat("==", round( (1- sum(is.na(pwp$check))/dim(pwp)[1] ) *100,2),"%", "\n")
}

```

```

        pwp$check[p]<-"Ok" # alterar o campo Check (de NA para OK) para o pixel calculado
    } # end if (substituir valor NDVI na imagem NDVI_JR2)
  } # end if (testando se pixel é NA)
} # end for (p)

DNA_after <- sum(is.na(as.matrix(NDVI_JR)))
DNA <- DNA_before - DNA_after

} # end while

# deletar variáveis
rm(list=c("DNA", "DNA_after", "DNA_before", "end", "i", "ini", "j", "msc", "ni",
        "novo", "opc", "p", "pv", "pwp", "reg", "result", "win"))

# Fim do Script -----

```