



Comparação de algoritmos de filtragem em séries temporais de NDVI/MODIS

Comparison of smoothing algorithms in NDVI MODIS time series

*Pedro José Farias Fernandes*¹
*Raúl Sanchez Vicens*²
*Luiz Felipe de Almeida Furtado*²

Recebido em setembro de 2017.
Aprovado em setembro de 2018.

RESUMO

O uso de séries temporais de Índices de Vegetação (IVs) necessita de uma etapa de pré-processamento, devido a ruídos causados pela dinâmica temporal. Para diminuir tal efeito, aplica-se técnicas de filtragem nas séries temporais, porém, não há consenso de qual é a técnica mais apropriada. Portanto, o objetivo é avaliar o desempenho dos filtros de média, mediana, Savitzky-Golay, 4253H twice, Whittaker 1 e 2, em séries temporais (2000-2015) de NDVI (filtros escolhidos por estarem implementados em R, exigirem poucos parâmetros de entrada, e indicação na literatura). Para cada ponto de referência das classes de cobertura agricultura, floresta, pastagem e silvicultura, foram geradas séries temporais ruidosas para cada nível de ruído (10%, 40% e 70%). O desempenho dos filtros foi comparado segundo o erro padrão encontrado. Os valores de RMSE dos filtros foram próximos para os três níveis de ruído. Nas simulações de 10% e 40% de ruído, o menor erro foi para o filtro de mediana e o segundo menor para o filtro 4253H twice. Para a simulação com 70% de ruído, o menor erro foi para o filtro de média, o segundo para o filtro de mediana e o terceiro para o 4253H twice. Entretanto, o filtro de mediana modificou mais a forma da série temporal em relação ao filtro 4253H twice. Os resultados geram conhecimento para futuros estudos que utilizem dados de IVs MODIS para análise da cobertura da terra no bioma Mata Atlântica.

PALAVRAS-CHAVE: Suavização. Ruído. Índice de vegetação.

¹ Universidade Federal Fluminense – UFF - Av. Gen. Milton Tavares de Souza s/n - Instituto de Geociências – Niterói – RJ, Brasil. E-mail: pj_fernandes@id.uff.br, rvicens@gmail.com

² Universidade Federal do Rio de Janeiro – UFRJ - Rua Athos da Silveira Ramos, n. 274, Bloco I, CEP: 21941-916 - Rio de Janeiro – RJ, Brasil. E-mail: furtadosere@gmail.com

ABSTRACT

The use of Vegetation Indices (VIs) time series requires a preprocessing step due to noise caused by temporal dynamics. To reduce this effect, filtering techniques are applied in the time series, however, there is no consensus as to which technique is most appropriate. The objective is to evaluate the performance of the mean, median, Savitzky-Golay, 4253H twice, Whittaker 1 and 2 filters in NDVI time series (2000-2015) (filters chosen because they are in R language, need of a few input parameters, and use in the literature). For each reference point of agriculture, forest, pasture and silviculture classes, noisy time series were generated for each noise level (10%, 40% and 70%). The filters performances were compared using the standard errors. The filters RMSE values were close in the three noise levels. In the 10% and 40% noise simulations, the smallest error was for the median filter and the second smaller for the 4253H twice. For the simulation with 70% of noise, the smallest error was for the medium filter, the second for the median filter and the third for the 4253H twice. However, the median filter modified the time series shape than the 4253H twice. The results generate knowledge for future studies that use MODIS VIs to analyze land cover in the Atlantic Forest biome.

KEYWORDS: Smoothing. Noise. Vegetation index.

* * *

Introdução

As séries temporais MODIS (*Moderate Resolution Imaging Spectroradiometer*) fornecem produtos para comparações da cobertura vegetal, úteis para o monitoramento da atividade fotossintética e da cobertura da terra. Entre estes produtos, há os índices de vegetação (IVs), como o NDVI (*Normalized Difference Vegetation Index*) (ROUSE et al., 1973) e o EVI (*Enhanced Vegetation Index*) (HUETE et al., 2002), que exploram o comportamento distinto da vegetação na faixa espectral do visível (VIS) e do infravermelho próximo (IVP) (PONZONI e SHIMABUKURO, 2009).

A disponibilidade desses dados possibilita capturar as dinâmicas das mudanças na vegetação e ajudar na interpretação de suas tendências (BEURS et al., 2015), uma vez que os fatores perturbadores da cobertura vegetal modificam os IVs, pois são influenciados pela temperatura da superfície, evapotranspiração e condições de umidade no geral (MILDREXLER et al., 2016; BEURS et al., 2015). Por isso, IVs como o NDVI

ou EVI, são relacionados com produtividade vegetal (MISHRA et al., 2015), variáveis biofísicas (biomassa, índice de área foliar, teor de clorofila, etc.) (CHEN et al., 2014), e vigor vegetativo (PONZONI e SHIMABUKURO, 2009).

Desse modo, pode-se utilizar esses dados para a detecção de tendências na cobertura vegetal e no entendimento dos fatores influenciadores dessa dinâmica, sejam eles antrópicos ou naturais (como tendências climáticas regionais). Por exemplo, Ju e Masek (2016) identificam tendências na vegetação do Alasca e Canadá, e fazem relação com mudanças globais. Estas tendências nos IVs podem ser positivas (chamada de *greening* na literatura, relacionada com aumento do vigor vegetativo e/ou regeneração) e negativas (*browning*, relacionada com redução do vigor vegetativo e/ou mudanças como desmatamento).

Porém, o uso de séries temporais de IVs necessita de uma etapa de pré-processamento, devido a ruídos causados por nuvens, sombras, variações no sensor e no tempo (SHAO et al., 2016). Para diminuir o efeito dos ruídos, aplica-se técnicas de filtragem/suavização nas séries temporais. Há diversos trabalhos na literatura que testam filtros nos IVs MODIS.

Shao et al. (2016) comparam os filtros *Savitzky-Golay*, Gaussiano assimétrico, duplo logístico, de *Whittaker* e de transformada de *Fourier* discreta para a classificação de cobertura da terra a partir de NDVI MODIS. Os autores concluem que a transformada discreta de *Fourier* e o filtro *Whittaker* melhoraram a separabilidade entre as classes e a exatidão global. Michishita et al. (2014) aplicam em dados NDVI MODIS os filtros de *Hanning*, *Savitzky-Golay*, mediana móvel, operação máxima, média e de processamento de *endpoint*. Estes autores concluem que os mais adequados para a filtragem foram os dois primeiros.

Borges et al. (2014) comparam os filtros duplo logístico, de *Savitzky-Golay* e gaussiano assimétrico em série temporal de EVI MODIS e mostram que o primeiro apresentou melhor desempenho. Hird e McDermid (2009) comparam as técnicas gaussiana assimétrica, duplo logística, *Savitzky-Golay*, 4253H *twice*, iteração do valor médio, e ARMD3-ARMA5. Os autores

mostraram que para seus dados, os filtros gaussianos assimétricos e duplo logístico tiveram desempenho superior em dados NDVI MODIS. Chen et al. (2004) apresentaram um método para reconstrução de série temporal NDVI MODIS baseada na técnica de *Savitzky-Golay*, e concluem que foi eficiente para obter uma série temporal de alta qualidade. Adami (2010) compara os filtros *4253H twice*, *Savitzky-Golay*, duplo logístico, gaussiano assimétrico, *Hants* e *Wavelets*, e concluiu que o último apresentou melhor desempenho para estimar datas de plantio de soja.

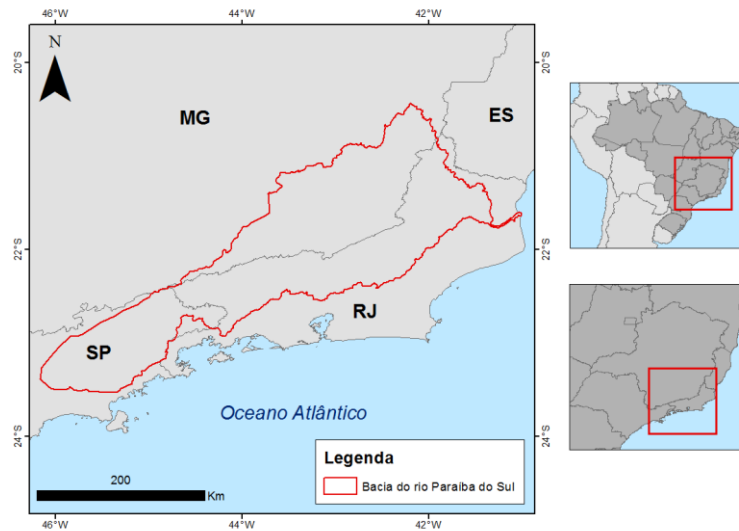
Diante dos diferentes resultados obtidos na literatura e das diversas áreas de estudo inseridas em biomas distintos nos trabalhos e por não haver consenso de qual é a técnica de suavização mais apropriada para sensoriamento remoto (ABADE et al., 2015), há a necessidade de comparar o desempenho de diferentes filtros para a posterior análise de séries temporais de IVs MODIS no bioma Mata Atlântica do sudeste brasileiro, em especial, da bacia do rio Paraíba do Sul. Portanto, o objetivo deste trabalho é avaliar o desempenho dos filtros de média, mediana, *Savitzky-Golay*, *4253H twice*, *Whittaker 1* e *2*, em séries temporais (2000-2015) de NDVI da bacia do rio Paraíba do Sul, e, conseqüentemente, propor um método para avaliação de testes de filtros. Tais filtros foram escolhidos por estarem implementados em linguagem R, operacionabilidade rápida, poucos parâmetros de entrada exigidos pelos algoritmos, e pelo seu desempenho em outros trabalhos.

2 Área de estudo

A área de estudo é a bacia do rio Paraíba do Sul (Figura 1) e está localizada no sudeste brasileiro nos Estados de São Paulo, Rio de Janeiro e Minas Gerais, com uma área de, aproximadamente, 56500 km². É uma área de estudo pouco pesquisada no campo do sensoriamento remoto e relevante para a realização de experimentos de análise de tendências na cobertura vegetal, por ser importante para a economia brasileira, abastecer a região

metropolitana do RJ e por ter sofrido processos de industrialização e degradação ambiental (OVALLE et al., 2013).

Figura 1 – Localização da área de estudo.



Fonte: Elaborado pelos autores.

3 Material e método

Neste item, são apresentados a área de estudo, os dados de sensoriamento remoto, os métodos utilizados para a execução do presente trabalho, e a descrição das técnicas de filtragem usadas.

3.1. Material

Foram utilizadas imagens NDVI do sensor MODIS a bordo do satélite Terra (produto MOD13Q1, cenas h13v11 e h14v11) entre 2000 e 2015, com resoluções espacial e temporal de 250 m e 16 dias, respectivamente disponível no sítio *Data Pool* do USGS (Serviço Geológico dos Estados Unidos). Este

produto é uma composição temporal de 16 dias gerada por um algoritmo que seleciona o melhor valor de refletância por pixel, com base na menor cobertura de nuvens possível, baixo ângulo de visada e o maior valor do IV (DIDAN, 2015). Também, foram utilizadas imagens Landsat do ano de 2000 (sensor TM) e 2015 (sensor OLI), com resolução espacial de 30 m, para interpretação visual, disponíveis no sítio *Earthexplorer* do USGS.

3.2. Método

Para a realização dos experimentos de filtragem de série temporal de NDVI MODIS, foram selecionados os filtros de média, mediana, *Savitzky-Golay*, *4253H twice*, *Whittaker 1* e *2*. Tal seleção foi feita com base em dois critérios: (i) utilização e revisão bibliográfica em outros trabalhos (SHAO et al., 2016; BORGES et al., 2014; MARIANO, 2015; MICHISHITA et al., 2014; ADAMI, 2010; CHEN et al., 2004); e (ii) rapidez de utilização na linguagem R. O filtro *Hants* e os baseados na transformada de *Fourier* ou de *Wavelets* não foram usados por conta de sua complexidade e exigência de diversos parâmetros de entrada (ZHOU et al., 2016; ZHOU et al., 2015). No caso das técnicas baseadas em *Fourier*, caso os parâmetros não sejam selecionados corretamente, a filtragem pode gerar oscilações espúrias em séries irregulares ou assimétricas de NDVI por necessitar de funções simétricas de seno e cosseno (HIRD; MCDERMID, 2009; CHEN et. al, 2004).

O fluxograma da Figura 2 mostra as etapas adotadas para o trabalho. (i) Primeiramente, a partir da série temporal de imagens de confiabilidade dos pixels (do produto MOD13Q1), foi calculada, para cada pixel, a porcentagem de quantos possuem o valor 0 (Good Data) dentro da série temporal, criando-se uma imagem de qualidade dos pixels.

(ii) Com a imagem de qualidade em conjunto com as imagens TM e OLI, foi feita a interpretação visual para a coleta de quatro pontos referentes às seguintes classes: agricultura, floresta, pastagem e silvicultura. Foram coletados pontos com a maior pureza espectral possível (realizando uma

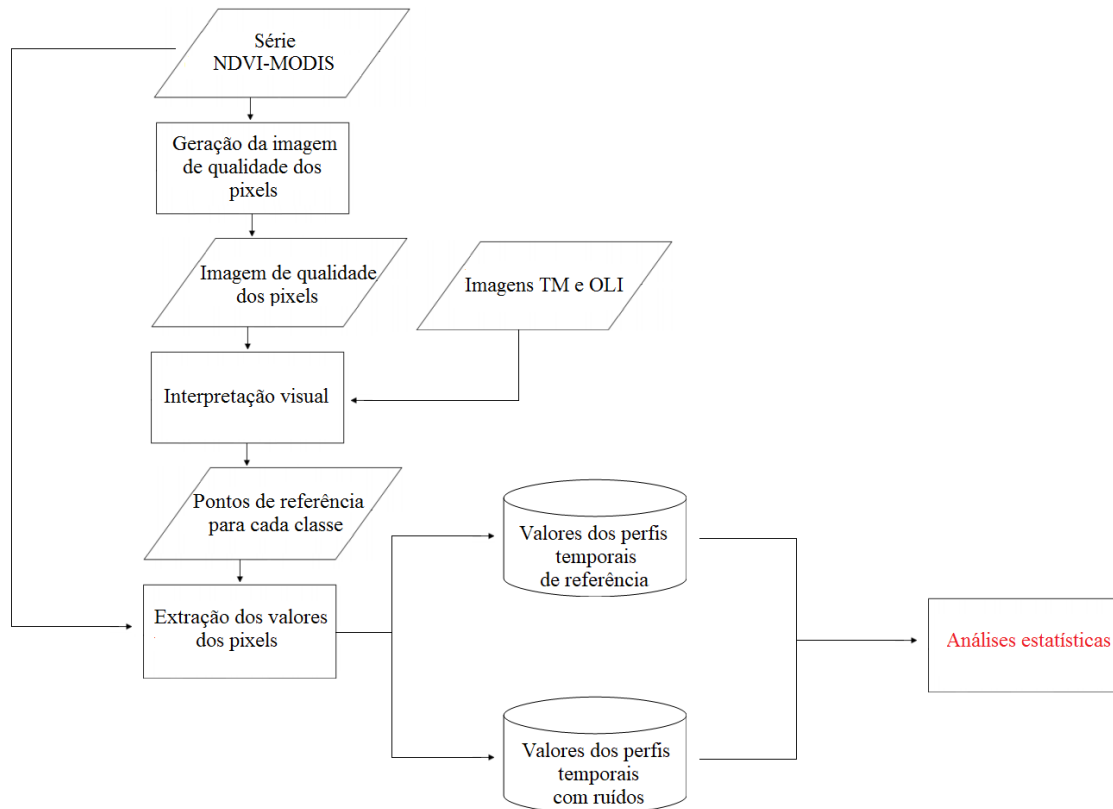
comparação entre Landsat com o tamanho das células MODIS) e com a maior porcentagem de pixels com boa qualidade.

(iii) Para cada ponto de referência (Tabela 1), extraiu-se os valores temporais de NDVI. Valores obtidos em condições com refletância na banda 3 (azul) maior que 0,10 e de ângulo zenital de visada maior que $32,25^\circ$ foram removidos conforme recomendação de Sakamoto et al. (2005) para atenuar ruídos atmosféricos, nuvem, entre outros, e outliers detectados em gráficos do tipo boxplot foram, também, excluídos. Essa pré-filtragem foi adotada por Mariano (2015), Adami et al. (2012), Adami (2010) e Thayn e Price (2008). Os valores removidos e outros pixels que vieram originalmente sem dados foram interpolados linearmente como feito em Chen et al. (2004).

Para cada localização dos pontos de referência, foram extraídos os valores brutos de NDVI, sem a pré-filtragem de Sakamoto et al. (2005), e geradas 1000 séries temporais ruidosas para cada nível de ruído: 10%, 40% e 70%, como feito por Hird e McDermid (2009). Os ruídos foram inseridos a partir de valores aleatórios oriundos de uma distribuição uniforme, para simular efeitos de nuvem e ângulo de visada, semelhante ao que foi feito por Adami (2010).

(iv) Para cada conjunto de 1000 séries temporais com ruídos, foi feita a análise de regressão linear simples com objetivo de obter o RMSE (raiz do erro médio quadrático) entre cada uma das 1000 séries e a série de referência, para, posteriormente, calcular o valor de RMSE médio das séries geradas, e comparar o desempenho dos filtros nas séries de NDVI para cada classe de cobertura da terra e nível de ruído.

Figura 2 – Fluxograma do trabalho.



Fonte: Elaborado pelos autores.

Tabela 1 – Coordenadas geográficas (WGS 84) dos pontos de referência.

Classe	Localização (lat/long)
Agricultura	-45,47 / -22,87
Floresta	-44,18 / -22,88
Pastagem	-44,94 / -22,62
Silvicultura	-45,41 / -22,83

Fonte: Elaborado pelos autores.

3.3. Descrição do funcionamento dos filtros testados

Aqui, são apresentadas as descrições de funcionamento e formulações matemáticas de cada uma das técnicas de suavização.

3.3.1. Filtros de média e mediana móveis

São os métodos mais simples de filtragem e trabalham no domínio do tempo, e com baixo custo computacional (MARIANO, 2015). O filtro de média móvel substitui um determinado valor pela média dos valores vizinhos dentro de uma janela temporal. O filtro de mediana trabalha da mesma forma, porém, com a mediana dos valores dentro de uma janela temporal. O primeiro conserva melhor a forma da série temporal, entretanto, é mais sensível a *outliers*, enquanto que o segundo, não conserva tão bem a forma da série e é menos sensível a valores extremos.

White et al. (2009) utilizam o filtro de média em série temporal AVHRR para o monitoramento de fenologia de primavera na América do Norte, enquanto Abade et al. (2015) utilizam o filtro de mediana, combinado com o filtro *Savitzky-Golay*, em série temporal NDVI MODIS para comparar classificações temporais de vegetação primária e exótica na transição Cerrado/Caatinga no Brasil Central. Os autores afirmam que o filtro de mediana é uma técnica que preserva tendências monotônicas e ruídos de curta duração. Além de ser útil na redução de ruídos impulsivos (CARVALHO JÚNIOR et al., 2012). No presente trabalho, utilizou-se janela temporal de tamanho 7, uma vez que Mariano (2015) indica esse tamanho em um filtro que combina média e mediana.

3.3.2. *Savitzky-Golay*

A suavização foi proposta por Savitzky e Golay (1964) e se trata de uma convolução pelo método dos mínimos quadrados gerando novos valores

através de um filtro ponderado em que há um ajuste - feito por coeficientes obtidos de Steiner et al. (1972) ou calculados pelas equações de Madden (1978) - de um polinômio de um determinado grau (CHEN et al., 2004).

A aplicação dessa técnica envolve 5 passos, como mostram Adami (2010) e Cerqueira et al. (2000): (i) definição do tamanho da janela temporal, (ii) remoção do valor central da janela, (iii) ajuste de um polinômio de um determinado grau a partir dos valores restantes, (iv) estimação do valor removido pelo polinômio, e (v) ir para o próximo valor e repetir o processo. A equação do filtro é dada por:

$$y_j = (\sum_{i=-m}^{i=m} C_i Y_{j+i})/N \quad (1)$$

Em que y_j representa o valor filtrado, c_i os coeficientes do filtro e N o tamanho da janela de suavização de tamanho $2m+1$.

Chen et al. (2004) comparam o filtro *Savitzky-Golay* na reconstrução de uma série temporal de NDVI com outros algoritmos (BISE e *Fourier*) e mostram maior eficiência da primeira técnica em relação às outras analisadas.

Os parâmetros do filtro, neste trabalho, foram selecionados de acordo com a observação dos dados de NDVI, conforme recomendações de Chen et al. (2004) e baseados em Brandt (2014).

3.3.3. 4253H *twice*

Esse filtro calcula uma série de medianas com base em diferentes janelas temporais (tamanhos 4, 2, 5 e 3) na série temporal e posteriormente é aplicada a função de Hanning, onde o ponto central da janela possui maior peso, nos dados e nos resíduos (VELLEMAN, 1980). Deve-se obter os coeficientes da função de *Hanning* pela Equação 2 (ADAMI, 2010; VELEMMAN, 1980):

$$w(n) = 0.5 \left(1 - \cos \left(2\pi \frac{n}{N} \right) \right) \quad (2)$$

Em que $0 \leq n \leq N$, sendo N o tamanho da série, n o ponto calculado, e w o coeficiente obtido para o ponto n .

Após a execução dos filtros de mediana na série temporal, aplica-se a função de *Hanning* (Equação 3) nos dados com uma janela de tamanho 3 (ADAMI, 2010). Os resíduos são calculados a partir dos valores originais e filtrados, e a suavização é realizada novamente nos resíduos. Finalmente, soma-se os resíduos filtrados a série temporal suavizada.

$$z_t = (1/4)y_{t-1} + (1/2)y_t + (1/4)y_{t+1} \quad (3)$$

Hird e McDermid (2009) testam seis filtros e verificam que o 4253H twice apresentou o terceiro melhor desempenho global (superior ao Savitzky-Golay), chegou a ser o primeiro colocado em duas das seis classes de paisagem analisadas no artigo, considerando uma análise estratificada por unidade de paisagem, e o melhor desempenho em série temporal com nível de ruído de 10%. Os autores concluem que tal filtro apresentou bom desempenho.

Adami (2010) testa seis filtros em séries EVI/MODIS (incluindo o filtro Hants e o baseado em Wavelets) e apresentou o menor RMSE para um perfil temporal com 10% de ruído, e considera, que de maneira geral, que o 4253H twice apresentou, junto com outros quatro filtros (Wavelets, Hants, Savitzky-Golay e duplo logístico), resultados superiores.

3.3.4. *Whittaker* 1 e 2

É um suavizador baseado no equilíbrio entre a fidelidade aos valores da série temporal (expressa pela diferença quadrática entre valores estimados e verdadeiros) e rugosidade (expressa como a diferença quadrática entre estimativas sucessivas); e esse equilíbrio é controlado por um parâmetro de rugosidade calculado (KANDASAMY et al., 2013). O filtro de *Whittaker*

possui boa adaptação às bordas, não assume periodicidade, é rápido computacionalmente e fornece controle sobre a suavização (ATZBERGER; EILERS, 2011).

De acordo com Atzberger e Eilers (2011), o objetivo da filtragem é ajustar uma série suavizada z aos dados originais y com ruído. Dessa forma, é estabelecido o equilíbrio entre dois fatores contrastantes, a fidelidade aos dados originais (S) e o parâmetro rugosidade de z dado pelo número k , como mostra a Equação 4. Sendo assim, o algoritmo define a série suavizada z que minimiza Q .

$$Q = S + kR \quad (4)$$

Onde,

$$S = \sum_i (y_i - z_i)^2 \quad (5)$$

$$R = \sum_i ((z_i - z_{i-1}) - (z_{i-1} - z_{i-2}))^2 \quad (6)$$

Kandasamy et al. (2013) testam oito filtros e demonstram que o filtro de Whittaker é eficiente para o preenchimento de falhas em série temporal MODIS de índice de área foliar e baixo valor de RMSE para a reconstrução de tal série. Atzberger e Eilers (2011) indicam que o filtro de *Whittaker* permite um rápido processamento de um grande volume de dados com apenas um único parâmetro (rugosidade), reduz o efeito pepita ao aumentar o sinal médio e melhora a separabilidade de classes de cobertura da terra com base na assinatura temporal.

Neste trabalho, foram utilizados dois tipos de suavização *Whittaker*, um com a diferença finita de penalidade com ordem 1, e o outro com ordem 2. A execução dos filtros foi feita com o valor do parâmetro de rugosidade utilizado por Marcos et al. (2015) (igual a 1) para preservar ao máximo as características da série temporal, ao mesmo tempo que reduz o ruído.

4 Resultados

Aqui são apresentados os perfis de referência para as classes de cobertura da terra selecionadas, e a análise dos erros dos filtros.

4.1. Apreciação visual dos perfis de referência

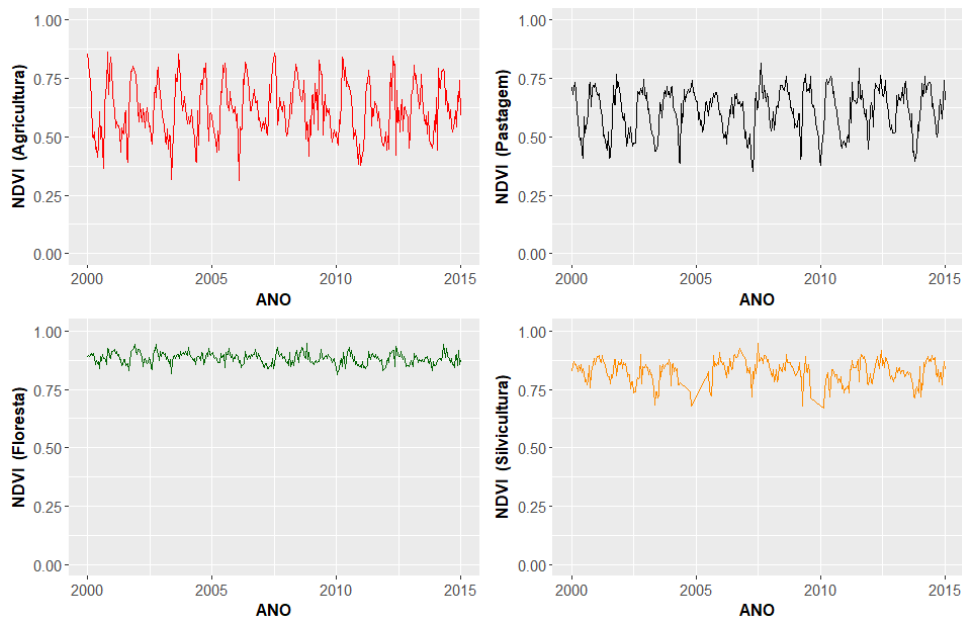
Os perfis de referência para as classes de cobertura na bacia do rio Paraíba do Sul estão na Figura 3. A forma dos gráficos possui, visualmente, comportamento temporal e formas semelhante aos perfis de referência apresentados em Borges et al. (2014), Michishita et al. (2014) e Hird e McDermid (2009). Cada trabalho possui classes de cobertura específicas para cada tipo de clima de diferentes áreas de estudo. Porém, a forma ruidosa dos gráficos de referência é semelhante, o que mostra que o processo de interpretação visual foi feito corretamente. Percebe-se que a variação do perfil temporal de floresta encontra-se com NDVI de 0,81 até 0,95, por conta da variação do vigor vegetativo conforme as estações do ano (sazonalidade entre estação úmida e seca). O perfil de agricultura possui a maior variação de NDVI, entre 0,31 até 0,87, por causa das épocas do início do plantio, passando pelo máximo desenvolvimento, até a colheita.

A pastagem apresenta a segunda maior variação de NDVI, com picos menores que a agricultura (em torno de 0,7-0,8) durante o verão, quando a vegetação rasteira está mais úmida e vigorosa, e valores menores durante a época mais seca, chegando a 0,35 - comportamento semelhante ao apresentado pelo gráfico da série temporal de gramínea no artigo de Hird e McDermid (2009).

Finalmente, a forma da série temporal da silvicultura apresentou padrão esperado, com menor teor de biomassa vegetal em relação à floresta (como mostra a Figura 3), e variação (0,67 e 0,95) maior que esta, e menor que o restante das classes. Por conta dos valores removidos na etapa de pré-

processamento, devido à qualidade dos pixels, uma maior quantidade de observações foi interpolada, o que se reflete na forma do próprio gráfico.

Figura 3 – Localização da área de estudo.



Fonte: Elaborado pelos autores.

4.2 Simulação com 10% de ruído

As Figuras 4,5,6 e 7 mostram, para cada classe de cobertura, o exemplo de uma das 1000 ruidosas geradas e os resultados após a filtragem. A Figura 8 apresenta os gráficos boxplot dos erros obtidos, para cada filtro por classe de cobertura da terra, entre as 1000 curvas ruidosas e a série temporal de referência (que consta na Figura 3). A Tabela 2 apresenta a média dos erros por classe e a média geral por filtro. Percebe-se, que de forma geral, os erros maiores foram para as classes de agricultura e pastagem, por conta da maior variação de NDVI ao longo das estações do ano, e os menores erros foram encontrados na silvicultura e floresta - esta última com os menores erros - por conta de sua menor variação de NDVI. Apesar das especificidades do clima e da vegetação, as métricas utilizadas por Borges et al. (2014) mostram melhor desempenho da filtragem para as classes de floresta, e um

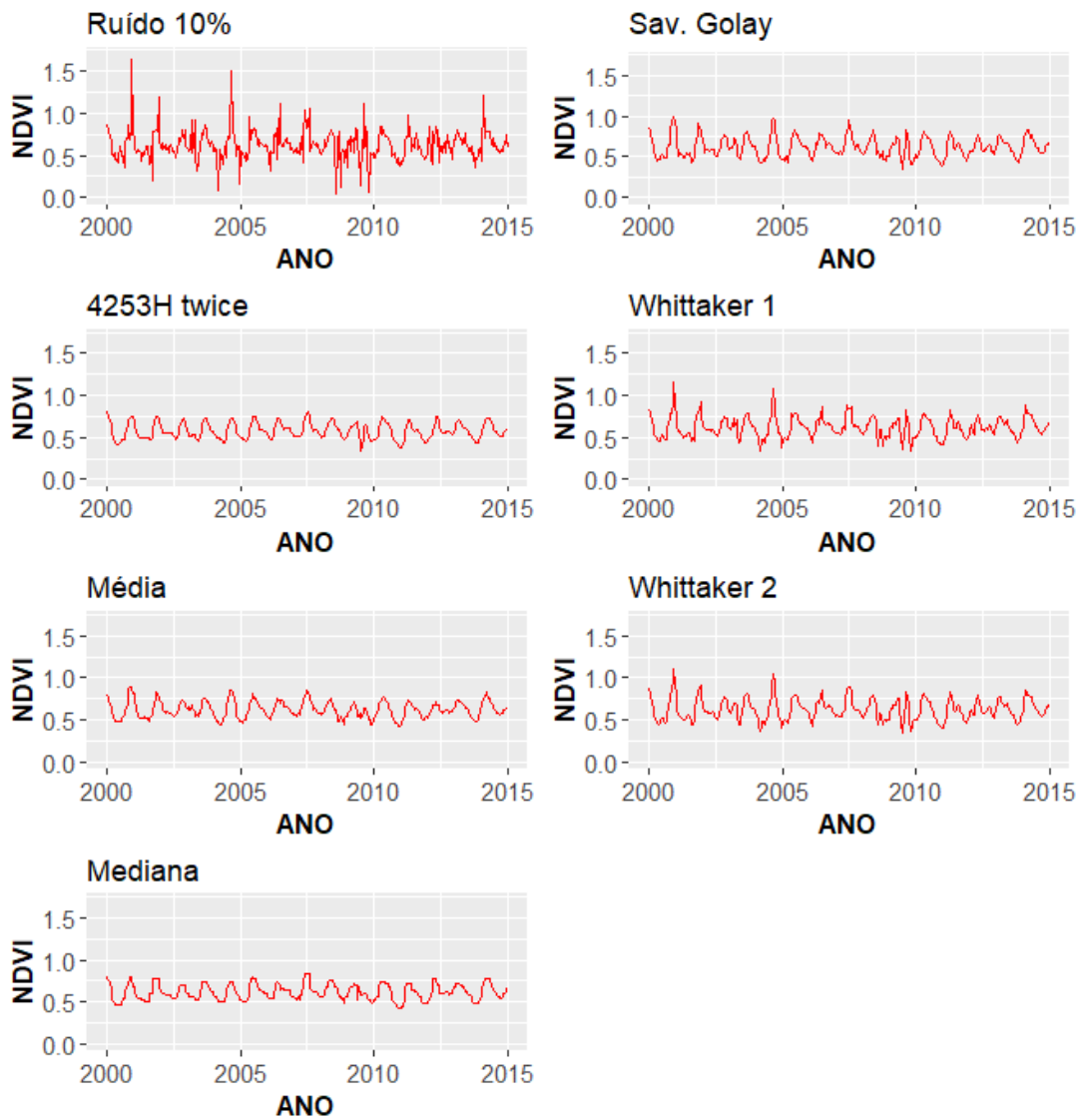
desempenho inferior para as classes de menor biomassa e/ou maior variação fenológica, como cerrado, pasto e agricultura. Além disso, com 10% de ruído, constata-se que os valores do RMSE (eixo y dos gráficos da Figura 8 e Tabela 2) são pequenos e próximos entre si, apesar dos menores erros dos filtros de mediana e 4253H para todas as classes, e os maiores erros do filtro Savitzky-Golay, em que na classe agricultura apresentou média de RMSE de 0,075. Visualmente (Figuras 4,5,6 e 7), percebe-se que o filtro de média realiza uma suavização que conserva a forma da série temporal, porém é sensível aos valores extremos (MARIANO, 2015). O filtro de mediana é menos sensível aos ruídos, porém, não conserva bem a forma da série temporal, como Mariano (2015) afirma. Entretanto, ele é utilizado com eficiência em dados de sensoriamento remoto, como em Abade et al. (2015) e Kovalsky et al. (2011).

Tabela 2 - Médias dos erros (*RMSE*) das 1000 curvas por classe e por filtro (10% de ruído).

Classe	Filtro Média	Filtro Mediana	<i>Savitzky-Golay</i>	<i>4253H twice</i>	<i>Whittaker 1</i>	<i>Whittaker 2</i>
Agricultura	0,075	0,062	0,075	0,0629	0,0691	0,0697
Floresta	0,024	0,018	0,024	0,022	0,02398	0,02398
Pastagem	0,0602	0,0437	0,0629	0,045	0,0596	0,0594
Silvicultura	0,047	0,027	0,050	0,034	0,0489	0,049
Média total para todas as classes	0,052	0,038	0,053	0,041	0,050	0,050

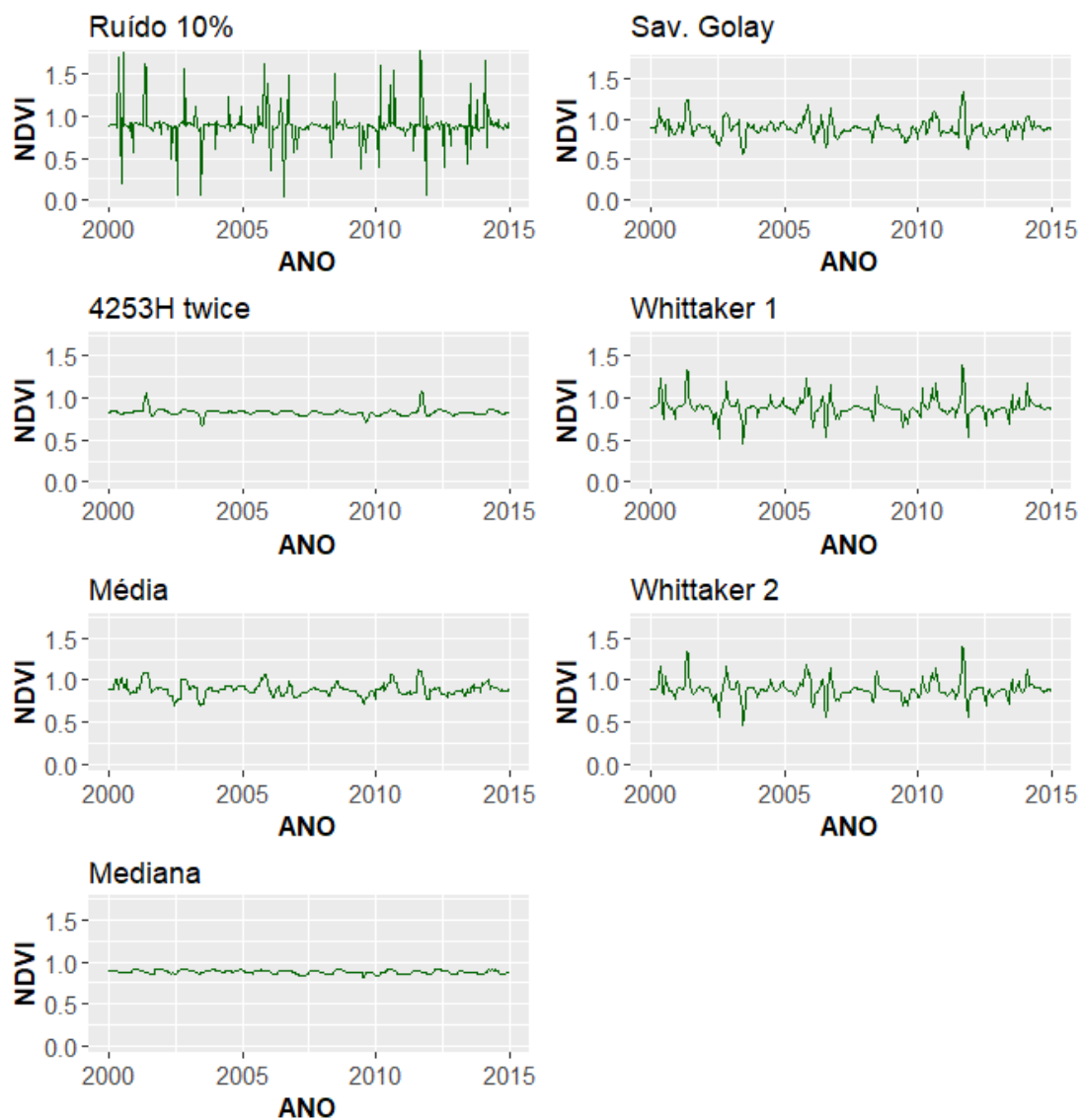
Fonte: Elaborado pelos autores.

Figura 4 – Simulação com 10% de ruído para agricultura.



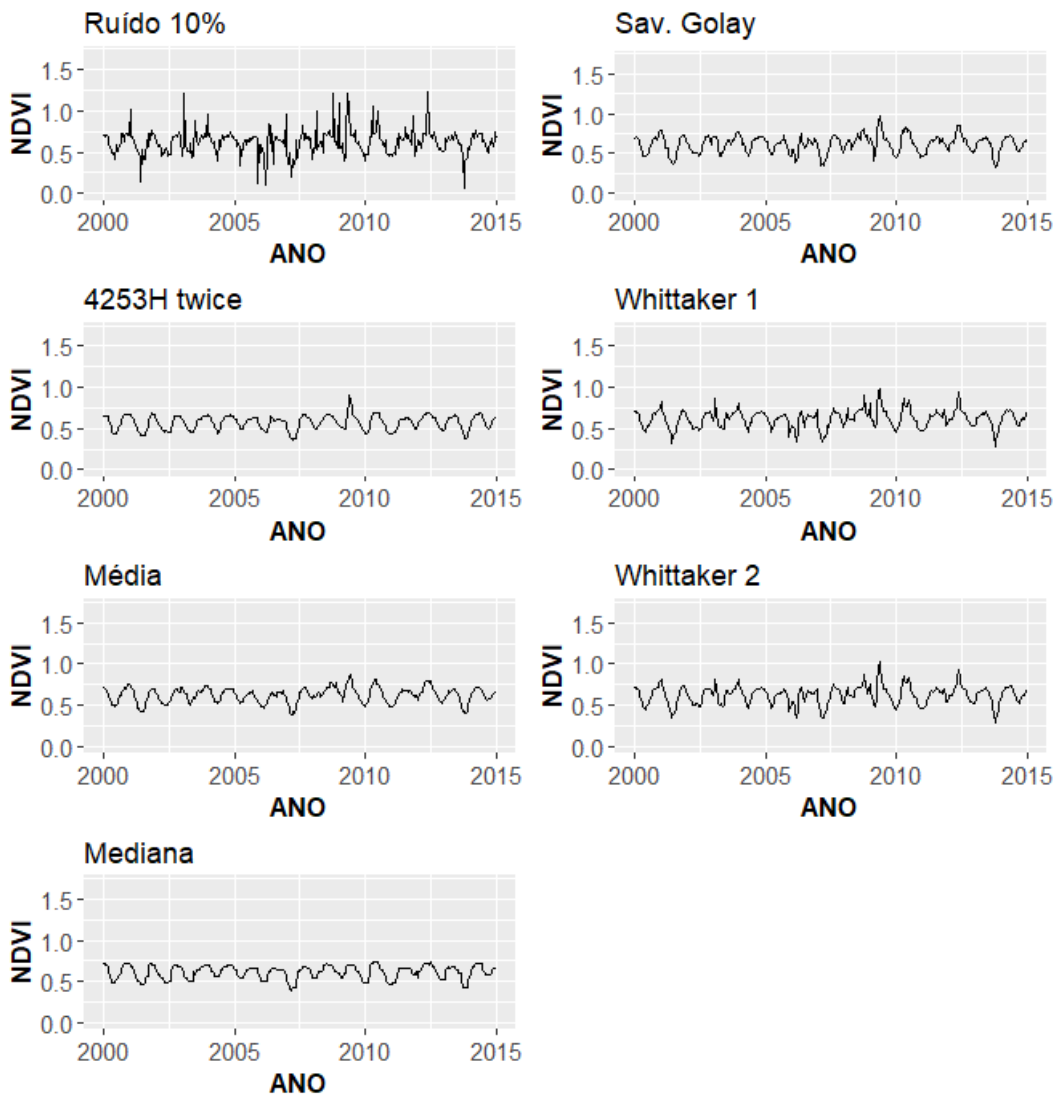
Fonte: Elaborado pelos autores.

Figura 5 - Simulação com 10% de ruído para floresta.



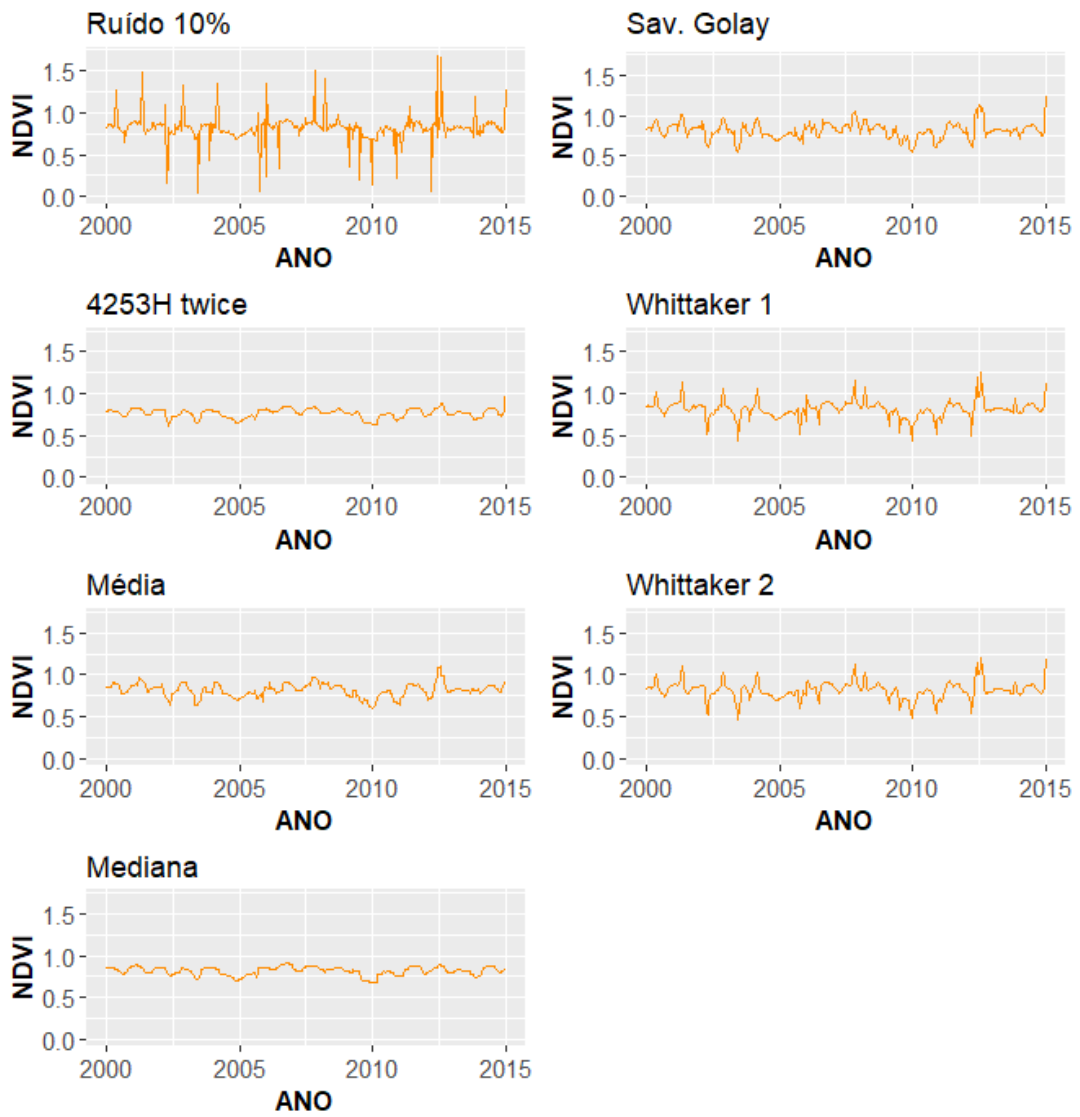
Fonte: Elaborado pelos autores.

Figura 6 - Simulação com 10% de ruído para pastagem.



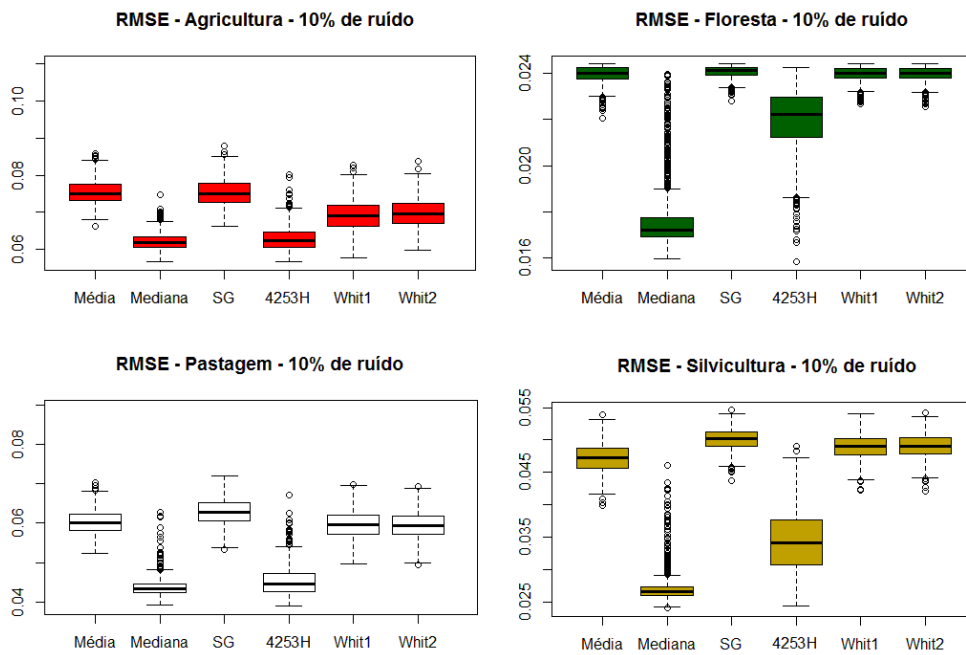
Fonte: Elaborado pelos autores.

Figura 7 - Simulação com 10% de ruído para silvicultura.



Fonte: Elaborado pelos autores.

Figura 8 - Gráficos *boxplot* dos erros para 10% de ruído.



Fonte: Elaborado pelos autores.

O filtro *Savitzky-Golay* apresentou o maior erro para todas as classes de cobertura, inclusive para classe floresta, com média de RMSE de 0,053 (Tabela 2), e para a silvicultura, apresentou o maior valor (0,050). Visualmente, os gráficos mostraram uma suavização com sensibilidade aos *outliers*, porém, seus erros podem ser considerados pequenos (como visto, média total de 0,053). Como em Hird e McDermid (2009), o filtro *Savitzky-Golay* apresentou a menor performance global (métrica utilizada pelo autor), para a simulação com 10% de ruído, em comparação aos outros filtros, já que não foi tão eficiente em manter a integridade da série, apesar de um RMSE baixo.

O resultado do presente trabalho em relação ao RMSE do filtro *Savitzky-Golay* não contradiz a efetividade da aplicação de Chen et al. (2004), até por causa dos valores baixos do erro, próximos aos RMSE encontrados por Michishita et al. (2014) - variando de 0,03 a 0,06 -, que considera o resultado para este filtro efetivo (porém para classes diferentes das do presente trabalho). Também, os valores de RMSE de agricultura e pastagem estão próximos aos mostrados por Adami (2010) - simulação com 10% de ruído - em

torno de 0,073 (nos outros casos, as médias de RMSE aqui apresentados foram menores).

Apesar da média do erro total maior do filtro 4253H *twice* em relação ao filtro de mediana, seu valor também é baixo, e foi o segundo melhor erro (0,041). A forma da série temporal filtrada por este suavizador foi, visualmente, a mais adequada na redução dos ruídos positivos e negativos, deixando o gráfico com uma adequada representação visual das variações e/ou fenologia das classes de cobertura. Diferente do filtro de mediana, que alterou a forma da série temporal não permitindo uma compreensão eficaz das variações temporais. Os valores dos erros estão próximos aos que Adami (2010) apresentou com uma simulação com 10% para o filtro 4253H *twice* (este autor obteve valores inferiores e superiores, obtidos pelo método *bootstrap*, de 0,0540 e 0,0541 respectivamente. Entretanto, o filtro 4253H *twice* no artigo de Hird e McDermid (2009) não foi considerado muito consistente na suavização, uma vez que em seus dados, esta técnica não preservou os valores máximos de NDVI, apesar da boa performance em certas classes de cobertura e a melhor performance com 10% de ruído. Esse fato mostra que com a variação da cobertura da terra e do tipo de dado, resultados contraditórios podem ser encontrados.

Porém, visualmente, os gráficos suavizados por Hird e McDermid (2009) estão adequados para a representação temporal. Os filtros de *Whittaker* apresentaram RMSE pouco menores que o filtro *Savitzky-Golay*, e apenas para agricultura e pastagem apresentaram, também, RMSE menores ao filtro de média. Entretanto, seus valores de erros podem ser considerados pequenos (como no caso do filtro *Savitzky-Golay*). Visualmente, pode-se considerar a suavização adequada (Figuras 4,5,6 e 7), em especial o filtro *Whittaker 2*, que por conta de uma função de maior ordem, permite uma suavização mais eficiente que o *Whittaker 1*.

E estes resultados são corroborados por Atzberger e Eilers (2011) que afirmam que o suavizador de *Whittaker* é útil no pré-processamento de séries temporais de sensoriamento remoto para melhorar sua qualidade, assim como

os de Shao et al. (2016), que mostram que o filtro de *Whittaker* melhora a separabilidade de classes de cobertura e mantém dados corretos de NDVI, sem uma suavização excessiva que remove informações temporais.

4.3. Simulação com 40% de ruído

As Figuras 9,10,11 e 12 mostram, para cada classe de cobertura, o exemplo de uma das 1000 curvas ruidosas geradas e os resultados após a filtragem. A Figura 13 apresenta os gráficos *boxplot* dos erros obtidos, para cada filtro por classe de cobertura, entre as 1000 curvas ruidosas e a série temporal de referência. A Tabela 3 apresenta a média dos erros por classe e a média geral por filtro.

Tabela 3 - Médias dos erros (*RMSE*) das 1000 curvas por classe e por filtro (40% de ruído).

Classe	Filtro Média	Filtro Mediana	<i>Savitzky-Golay</i>	<i>4253H twice</i>	<i>Whittaker 1</i>	<i>Whittaker 2</i>
Agricultura	0,0932	0,084	0,096	0,087	0,093	0,0927
Floresta	0,02421	0,0240	0,0241	0,0242	0,02424	0,02423
Pastagem	0,076	0,067	0,081	0,072	0,079	0,0788
Silvicultura	0,052	0,048	0,054	0,052	0,0535	0,0535
Média total para todas as classes	0,0618	0,0587	0,0639	0,059	0,0626	0,0623

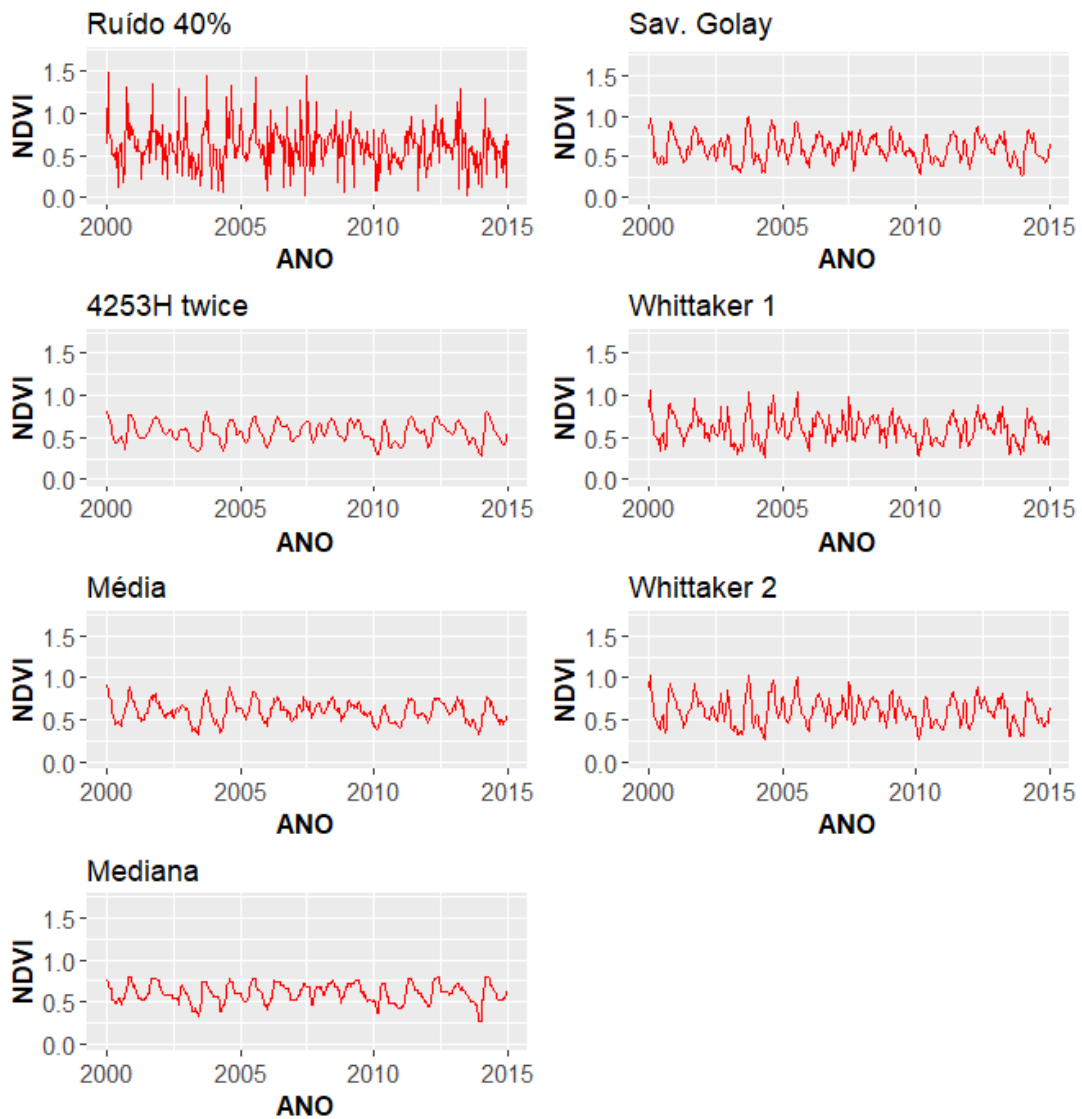
Fonte: Elaborado pelos autores.

Os gráficos *boxplot*, e conseqüentemente as médias de erros (por classe e geral) apresentaram padrão semelhante e valores próximos. Erros maiores para as classes com maior variação temporal dentro do ano (agricultura e pastagem) e menores para as classes com comportamento mais constante (floresta e silvicultura). Mais uma vez o filtro de mediana apresentou a menor média geral de *RMSE* (0,0587), e o *4253H twice*, a segunda menor (0,059). Já, o resultado com maior *RMSE*, foi, novamente, do filtro *Savitzky-Golay* (em especial, agricultura com média de *RMSE* de 0,096) com média total de 0,0639.

O padrão visual dos gráficos (Figuras 9,10,11 e 12) permaneceram os mesmos, com um melhor comportamento do filtro *4253H twice* para a manutenção das características temporais e/ou fenológicas. Como a discussão mais detalhada foi feita no item anterior, não será repetida nesse item. Vale ressaltar que diferente dos valores de RMSE (pelo método *bootstrap*) encontrados por Adami (2010) em uma simulação com 40% de ruído foram um pouco maiores para o filtro *Savitzky-Golay* (em torno de 0,09) e razoavelmente maiores que o filtro *4253H twice* (por volta de 0,16). Entretanto, uma comparação direta não pode ser feita por conta de objetivos diferentes entre os trabalhos, e também, diferenças vegetacionais das classes e climáticas, mas permite dizer que os resultados do presente trabalho estão adequados.

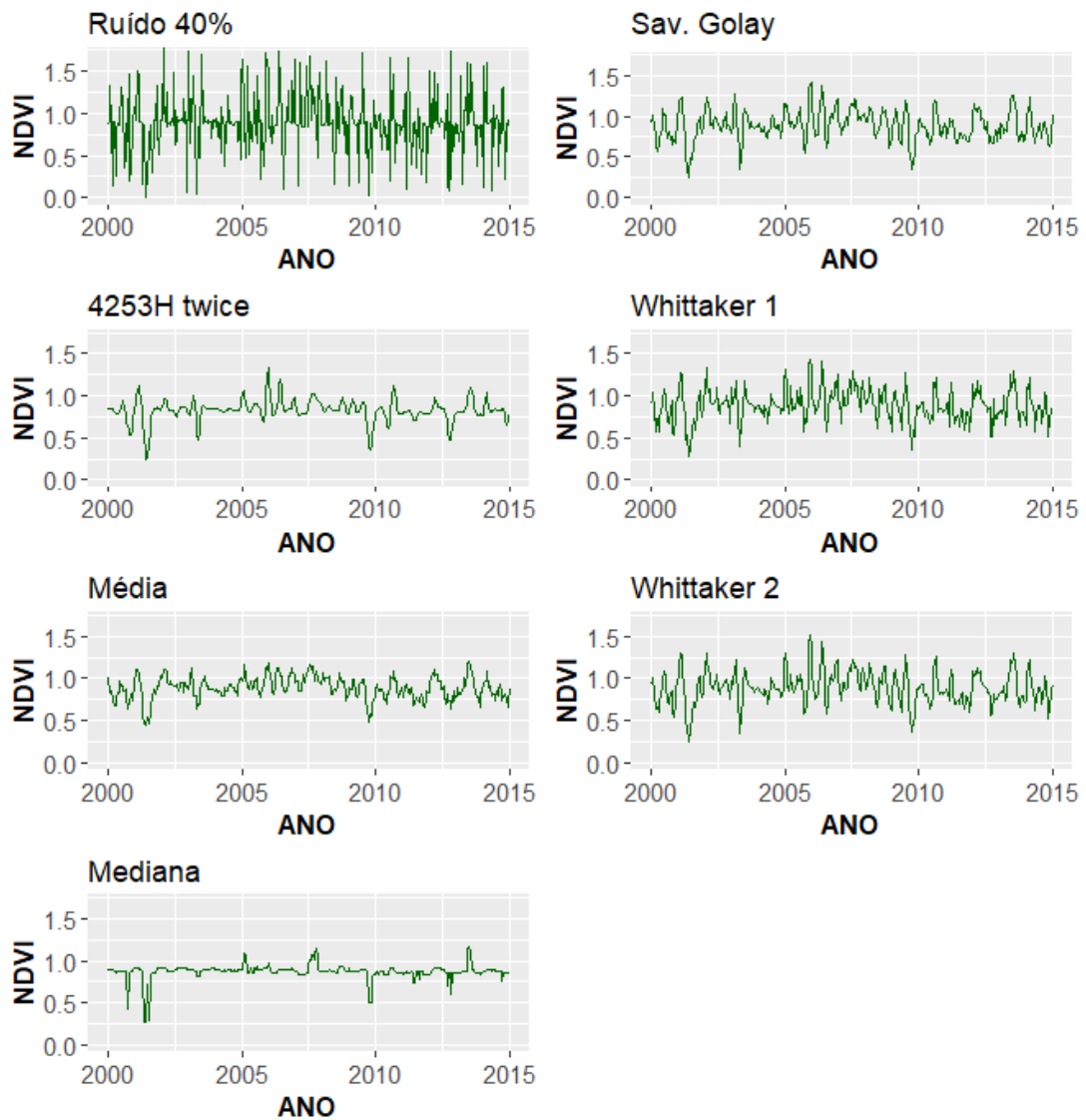
Também, nota-se que o filtro *Savitzky-Golay* apresentou desempenho total geral pouco inferior que o filtro *4253H twice* no trabalho de Hird e McDermid (2009), e este último suavizador apresentou o segundo melhor resultado (empatado com o filtro Gaussiano assimétrico), em relação às outras técnicas, para uma simulação com 40% de ruído.

Figura 9 - Simulação com 40% de ruído para agricultura.



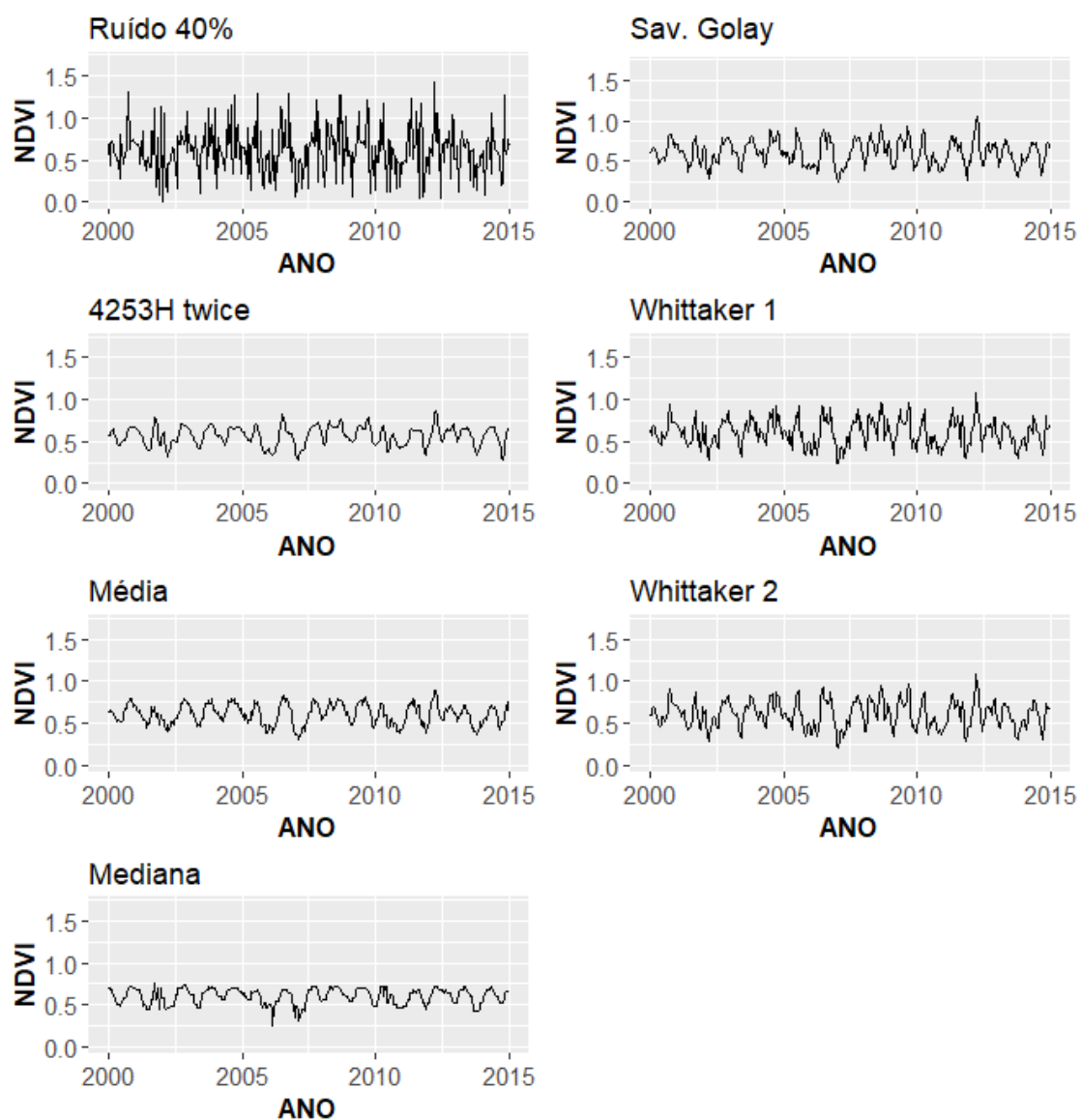
Fonte: Elaborado pelos autores.

Figura 10 - Simulação com 40% de ruído para floresta.



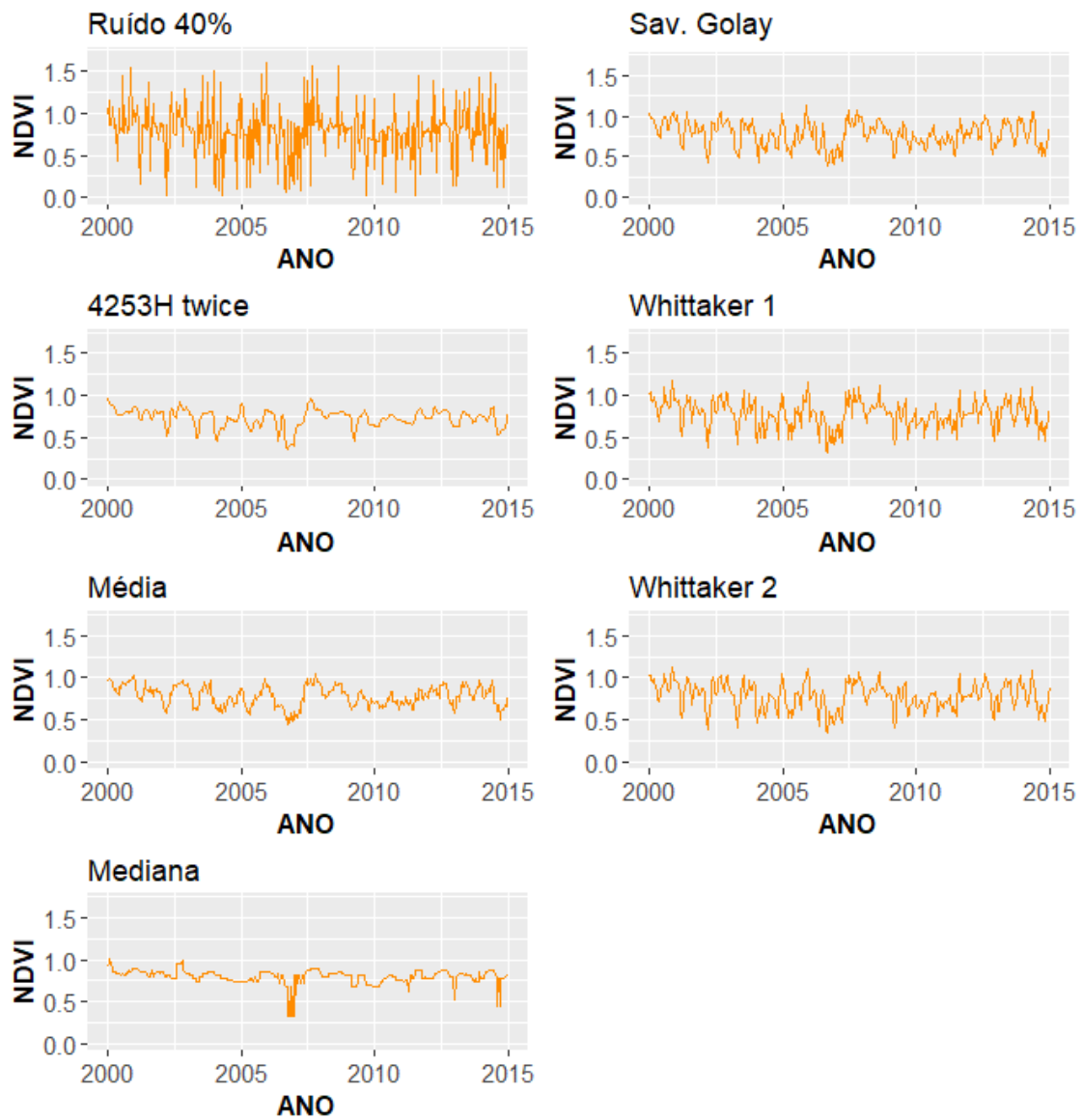
Fonte: Elaborado pelos autores.

Figura 11 - Simulação com 40% de ruído para pastagem.



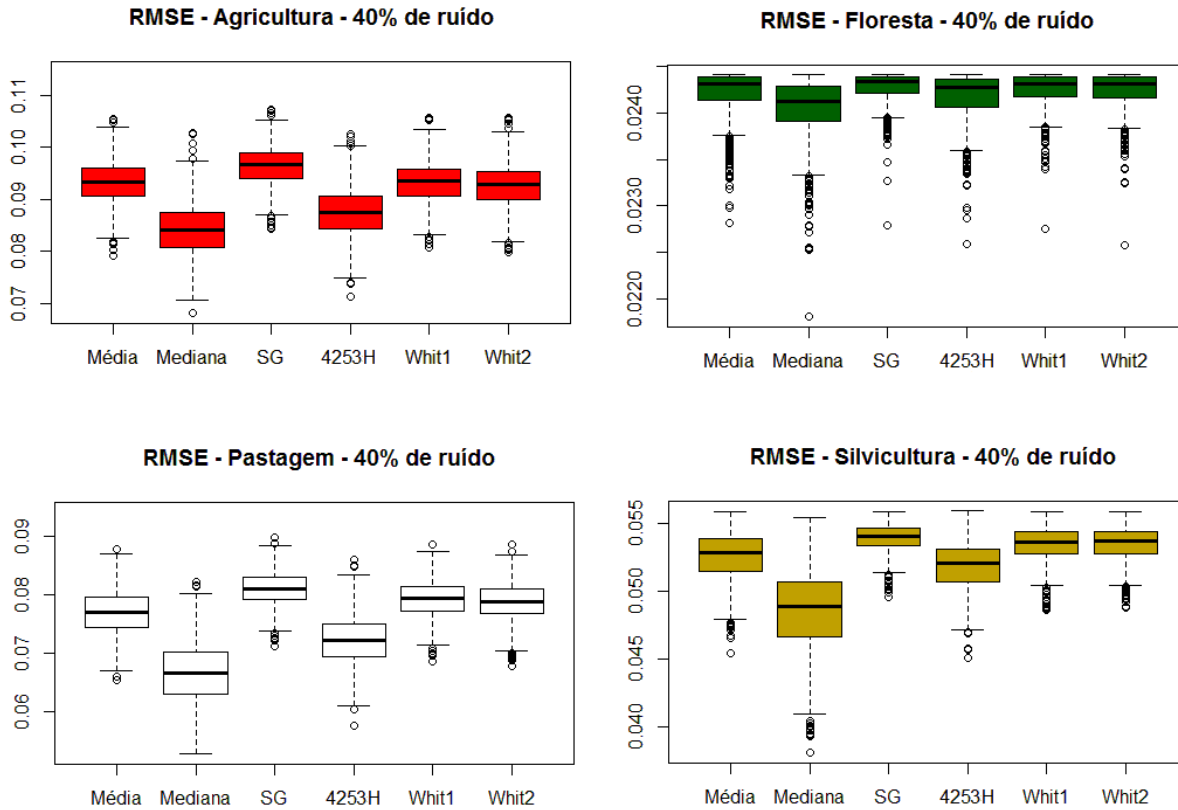
Fonte: Elaborado pelos autores.

Figura 12 - Simulação com 40% de ruído para silvicultura.



Fonte: Elaborado pelos autores.

Figura 13 - Gráficos *boxplot* dos erros para 40% de ruído.



Fonte: Elaborado pelos autores.

4.4. Simulação com 70% de ruído

As Figuras 13,14,15 e 16 mostram, para cada classe de cobertura, o exemplo de uma das 1000 curvas ruidosas geradas e os resultados após a filtragem. A Figura 17 apresenta os gráficos *boxplot* dos erros obtidos, para cada filtro por classe de cobertura, entre as 1000 curvas ruidosas e a série temporal de referência (que consta na Figura 3) e a Tabela 4, a média dos erros por classe e a média geral por filtro.

Novamente, pela análise visual das séries temporais das classes agricultura e pastagem, percebe-se que o filtro *4253H twice* permitiu uma suavização capaz de conservar detalhes da variação temporal, o filtro de mediana alterou a forma do gráfico (apesar de um RMSE baixo). O filtro *Savitzky-Golay* causou uma suavização que manteve os ruídos da série de NDVI, o que pode dificultar a interpretação das variações temporais.

Entretanto, para os perfis das classes que possuem séries temporais com menor variação (floresta e silvicultura), não foi possível o entendimento correto do comportamento temporal (para esse nível de ruído), apesar de gráficos com formatos mais suaves gerados pelo filtro *4253H twice*.

Pelos números da Tabela 4 e gráficos *boxplot*, nota-se os menores valores de erros para o filtro de média em agricultura, floresta e pastagem e os maiores para os erros do filtro *Savitzky-Golay*. De forma geral, as médias totais de RMSE possuem valores próximos, com o menor valor para média (0,0649), o segundo menor para o *4253H twice* (0,0657), e o maior para o *Savitzky-Golay* (0,0668).

Nota-se que para 70% de ruído, o filtro *4253H twice* teve melhor desempenho que o de mediana (que nas outras simulações foi melhor), e na pastagem, estes filtros mostraram valores iguais. Fazendo uma comparação com os resultados de Hird e McDermid (2009) (com ressalvas por conta de áreas de estudo diferentes e filtros utilizados diferentes), o filtro *4253H twice* para uma simulação com 70% de ruído teve performance inferior a outros 3 filtros (de um total de 6), inclusive do filtro *Savitzky-Golay*. Adami (2010) também encontrou erros maiores do filtro *4253H twice* em relação ao *Savitzky-Golay*.

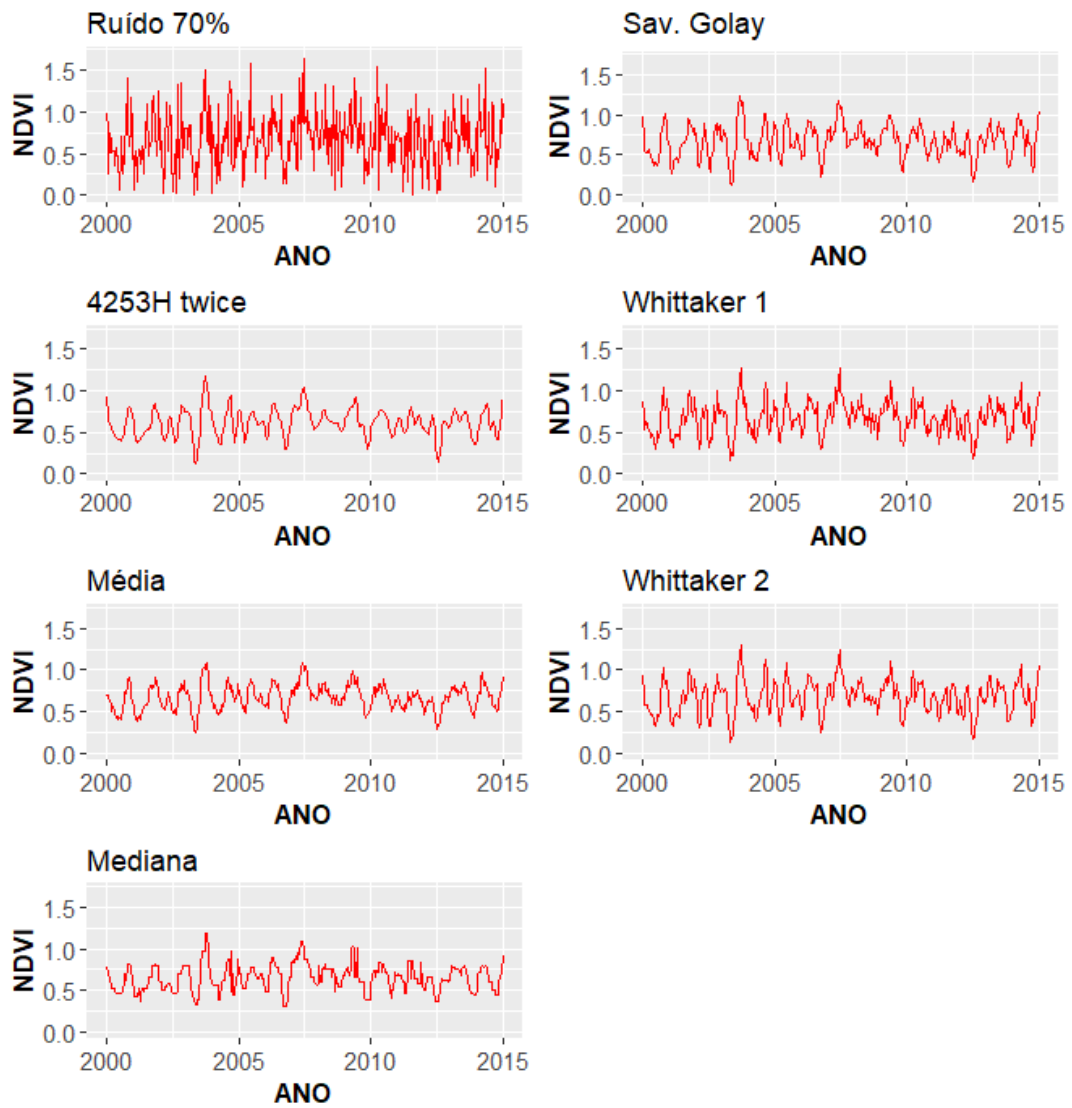
Diferente do presente trabalho, que, apesar dos valores de RMSE próximos, o filtro *4253H twice* apresentou média total (0,0657) menor que da técnica de *Savitzky-Golay* (0,0668). Assim como Adami (2010), no presente trabalho, considera-se o filtro *4253H twice* eficiente na filtragem dos dados MODIS.

Tabela 4 - Médias dos erros (*RMSE*) das 1000 curvas por classe e por filtro (70% de ruído).

Classe	Filtro Média	Filtro Mediana	<i>Savitzky-Golay</i>	<i>4253H twice</i>	<i>Whittaker</i> 1	<i>Whittaker</i> 2
Agricultura	0,0996	0,1016	0,1025	0,100	0,1005	0,09976
Floresta	0,02423	0,02426	0,02429	0,02452	0,02427	0,02427
Pastagem	0,0822	0,0836	0,0857	0,0836	0,0844	0,0840
Silvicultura	0,0537	0,0541	0,0547	0,05426	0,05439	0,05439
Média total para todas as classes	0,0649	0,0659	0,0668	0,0657	0,0659	0,0656

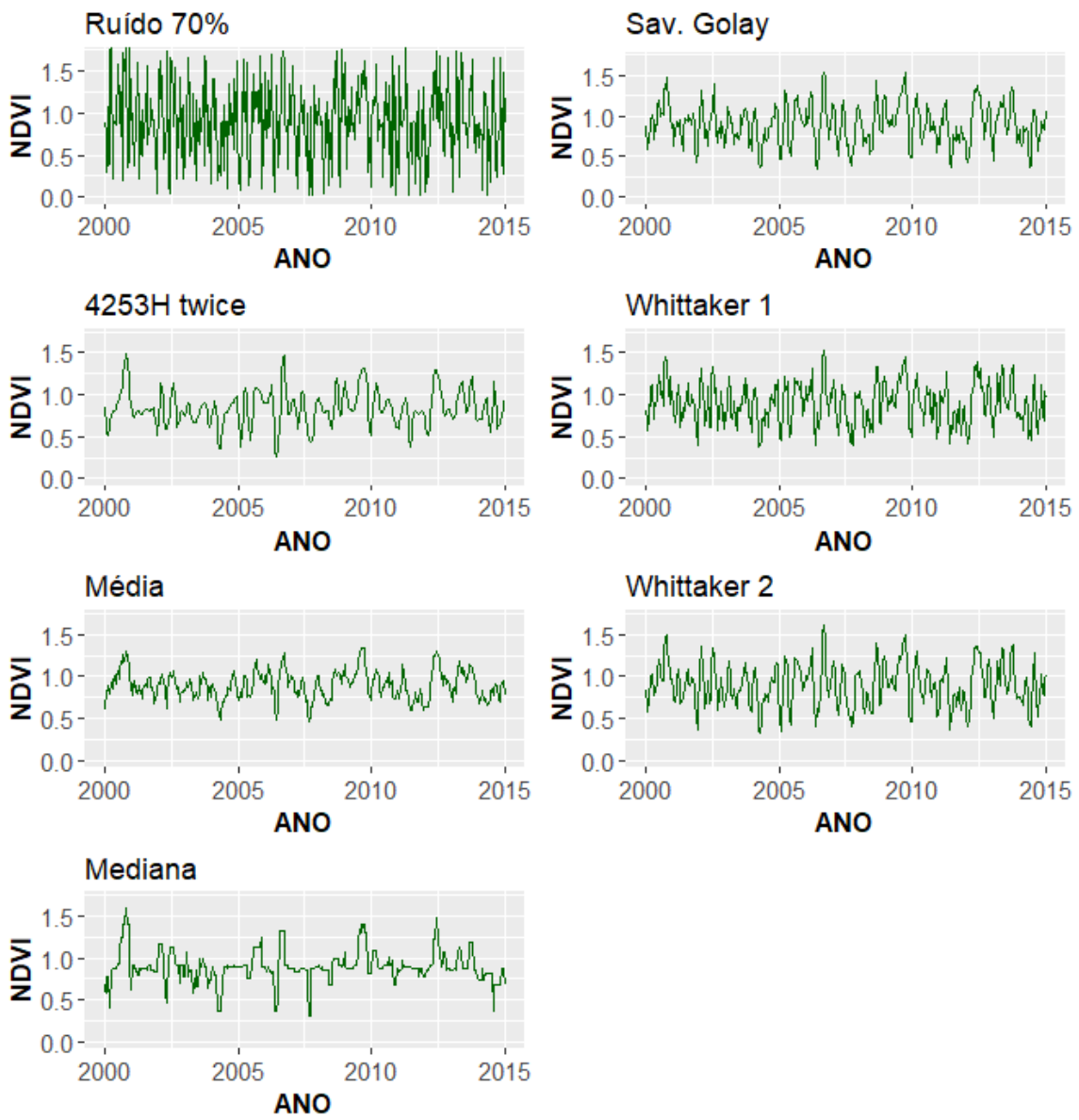
Fonte: Elaborado pelos autores.

Figura 14- Simulação com 70% de ruído para agricultura.



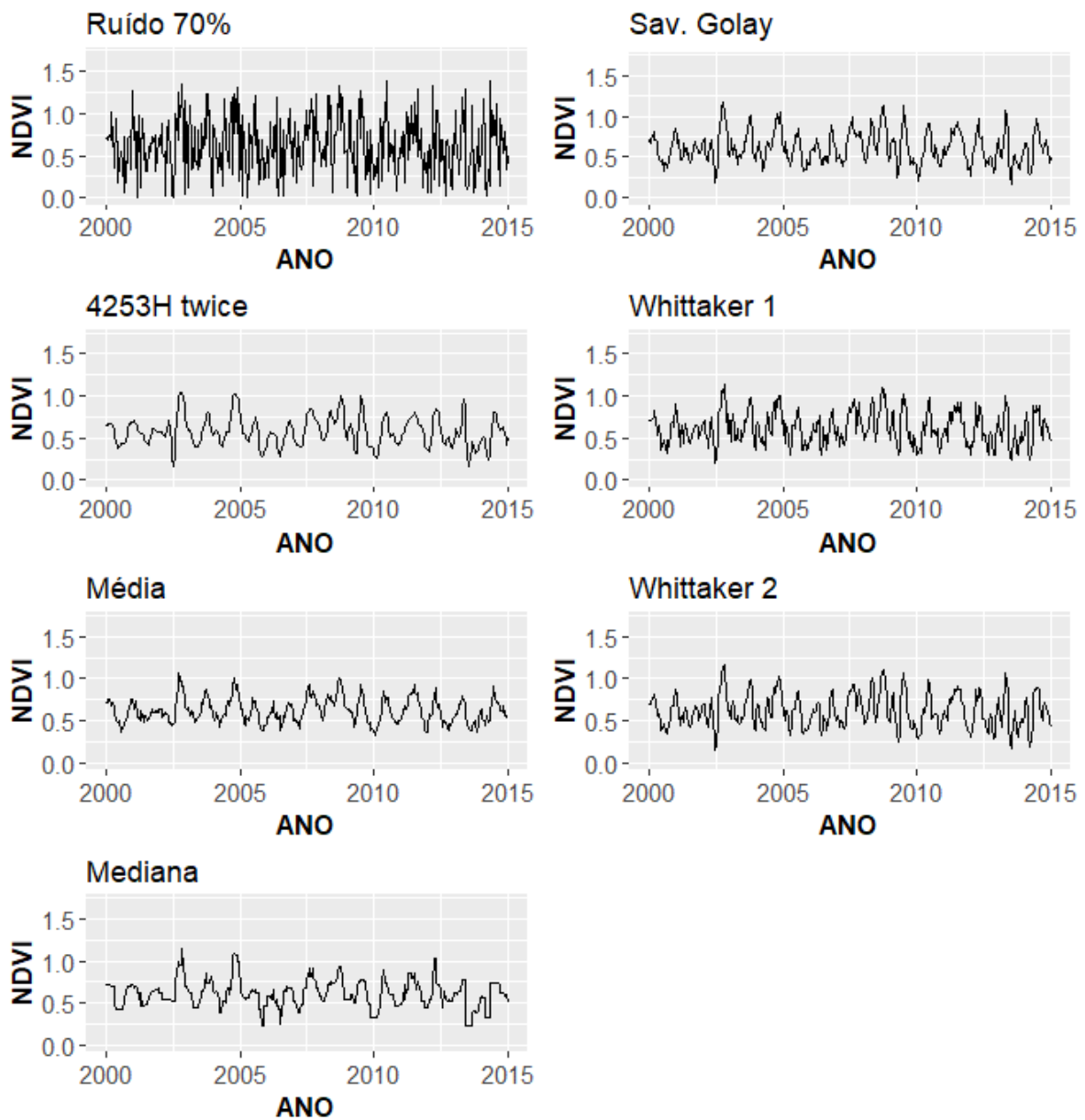
Fonte: Elaborado pelos autores.

Figura 15- Simulação com 70% de ruído para floresta.



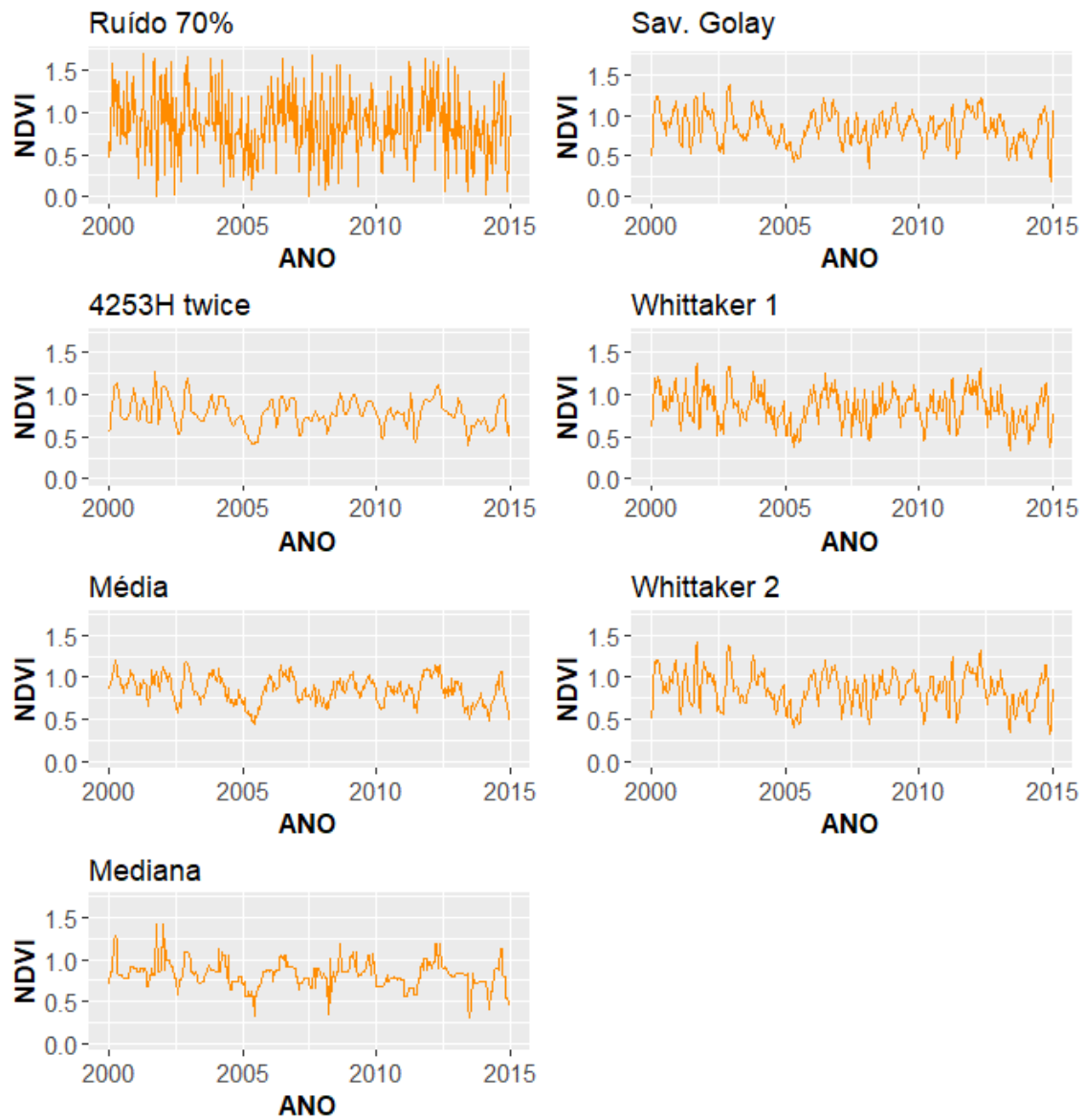
Fonte: Elaborado pelos autores.

Figura 16 - Simulação com 70% de ruído para pastagem.



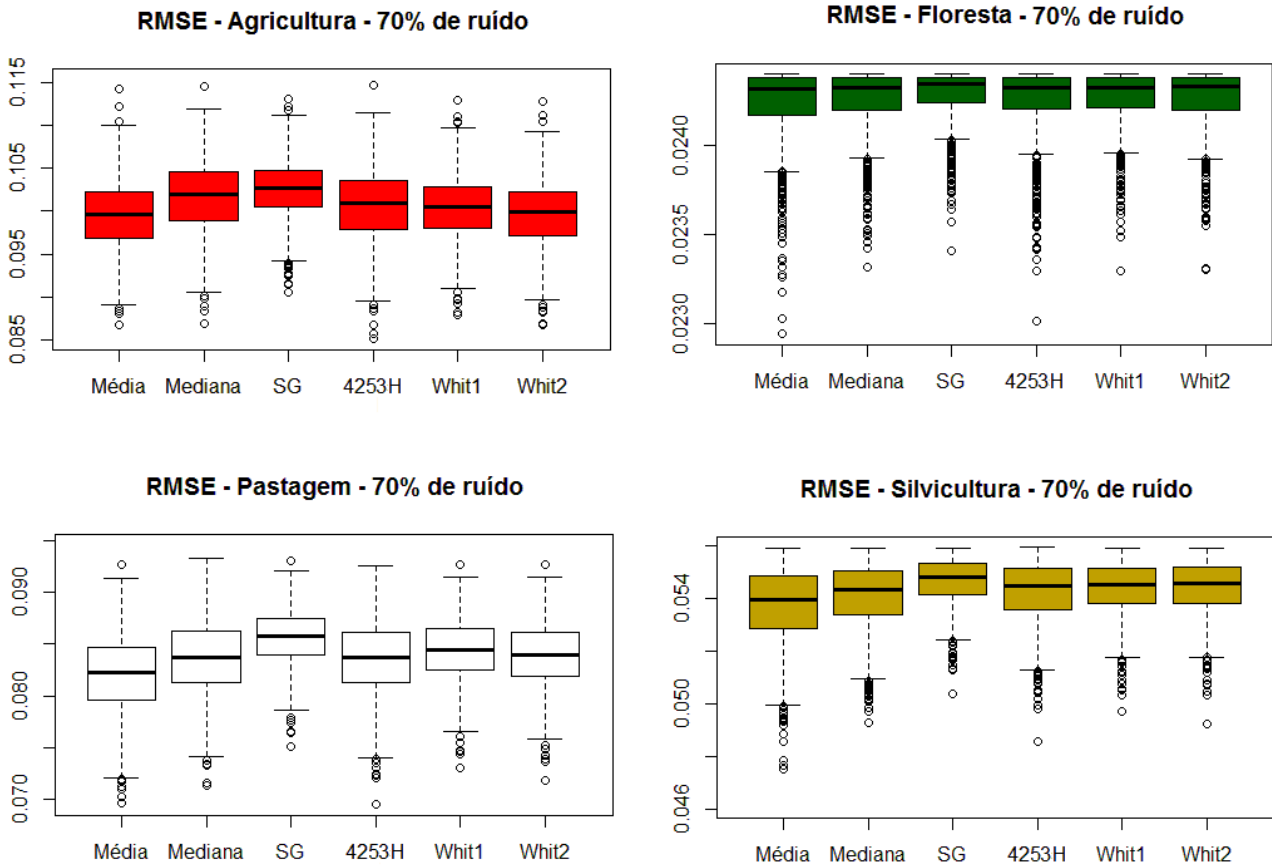
Fonte: Elaborado pelos autores.

Figura 17 - Simulação com 70% de ruído para silvicultura.



Fonte: Elaborado pelos autores.

Figura 18 - Gráficos *boxplot* dos erros para 70% de ruído.



Fonte: Elaborado pelos autores.

Portanto, testes de filtragem foram feitos para dados de NDVI MODIS na bacia do rio Paraíba do Sul, uma área de estudo que possui entre suas classes de cobertura, a agricultura, a floresta, a pastagem e a silvicultura. Os resultados do presente trabalho geram conhecimento para futuros estudos que utilizem dados de IVs MODIS para análise da cobertura da terra no domínio morfoclimático de mares de morros que coincide com o bioma Mata Atlântica (especificamente, o sudeste brasileiro).

5 Conclusões

Os valores de RMSE dos filtros foram próximos para os três níveis de ruído, assim como os padrões de erros nos gráficos *boxplot* foram semelhantes nos três níveis de ruído, e para as quatro classes de cobertura.

Os filtros testados permitiram a obtenção de gráficos visualmente adequados para a representação temporal, conservando a forma e os detalhes temporais, em especial o filtro *4253H twice*.

Nas simulações de 10% e 40% de ruído, o menor erro foi para o filtro de mediana e o segundo menor para o filtro *4253H twice*.

Para a simulação com 70% de ruído, o menor erro foi para o filtro de média, o segundo para o filtro de mediana e o terceiro para o *4253H twice*.

Apesar de o menor valor de RMSE, o filtro de mediana modificou mais a forma da série temporal em relação ao filtro *4253H twice*, o que permite sugerir este segundo para a filtragem de dados MODIS. Este fato também é corroborado pelo bom desempenho do filtro *4253H twice* na literatura.

Os erros são maiores para as classes com menores variações de NDVI (pastagem e agricultura), enquanto que as classes floresta e silvicultura apresentaram menores valores de RMSE.

Referências

- ABADE, N. A.; CARVALHO JUNIOR, O. A.; GUIMARÃES, R. F.; OLIVEIRA, S. N. Comparative Analysis of MODIS Time-Series Classification Using Support Vector Machines and Methods Based upon Distance and Similarity Measures in the Brazilian Cerrado-Caatinga Boundary. **Remote Sensing**, v. 7, 2015. pp. 12160-12191.
- 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. Site: <<http://urlib.net/8JMKD3MGP7W/3897URE>>. Acessado em março de 2017.
- ADAMI, M.; RUDORFF, B. F. T.; FREITAS, R. M.; AGUIAR, D. A.; SUGAWARA, L. M.; MELLO, M. P. Remote Sensing Time Series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil. **Sustainability**, v. 4, 2012. pp. 574-585.

- 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, 2011. pp. 3689-3709.
- BEURS, K. M.; HENEGBRY, G. M.; OWSLEY, B. C.; SOKOLIK, I. Using multiple remote perspectives to identify and attribute land surface dynamics in Central Asia 2001-2013. **Remote Sensing of Environment**, v. 170, 2015. pp. 48-61.
- BORGES, E. F.; SANO, E. E.; MEDRADO, E. Radiometric quality and performance of TIMESAT for smoothing moderate resolution imaging spectroradiometer enhanced vegetation index time series from western Bahia State, **Brazil. J. Appl. Remote Sens.**, v. 8, n. 1, 2014. pp. 1-18.
- BRANDT, M. **Smoothing/Filtering a NDVI time series using a Savitzky Golay filter and R.** Site: <https://matinbrandt.wordpress.com/2014/12/02/smoothingfiltering-a-ndvi-time-series-using-a-savitzky-golay-filter-and-r/> Acessado em abril de 2018.
- CARVALHO JUNIOR, O.; SILVA, N.C.; CARVALHO, A.P.F.; COUTO JUNIOR, A.F.; SILVA, C.R.; SHIMABUKURO, Y.E.; GUIMARÃES, R.F.; GOMES, R.A.T. Combining noise-adjusted principal components transform and median filter techniques for denoising modis temporal signatures. **Revista Brasileira de Geofísica**, vol.30, n. 2, 2012. pp. 147-157.
- 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, 2000. pp. 690-698.
- CHEN, B.; XU, G.; COOPS, N. C.; CIAIS, P.; INNES, J. L.; WANG, G.; MYENI, R. B.; WANG, T.; KRZYZANOWSKI, J.; LI, Q.; CAO, L.; LIU, Y. Changes in vegetation photosynthetic activity trend across the Asia-Pacific region over the last three decades. **Remote Sensing of Environment**, v. 144, 2014. pp. 28-41.
- CHEN, J.; JONSSON, P.; TAMURA, M.; GU, Z.; MATSUSHITA, B.; EKLUNDH, E. 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, 2004. pp. 332-344.
- DIDAN, K. **MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006**. NASA EOSDIS Land Processes DAAC, 2015.

- 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, 2009. pp. 248-258.
- 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, 2002. pp. 195-213.
- JU, J.; MASEK, J. G. The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data. **Remote Sensing of Environment**, v. 176, 2016. pp. 1-16.
- KANDASAMY, S.; BARET, F.; VERGER, A.; NEVEUX, P.; WEISS, M. A comparison of methods for smoothing and gap filling time series of remote sensing observations – application to MODIS LAI products. **Biogeosciences**, v. 10, 2013. pp. 4055-4071.
- KOVALSKYY, V.; ROY, D. P.; ZHANG, X. Y.; JU, J. The suitability of multi-temporal web-enabled Landsat data NDVI for phenological monitoring – a comparison with flux tower and MODIS NDVI. **Remote Sensing Letters**, v. 3, n.4, 2012. pp. 325-334.
- MADDEN, H. H. Comments on the Savitzky-Golay convolution method for leastsquares-fit smoothing and differentiation of digital data. **Analytical Chemistry**, v.50, n.9, 1978. pp. 1383-1386.
- MARCOS, B.; GONÇALVES, J. F.; ALCARAZ-SEGURA, D.; CUNHA, M. REMOTE SENSING INDICATORS OF CHANGES IN ECOSYSTEM FUNCTIONING RELATED TO WILDFIRE DISTURBANCES. In: 2nd EARSeL International Workshop on Temporal Analysis of Satellite Images, 2., 2015, Estocolmo. **Anais**, 2015, pp. 1-8.
- MARIANO, D. A. **Detecção e avaliação de seca agrônômica através da análise de séries temporais de dados MODIS e PERSIANN**. 2015. 110 p. (sid.inpe.br/mtc-m21b/2015/02.12.20.29-TDI). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2015.

- MILDREXLER, D.; YANG, Z.; COHEN, W. B.; BELL, M. D. A forest vulnerability index based on drought and high temperatures. **Remote Sensing of Environment**, v. 173, 2016. pp. 314-325.
- MISHRA, N. B.; CREWS, K. A.; NEETI, N.; MEYER, T.; YOUNG, K. R. MODIS derived vegetation greenness trends in African Savanna: deconstructing and localizing the role of changing moisture availability, fire regime and anthropogenic impact. **Remote Sensing of Environment**, v. 169, 2015. pp. 192-204.
- MICHISHITA, M.; JIN, Z.; CHEN, J.; XU, B. Empirical comparison of noise reduction techniques for NDVI time-series based on a new measure, ISPRS **Journal of Photogrammetry and Remote Sensing**, v. 91, 2014. pp. 17-28.
- OVALLE, A. R. C.; SILVA, C. F.; REZENDE, C. E.; GATTS, C. E. N.; SUZUKI, M. S.; FIGUEIREDO, R. O. Long-term trends in hydrochemistry in the Paraíba do Sul River, southeastern Brazil. **Journal of Hydrology**, v. 481, 2013. pp. 191-203.
- PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. p.144. São José dos Campos - SP: Editora Parêntese, 2009.
- SAKAMOTO, T., M. YOKOZAWA, H. TORITANI, M. SHIBAYAMA, N. ISHITSUKA e H. OHNO. A crop phenology detection method using time-series MODIS data. **Remote Sensing of Environment**, v.96, 2005. pp. 366-374.
- ROUSE JR., J. W., R. H. HASS, J. A. SCHELL e D. W. DEERING. Monitoring vegetation systems in the great plains with ERTS. Earth Resources Technology Satellite-1 Symposium. 3., 1973. Washington. **Proceedings...Washington**: NASA: 309317 p. 1973.
- SAKAMOTO, T. et al. A crop phenology detection method using time-series MODIS data. **Remote Sensing of Environment**, v. 96, n. 3-4, 2005. pp. 366-374.
- SAVITZKY, A.; GOLAY, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v.36, n.8, 1964. pp. 1627-1639.
- SHAO, Y.; LUNETTA, R. S.; WHEELER, B.; IIAMES, J. S.; CAMPBELL, J. B. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. **Remote Sensing of Environment**, v. 174, 2016. pp. 258-265.

- STEINIER, J.; TERMONIA, Y.; DELTOUR, J. Smoothing and differentiation of data by simplified least square procedure. **Analytical Chemistry**, v.44, n.11, 1972. pp. 1906-1909.
- THAYN, J. B.; PRICE, K. P. Julian dates and introduced temporal error in remote sensing vegetation phenology studies. **Remote Sensing Letters**, 2008. pp. 6045-6049.
- VELLEMAN, P. F. Definition and comparison of robust non-linear data smoothing algorithms. **Journal of the American Statistical Association**, v.75, n.371, 1980. pp. 609-615.
- WHITE, M. A. et al. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. **Global Change Biology**, v. 15, n. 1, 2009. pp. 2335-2359.
- ZHOU, J.; JIA, L.; VAN HOEK, M.; MENENTTI, M.; LU, J.; HU, G. AN OPTIMIZATION OF PARAMETER SETTINGS IN HANTS FOR GLOBAL NDVI TIME SERIES RECONSTRUCTION. In: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, **Proceedings...** 2016.