

Avaliação de métodos para Reconstrução de Pixel com baixa qualidade em séries temporais de NDVI

Julio Cesar de Oliveira¹
José Carlos Neves Epiphânio²
Camilo Daleles Rennó²
Marcos Adami²

¹Professor Assistente da Universidade Federal de Viçosa – UFV, Pós-Graduando em Sensoriamento Remoto no Instituto Nacional de Pesquisas Espaciais – INPE
oliveirajc@ufv.br

²Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{epiphânio, camilo, adami}@dsr.inpe.br

Abstract. Normalized Difference Vegetation Index is a vegetation index widely applied in research. However, noise induced by cloud contamination and atmospheric variability affect the data quality. The literature contains reference to a broad variety of strategies designed to reduce the impacts of these factors. This paper inspects four NDVI time series noise-reduction techniques to reduce the impacts of such factors. Performance of techniques was evaluated based on reconstruction of the original record or on minimizing the Mean Absolute Percentage Error (MAPE).

An empirical comparison of four strategies revealed the general superiority of the spatial-temporal analysis used by Window Regression technique. This study demonstrated the strong influence of noise level on the performance of the techniques based on time domain. The results indicate that the spatial-temporal information is a promising method for reconstructing high-quality MODIS NDVI time series, but users need to consider both their objectives and the nature of the noise present in an NDVI data set when selecting a technique for noise reduction.

Palavras-chave: Time Series, MODIS NDVI, NoiseReduction.

1. Introdução

Séries temporais do *NormalizedDifferenceVegetation Index* (NDVI) provenientes do *ModerateresolutionImagingSpectroradiometer* (MODIS) têm mostrado ser adequadas tanto para a modelagem de ecossistemas terrestres como para o monitoramento e obtenção de características biofísicas da vegetação. 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 séries temporais de índice de vegetação devem ser minimizados antes da sua utilização (HIRD; MCDERMID, 2009; HUETE *et al.*, 1999)

Adami (2010) define uma série temporal como um conjunto de observações ordenadas no tempo, onde a forma de análise é a modelagem da autocorrelação entre as observações. Esta modelagem pode ser realizada no domínio da frequência, por meio de modelos não paramétricos (como por exemplo, a transformada de Fourier), ou no domínio do tempo por meio de modelos paramétricos, como o *AutoregressiveIntegratedMovingAverage* (ARIMA) (MORETTIN, 1999).

Trabalhos como os de Adami (2010), 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, Ma e Veroustraete (2006) e Hird e Mcdermid (2009) destacam a necessidade de reduzir os ruídos presentes no perfil temporal de NDVI.

Estão disponíveis na literatura inúmeras metodologias com a finalidade de reduzir a presença desses ruídos e reconstruir as séries temporais de NDVI com melhor qualidade (BRADLEY *et al.*, 2007; CHEN *et al.*, 2004; HIRD; MCDERMID, 2009; MA; VEROUSTRAETE, 2006; MUSIAL *et al.*, 2011; SWETS *et al.*, 1999). Basicamente, essas metodologias de filtragem trabalham no domínio da frequência ou do tempo, ou seja, analisam o comportamento espectral do pixel em uma dimensão da informação. Outras estratégias, como a proposta por Oliveira *et al.* (2012) utilizam uma combinação do domínio espaço-temporal para a reconstrução da série temporal de NDVI.

O presente estudo tem como objetivo comparar o desempenho de técnicas de filtragem na redução de ruídos em séries temporais de NDVI. A avaliação será realizada por meio de uma análise comparativa das discrepâncias entre o dado filtrado e o real, conforme a intensidade de ruído aplicado nas séries.

As quatro técnicas de filtragem avaliadas neste estudo são os filtros 4253H twice (VELLEMAN, 1980); Savitzky–Golay (SAVITZKY; GOLAY, 1964); Valor Médio da Iteração (MA; VEROUSTRAETE, 2006) e Janela de Regressão (OLIVEIRA *et al.*, 2012). Uma diferença metodológica entre filtros é que os três primeiros trabalham no domínio do tempo, enquanto que a quarta técnica de filtragem utiliza a informação espaço-temporal para redução de ruídos em séries temporais.

2. Materiais e Métodos

A base de dados analisada é constituída por uma série temporal do produto MOD13Q1 (Tabela 1) com um recorte equivalente a 8x8 pixels sobre uma área agrícola. De acordo com produto *Science Datasets* (SDS), disponibilizado no produto MOD13Q1, todas as 44 datas selecionadas apresentam excelente confiabilidade do pixel.

Tabela 1: Datas das imagens que compõem a série temporal MOD13Q1 analisada.

MOD13Q1 – Índice de Vegetação (Composição de 16 Dias - 250m - tile h13v11)																
2003	Tempo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Dia do ano	033	049	065	081	097	113	129	145	161	177	193	209	225	241	257
	Tempo	16	17	18	19	20	21									
	Dia do ano	273	289	305	321	337	353									
2004	Tempo	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	Dia do ano	001	017	033	049	065	081	097	113	129	145	161	177	193	209	225
	Tempo	37	38	39	40	41	42	43	44							
	Dia do ano	241	257	273	289	305	321	337	353							

No presente estudo a avaliação do desempenho das técnicas para redução de ruídos descritas apoia-se em medidas de discrepância entre o valor original de NDVI e o correspondente valor filtrado. Desta forma, a metodologia para realização desta pesquisa foi subdividida em três etapas: seleção de curvas temporais; aplicação de ruídos sobre as curvas e filtragem delas; e, por fim, a avaliação das discrepâncias (Figura 1).

Etapa 1 – Seleção dos pontos para análise

A primeira etapa está ilustrada na Figura 1 e consiste na elaboração de métodos de amostragem de pixels e datas a serem analisados dentro de uma série temporal de NDVI. Com o objetivo de verificar a influência espaço-temporal na filtragem das séries, principalmente no método de redução de ruídos que utilizam a informação espacial em sua análise, o processo de amostragem foi subdividido em 2 tipos: (A) amostragem de datas ao longo da série temporal (amostragem temporal); (B) amostragem de datas e pixels (amostragem espaço-temporal).

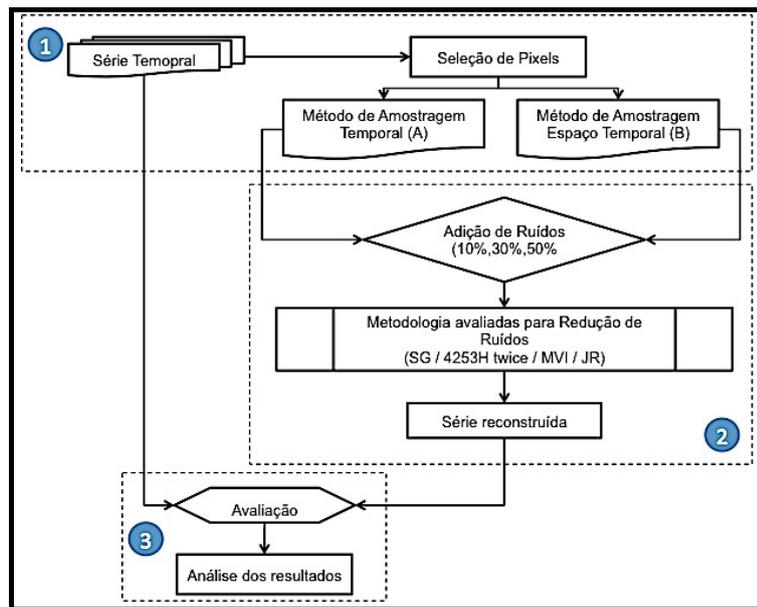


Figura 1 – Fluxograma metodológico das etapas desenvolvidas.

O método de amostragem “A” consiste na aplicação de ruído em uma porcentagem das datas que compõem a série temporal de um pixel da área de estudo. Destaca-se que os pixels presentes nas bordas da imagem não foram utilizados em ambos os procedimentos de amostragem. Sendo assim, dos 64 pixels (matriz de 8x8 pixels) presentes na área de estudo 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 apto para amostragem somente o intervalo da 4a até a 41a data que compõe a série temporal (totalizando 38 datas possíveis). Desta forma, os pontos para análise nomeados pela amostragem “A” consistem na seleção de 30% das 38 datas possíveis em cada curva espectral dos 36 pixels da área de estudo.

O segundo método de amostragem utiliza o mesmo intervalo de datas e pixels possíveis descritos anteriormente, porém para cada data selecionada foi aplicado ruído para mais de um pixel. Desta forma, foram selecionadas 30% das datas possíveis da série temporal, e para cada data selecionada foram escolhidos aleatoriamente 4 pixels para aplicação de ruídos. A Figura 2 ilustra os dois procedimentos de amostragem.

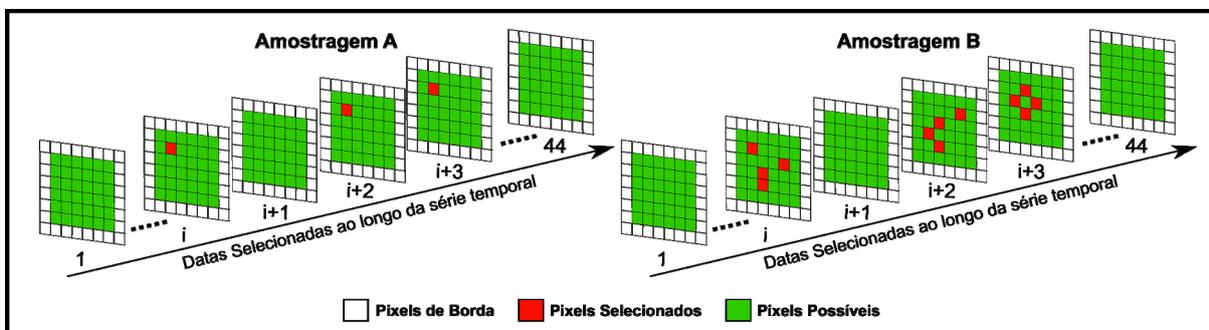


Figura 2 – Ilustração da amostragem temporal (A) e espaço-temporal (B).

Etapa 2 – Aplicação de ruídos e filtragem das curvas

Para cada pixel (X,Y) da data i selecionado na etapa 2 foram aplicados 3 níveis de ruído (10%, 30% e 50%) conforme proposto por Adami (2010), e Hird e McDermid (2009). Os níveis de ruídos de 10%, 30% e 50% representam a porcentagem do valor de NDVI que será

adicionado ou subtraído no valor original do pixel $(X,Y)_i$. Destaca-se que tanto para o processo de amostragem da etapa 2 como para a definição se o ruído é aditivo ou subtrativo, foram utilizados procedimentos de escolha aleatória.

Sobre as séries degradadas foram aplicados os filtros descritos no presente trabalho com o objetivo de reduzir ou eliminar os ruídos introduzidos.

Etapa 3 – Avaliação dos resultados

O objetivo nesta etapa do trabalho foi avaliar quantitativamente o desempenho dos filtros escolhidos para a redução de ruídos. Sendo assim, para cada pixel $(X,Y)_i$ onde foi adicionado ruído obteve-se a discrepância entre o valor filtrado e o original.

2.1. Qualidade do ajuste

Musialet *al.* (2011) destacam 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 Matura (2005) sugerem o uso da média do erro absoluto (*MeanAbsoluteError - 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 reais 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 a performance geral da média, enquanto que o RMSE penaliza grandes diferenças individuais.

As diferentes escalas presentes entre séries temporais podem induzir a comparações inexpressivas. Por isso, foi utilizado o erro médio absoluto relativo (*MeanAbsolutePercentageError- MAPE*) para comparar o desempenho das técnicas, além de que Willmott e Matura (2005) classificam o erro médio absoluto como uma medida de dispersão não ambígua da magnitude do erro médio. A formulação das medidas de dispersão estão apresentadas na Equação 1.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - x_i)^2} \quad MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - x_i| \quad MAPE = \frac{1}{n} \cdot \sum_{i=1}^n \left(\left| \frac{y_i - x_i}{x_i} \right| \cdot 100 \right) \quad (1)$$

onde o n é o número de pontos avaliados, y_i o dado filtrado e x_i as observações presentes na série temporal.

3. Resultados e Discussão

A amostragem “A” propõem uma degradação em 30% das datas disponíveis de uma curva temporal de NDVI, ou seja, em 12 datas da série temporal de um pixel foi aplicado 3 níveis de ruídos conforme ilustrado na Figura 3.

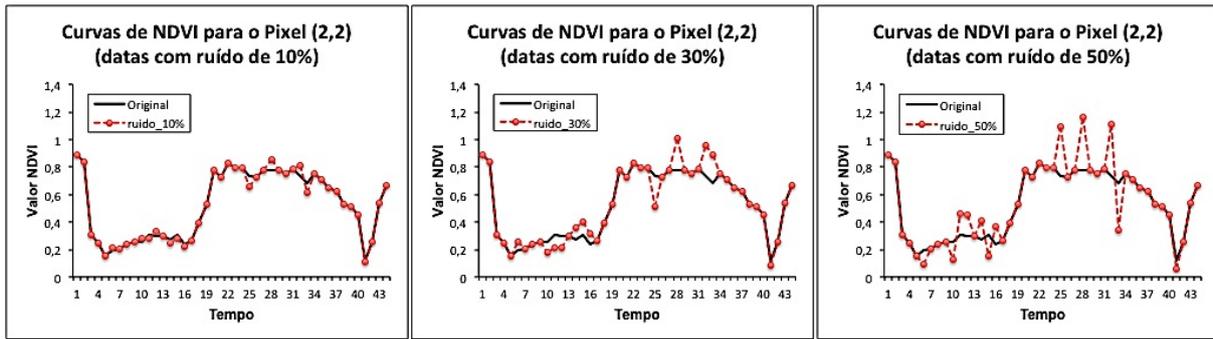


Figura 3 – Curvas original e degradada com 3 níveis de ruídos para o pixel (2,2).

Após a degradação da curva da série temporal foi efetuada a redução dos ruídos com base nas 4 técnicas de filtragem selecionadas. Posteriormente foram calculados as discrepâncias absolutas entre o valor de NDVI da curva original e filtrada para as datas degradadas. Um exemplo deste procedimento está ilustrado na Figura 4.

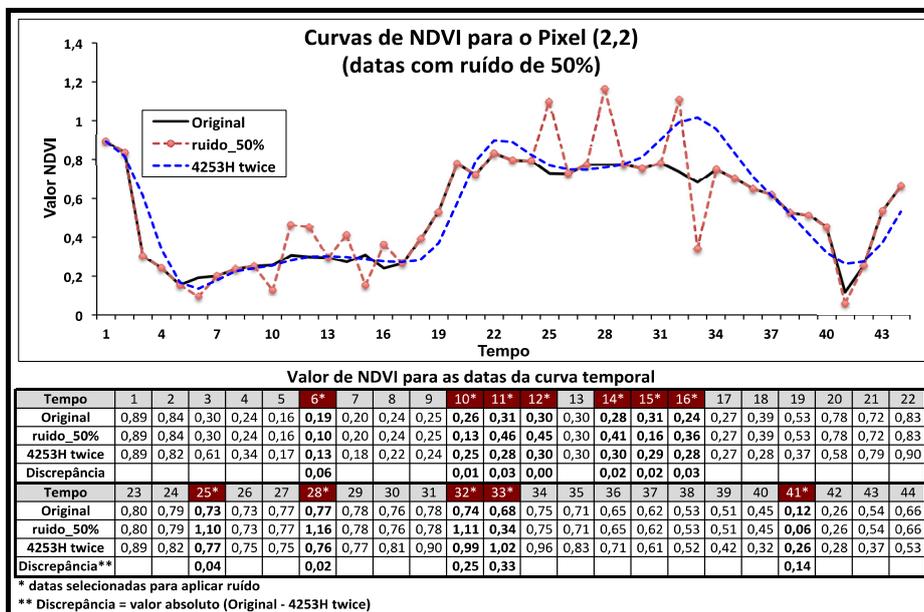


Figura 4 – Valores das discrepâncias para as datas degradadas com base nas curvas original e filtrada (filtro 4253H twice).

Para a amostragem “A” o mesmo procedimento realizado no pixel (2,2) foi repetido para os demais pixels da imagem (exceto os pixels de borda). Portanto, em única iteração foram gerados 432 pontos (36 pixels x 12 datas/pixel) para a análise da performance das técnicas de redução dos ruídos em cada um dos níveis de ruídos propostos. Porém, todo o procedimento de amostragem, filtragem e cálculo das discrepâncias foi repetido 50 vezes, o que resultou em 21000 pontos analisados para a amostragem “A”.

No procedimento de amostragem “B” foram selecionados 4 pixels para cada data e um total de 48 pontos por iteração, de modo que após 500 iterações foram analisados 24000 pontos. A qualidade do ajuste entre os filtros (4253H twice; Valor Médio da Iteração (MeanValueIteration - MVI); Savitzky–Golay (SG); Janela de Regressão (JR)) para cada método de amostragem e nível de ruído está ilustrada na Figura 5.

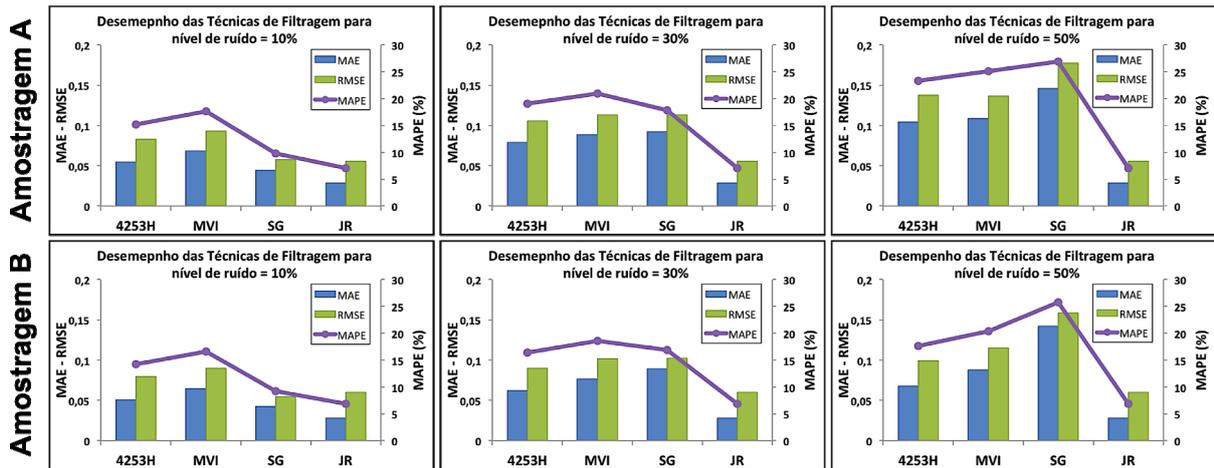


Figura 5 – Medidas de dispersão das discrepâncias dos filtros (4253H– MVI – SG – JR)de acordo com os níveis de ruído de 10%, 30% e 50%.

As medidas de dispersão que avaliaram a performance das técnicas de filtragem apresentaram o mesmo comportamento para ambas as metodologias de amostragem (Figura 5). 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 no comportamento destes filtros. Já para a técnica Janela de Regressão, que utiliza a informação espacial para estimar o valor de um determinado ponto, a degradação de 4 pixels para uma mesma data não influenciou na qualidade do ajuste desta técnica.

Pode-se observar que quanto maior a intensidade do ruído pior é o desempenho das técnicas de filtragem 4253H twice, MVI e SG. Este comportamento pode ser verificado em todas as medidas de dispersão (RMSE, MAE e MAPE) que avaliaram os resultados. Já a técnica Janela de Regressão não é influenciada pelo nível de ruído, 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).

A Figura 6 apresenta a variação do erro médio absoluto relativo (MAPE) para cada técnica de filtragem, nível de ruído e método de amostragem implementados. Pode-se observar que a técnica JR apresentou o melhor desempenho (menor variação do MAPE) entre as técnicas avaliadas, ao mesmo tempo que os filtros 4253H twice, MVI e SG resultaram em desempenhos piores com o aumento do nível de ruído. Destaca-se que na Figura 6 não foram apresentados os pontos considerados *outliers* pelo fato de que em todas as técnicas o número total desses pontos ter sido bem inferior quando comparado ao número total analisado (acima de 21000 pontos em cada tipo de amostragem).

5. Conclusão

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

Neste trabalho quatro técnicas de filtragem foram avaliadas em termo da performance na redução de ruídos em séries temporais de NDVI. O erro médio absoluto relativo foi escolhido para comparar o ajuste das técnicas, além de dois métodos de amostragem (temporal e espaço-temporal) implementados para identificar limitações das técnicas.

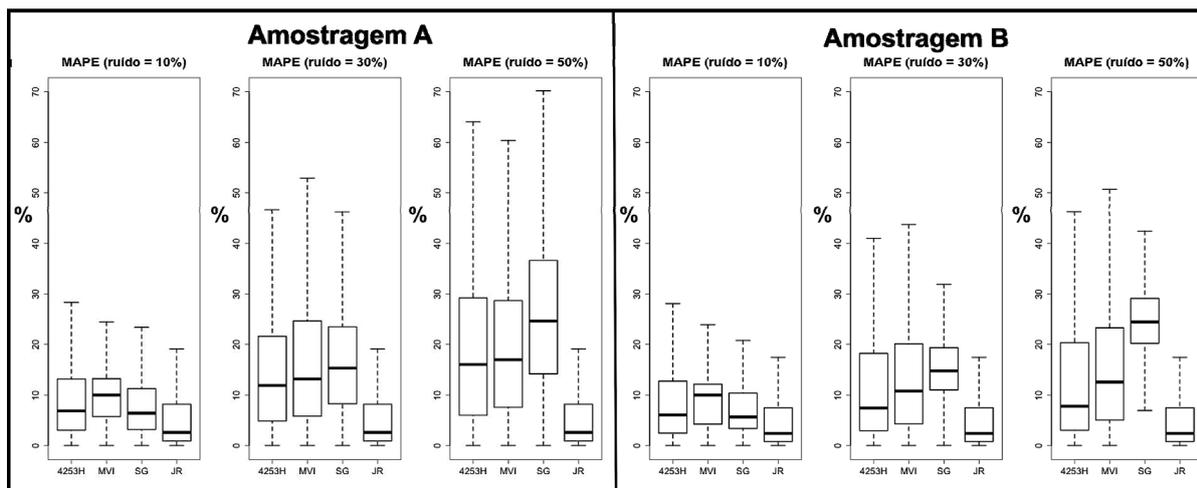


Figura 6–Boxplot do erro médio absoluto relativo para as técnicas de filtragem (4253H, MVI, SG e JR) e respectivos níveis de ruído (10%, 30% e 50%).

O desempenho das técnicas 4253H twice, Valor Médio da Iteração (MVI) e Savitzky–Golay não foi influenciado pelo processo de amostragem, pois elas mesmas não dependem da informação espacial no procedimento para redução de ruídos. Porém, estas técnicas apresentam queda no desempenho conforme o aumento do nível de ruído implementado, com destaque para o filtro SG que apresentou a maior queda de desempenho.

Já o desempenho da técnica Janela de Regressão (JR) não foi influenciado pelas amostragens realizadas, ou seja, tanto para seleção aleatória de 30% da datas ao longo de uma curva espectral, como para degradação em 4 pixels em uma mesma imagem/data, o desempenho desta técnica foi o melhor. Apesar de esta técnica manter o mesmo desempenho para todos os níveis de ruídos ou tipos de amostragem, a qualidade do ajuste entre o dado filtrado e o original que esta técnica proporciona está limitado ao número de vizinhos com boa garantia de qualidade para um determinado pixel a ser analisado. Sendo assim, o desempenho da técnica Janela de Regressão será influenciado negativamente quanto maior for o intervalo de datas consecutivas degradadas (gap) ou número de pixels em uma mesma data a serem classificados como de baixa qualidade.

Existem vários métodos para reconstrução de séries temporais de NDVI, porém poucos utilizam a informação da qualidade do pixel no procedimento de redução de ruídos. 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 cautela.

Musialet *al.* (2011) destacam que a escolha de uma abordagem para estimar valores perdidos (ou de baixa qualidade) em séries temporais depende do entendimento do sinal, da distribuição das datas com gaps, e da expectativa do analista em obter dados próximos aos existentes (baixo MAE) ou uma representação das características gerais da série histórica.

6. Referências Bibliográficas

ADAMI, M. **Estimativa da data de plantio da soja por meio de séries temporais de imagens MODIS**. [S.l.]: Instituto Nacional de Pesquisas Espaciais, 2010.

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 noisysatellite NDVI data. **Remote Sensing of Environment**, v. 106, n. 2, p. 137-145, 30 jan 2007.

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.

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.

HUETE, A.; JUSTICE, C. O.; LEEUWEN, W. J. D. VAN. **MODIS Vegetation Index (MOD13): Algorithm Theoretical Basis Document**. [S.l.: s.n.]. Disponível em: <http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf>. , 1999

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.

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.

MORETTIN, P. A. **Ondas e ondaletas: Da análise de Fourier a análise de Ondaletas**. Editora US ed. São Paulo: [s.n.], 1999. p. 272

MUSIAL, J. P.; VERSTRAETE, M. M.; GOBRON, N. Comparing the effectiveness of several algorithms to fill and smooth incomplete and noisy time series. **Atmospheric Chemistry and Physics Discussions**, v. 11, n. 5, p. 14259-14308, 10 maio 2011.

OLIVEIRA, J. C. DE; EPIPHANIO, J. C. N.; RENNÓ, C. D. **The use of spatial-temporal analysis for noise reduction in MODIS NDVI time series data**. 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2012). **Anais...** [S.l.: s.n.]. Disponível em: <<http://spatial-accuracy.org/system/files/OliveiraAccuracy2012.pdf>>. , 2012

ROY, D. P.; BORAK, J. S.; DEVADIGA, S.; WOLFE, R. E. The MODIS land product quality assessment approach. **Remote Sensing of Environment**, v. 83, p. 62-76, 2002.

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, 30 jun 2005.

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, 1 jul 1964.

SWETS, D. L.; REED, B. C.; ROWLAND, J. D.; MARKO, S. E. **A Weighted Least-Squares Approach to Temporal NDVI Smoothing**. 1999 ASPRS Annual conference □: from image to information. **Anais...** [S.l.]: American Society for Photogrammetry and Remote Sensing. , 1999

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.

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, 2005.