Evaluation of smoothing methods on Landsat-8 EVI time series for crop classification based on phenological parameters

Hugo do Nascimento Bendini [1]
Leila Maria Garcia Fonseca [1]
Thales Sehn Körting [1]
Ieda Del Arco Sanches [1]
Rennan Freitas Bezerra Marujo [1]

[1] Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
hbendini@dsr.inpe.br
{leila.fonseca, thales.korting, ieda.sanches, rennan.marujo}@inpe.br

Abstract. This study aims to evaluate three different time series smoothing methods, combined or not with filtering techniques, and their impact on the agricultural land use classification, in a region of the Brazilian Cerrado, using phenological parameters extracted from Enhanced Vegetation Index (EVI) Landsat-8 image time series. We extracted the time series from pixels located on well know polygons delimited on agricultural lands and monitored on a field campaign during August 2015 and August 2016. For the classification we considered the following classes: Annual Agriculture, Natural Forest, Perennial Agriculture, Semi-Perennial Agriculture and Grassland. The three smoothing algorithms were implemented through the TIMESAT software package including the: Savitzky-Golay (SG), asymmetric Gaussian function (AG) and double-logistic function (DL), and then the phenological attributes were extracted. For each method the phenological attributes were subjected to data mining using the Random Forest (RF) algorithm. The results were evaluated by the confusion matrix analysis, including global accuracy, producer´s accuracy and kappa. The intra-class variability was measured by calculating the mean standard deviation for samples within the different classes. The best classification accuracy with the different smoothing methods was the SG applied to the raw time series, with a global accuracy of 86% and kappa of 0.82.

Key-words: remote sensing, data mining, Random Forest, TIMESAT, multi-temporal analysis.

1. Introduction

With the availability of free and continuous satellite imagery, which allows the construction of consistent time series of vegetation index images, remote sensing undergoes a paradigm shift with regard to monitoring changes in land use and land cover.

In the time series, each image pixel can be treated as a signal, so that signal processing techniques and econometrics can be applied, such as the decomposition of time series in trend and seasonality components, extraction of parameters of these components to land cover and land use classification (Arvor et al., 2011; Zheng et al., 2015), and change detection and trajectories analysis (Verbesselt et al., 2010). One of the important steps for this type of study is the preprocessing of the time series for noise removal, usually caused by the presence of clouds. Several algorithms for noise removal in time series have been used, among them, the Savitzky-Golay smoother (Chen et al., 2004), asymmetric Gaussian functions (Jönsson and Eklundh, 2004) and Double-logistic (Zhang et al., 2003; Jönsson and Eklundh, 2004). In addition to the application of smoothing algorithms, cloud masks combined with outliers interpolation techniques as well as multi-sensor approaches for replacing contaminated pixels by cloud-free pixels have been used (Hamunyela et al., 2013; Bendini et al., 2016a; Bendini et al., 2016b). Some studies have compared these smoothing approaches, but most of them focuses on coarse spatial resolution satellite image time series, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Atzberger and Eillers, 2011; Borges and Sano, 2014; Shao et al., 2016). Furthermore, they do not take into account the filtering effect on the classification performance.

ISBN: 978-85-17-00088-1

Bendini et al. (2016) evaluated a method for integration of OLI/Landsat-8 and MUX/CBERS-4 images to obtain vegetation indices time series with lower cloud contamination to improve a crop classification in the Cerrado Biome. They used phenological attributes extracted from time series smoothed by the double-logistic algorithm, since this algorithm was referenced by Borges and Sano (2014) as being the most suitable for smoothing series in agricultural areas in the Cerrado, considering MODIS images. However, there is a lack of information to confirm if this comparison is valid for Landsat-like image time series. Thus, the objective of this study was to evaluate three time series smoothing methods, Savitzky-Golay, asymmetric Gaussian and Double-logistic functions, combined or not with filtering techniques, and also evaluate the impact of their use on agricultural use classification in a region of the Cerrado, using phenological parameters extracted from EVI Landsat-8 image time series.

2. Work methodology

We conducted our study in the Itobi municipality, in São Paulo state, Brazil (Figure 1). Field campaigns were carried out for collecting training samples, in turn of 100 pixels. The classes considered were Annual Agriculture (potato, soybean, corn, onions and sugar beet on a double crop system), Semi-Perennial Agriculture (sugarcane), Perennial Agriculture (avocado, mango and Brazilian grapetree), Natural Forest and Grassland. A total of 24 scenes of Landsat-8 OLI (WRS 2, Path/Row 219/75), acquired from August 2015 to August 2016, were processed to Level 1 Terrain Corrected. The images were corrected for atmospheric conditions to identify and mask cloud and cloud shadows by the USGS EROS Science Processing Architecture (ESPA) (DeVries et al. 2015; DeVries et al. 2015a). Landsat-8 data were corrected using the L8SR algorithm (U.S. Geological Survey, 2015; Vermote, 2016).

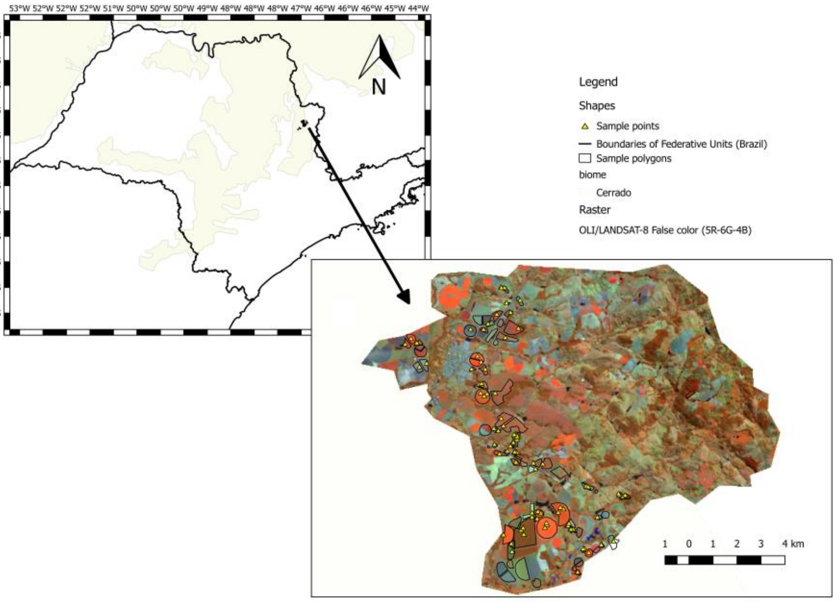

Figure 1. Location of Itobi municipality, in São Paulo state, Brazil.

We first applied a combined filtering approach for noise removal on the Landsat time series by replacing the noise values by the average of their nearest neighbors values in time, considering the Fmask quality data and negative outliers based on a threshold recommended by Hamunyela et al. (2013). This method, however, is not capable of removing consecutive outliers. For both cases (filtered and raw data) we implemented three smoothing algorithms through the TIMESAT software package including the Savitzky-Golay, asymmetric Gaussian and double-logistic functions (Jönsson and Eklundh, 2004). Afterwards, the phenological attributes were extracted using TIMESAT.

The Savitzky–Golay approach applies a moving window to a given time-series dataset. Within a moving window (e.g., 2n + 1 points, n is a user defined window width), a quadratic polynomial function is used to fit all points and then the value of the central point is replaced by the fitted value. The asymmetric Gaussian algorithm relies mainly on five parameters to fit time-series data including the time of the minimum or maximum EVI, the width and flatness of the right side of the function, and the width and flatness of the left side of the function. The double logistic function estimates four parameters to determine the left inflection point, the right inflection point, and rates of changes at two inflection points. Both the asymmetric Gaussian algorithm and the double logistic function are modifications of local model functions, which have been proven to be effective in capturing phenological cycle events defined by EVI curves (Shao et al., 2016). Details about these processing techniques can be seen in Jönsson and Eklundh (2015). Considering the phenological attributes, a Random Forest (RF) algorithm (Breiman, 2001) was applied to the raw Landsat EVI time series smoothed by 1) Savitzky-Golay function, 2) Asymmetric Gaussian, and 3) Double logistic function, and also to the filtered Landsat EVI time series smoothed by 4) Savitzky-Golay function, 5) Asymmetric Gaussian and 6) Double- logistic function. The RF algorithm has been widely used in remote sensing applications (Müller et al, 2015; Peña et al, 2015) because it efficiently handle large databases, providing estimates on the most relevant variables, and also allowing the identification of outliers (Rodriguez-Galiano et al., 2012). There were a total of 31 training pixels for the annual agriculture class, 15 pixels for perennial agriculture, 26 pixels for semi-perennial agriculture, 14 pixels for grassland and 14 pixels for native forest. The results were evaluated through the confusion matrix index, as global accuracy (GA), Kappa and producer's accuracy (PA) (Witten et al, 2011). The models considered a 10-fold cross validation method. The intra-class variability was measured by calculating the mean standard deviation for samples within the different classes (Shao et al., 2016). The classification results were obtained using the software package WEKA (Hall et al., 2009).

3. Results and discussions

The classification accuracy assessment, considering different smoothing algorithms applied to EVI raw time series (DL, AG, SG), is presented in Table 1. It includes the global accuracy (GA), kappa statistic, and the producer's accuracy (PA) for each class.

Table 1. Classification accuracy assessment, considering different smoothing methods.

	Producer´s Accuracy (%)					
Smoothing methods	*Annual Agriculture*	*Natural Forest*	*Perennial Agriculture*	*Semi-perennial Agriculture*	*Grassland*	***GA% (kappa)***
DL	84.00%	75.00%	53.57%	90.00%	61.90%	70.00% (0.62)
AG	89.66%	78.57%	76.92%	73.68%	66.67%	79.00% (0.73)
SG	92.86%	85.71%	78.57%	82.35%	92.31%	86.00% (0.82)

The best classification result was obtained by the Savitzky-Golay smoothed data, with an overall accuracy of 86% (kappa of 0.82). This result is followed by the Asymmetric Gaussian method, with overall accuracy of 79% (kappa of 0.73), and by the Double-logistic function, with overall accuracy of 70% (kappa of 0.62). We can observe that the only class that did not obtained the highest producer´s accuracy using the Savitzky-Golay smoothed time series was the Semi-perennial agriculture (82.35%). On the other hand, the highest producer´s accuracy for this class was found with the Double-logistic method (90%). Figure 2 shows the smoothed EVI time series, using different smoothing algorithms applied to the raw time series. It also shows the points of start and end of seasons detected by the TIMESAT´s algorithm to extract the phenological attributes. The classification results considering the different smoothing methods on the raw EVI time series can be observed in Table 2.

ISBN: 978-85-17-00088-1

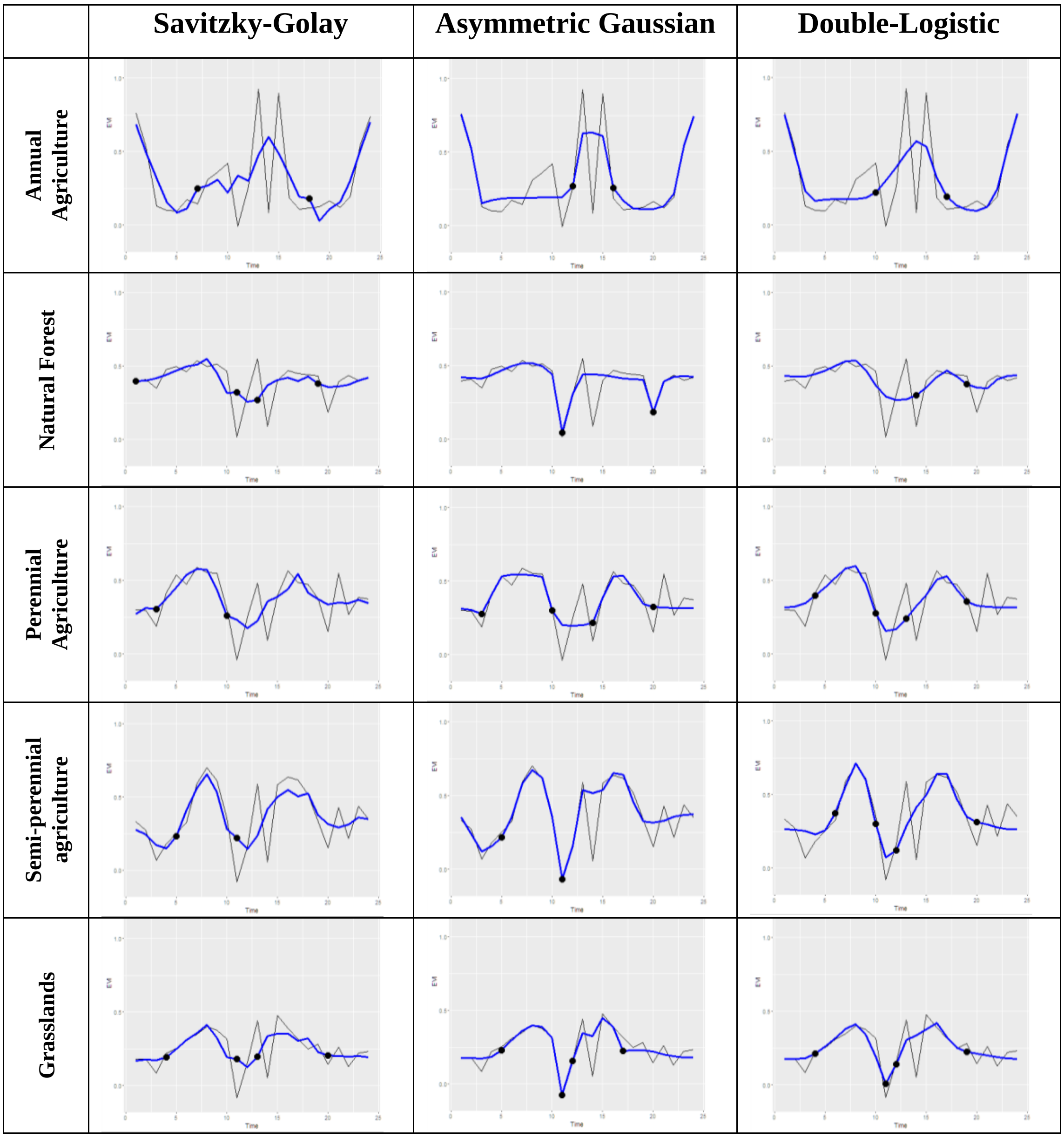

Figure 2. Smoothed EVI time series with the different algorithms using the raw data. The blue line is the smoothed time series, the black thin line is the raw time series, and the points are the start and end of seasons detected by the TIMESAT´s algorithm to extract the phenological attributes.

Figure 3 shows the Smoothed EVI time series with the different smoothing algorithms using the filtered time series and the points of start and end of seasons detected by the TIMESAT´s algorithm to extract the phenological attributes.

Table 2. Accuracy assessment statistics including global accuracy and (kappa statistic) for classifications, considering the different smoothing algorithms using the EVI filtered time series, and the producer's accuracy of each class.

	Producer´s Accuracy (%)					
Smoothing methods	*Annual Agriculture*	*Natural Forest*	*Perennial Agriculture*	*Semi-perennial Agriculture*	*Grassland*	***GA% (kappa)***
DL	86.67%	68.75%	76.00%	69.23%	68.75%	76.00% (0.69)
AG	90.32%	71.43%	78.57%	61.54%	92.86%	81.00% (0.75)
SG	78.79%	57.14%	64.52%	72.73%	72.73%	70.00% (0.61)

ISBN: 978-85-17-00088-1

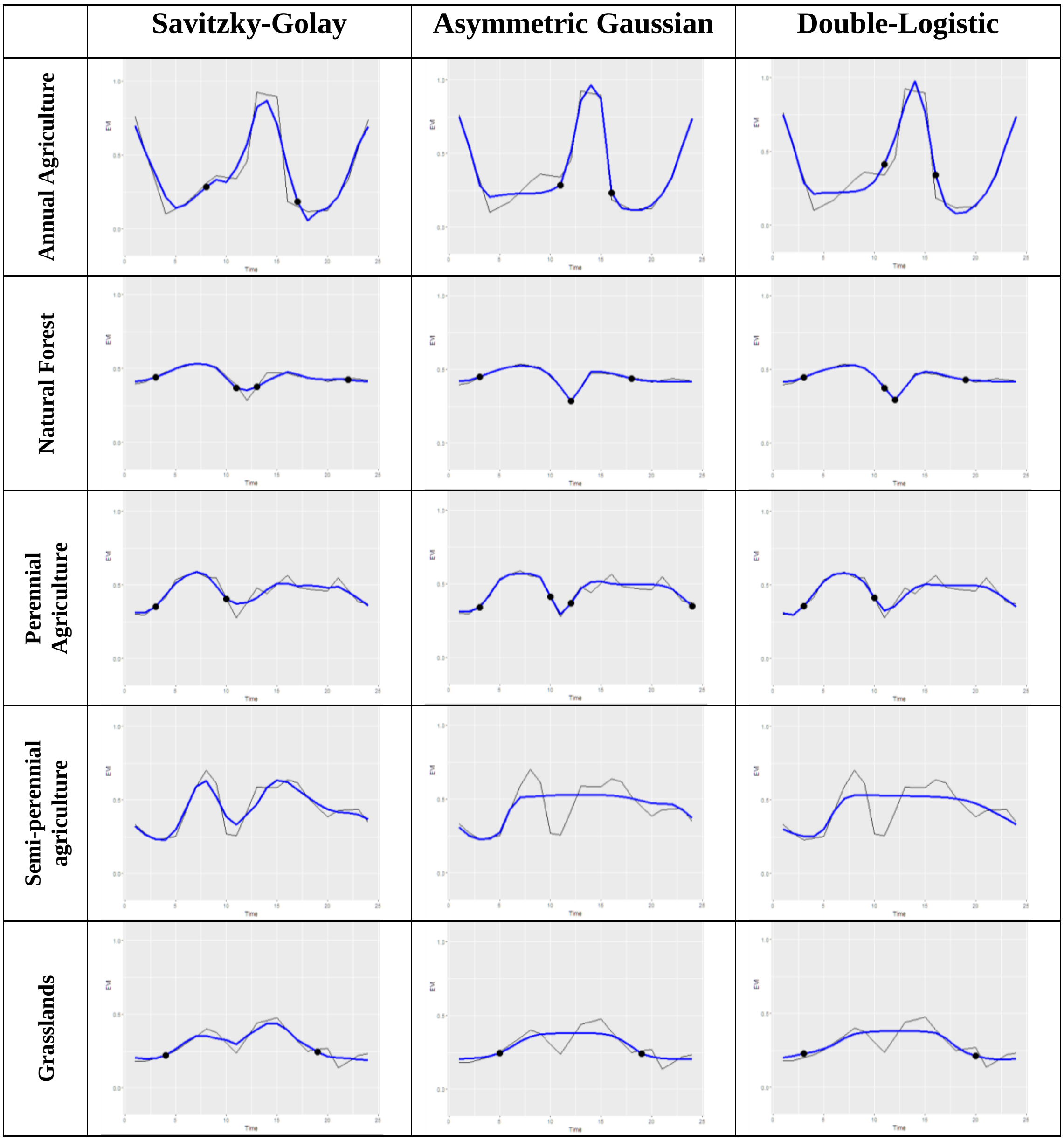

Figure 3. Smoothed EVI time series with the different smoothing algorithms using the filtered EVI time series. The blue line is the smoothed time series, the black thin line is the raw time series, and the points are the start and end of seasons detected by the TIMESAT´s algorithm to extract the phenological attributes.

The best classification result was found using the Asymmetric Gaussian smoothed data, with an overall accuracy of 81% (kappa of 0.75), as can be seen in Table 2. This results are followed by 76% (kappa of 0.69) with Double-logistic method, and 70% (kappa of 0.61) using the Savitzky-Golay function. In respect to per-class classification results, we can see that the only class that had lower producer´s accuracy using the Asymmetric Gaussian smoothed time series it was the Semi-perennial agriculture class, with 61.54%. Most misclassification happened between perennial crop and annual crop, and between Semi-perennial agriculture and Natural Forest class.

Besides, we can observe at Figures 2 and 3, that when the filtering approach was applied before the smoothing, the model fits better the time series. However, contrary to what we were expecting, this probably have no influence on the improvement of the classification results.

In the classification results, considering the different smoothing algorithms using the EVI raw time series (Table 1), we observed that the lowest producer's accuracy was also obtained for Semi-perennial agriculture. This could be related to the fact that we used only one year EVI time series for the phenological parameters detection, while the complete cycle of the semi-perennial crops, such as sugarcane, lasted more than a year. This might explain the confusion between the Semi-perennial agriculture and both Natural Forest and Perennial crops classes.

We calculated the mean standard deviation considering all the pixel time-series within each class in order to measure the intra-class variability. Figures 4 and 5 present the within-class mean standard deviation for the different classes, considering data from different smoothing algorithms using the raw and the filtered EVI time series.

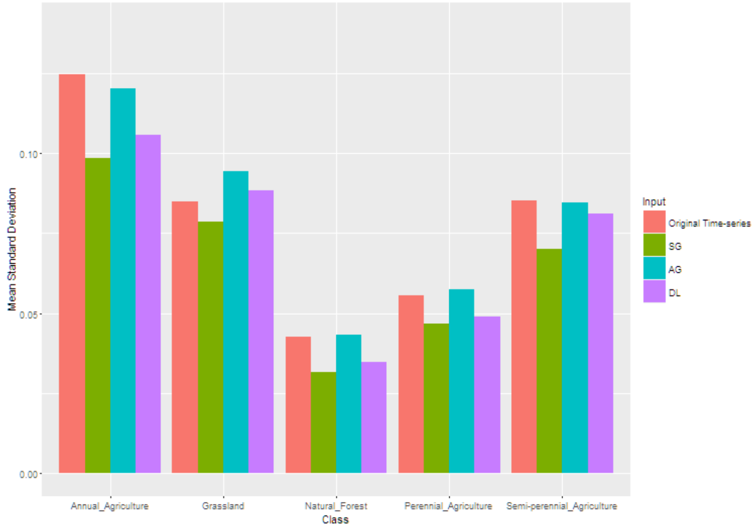

Figure 4. Within-class mean standard deviation for the different classes using the different smoothing algorithms with the raw EVI time series.

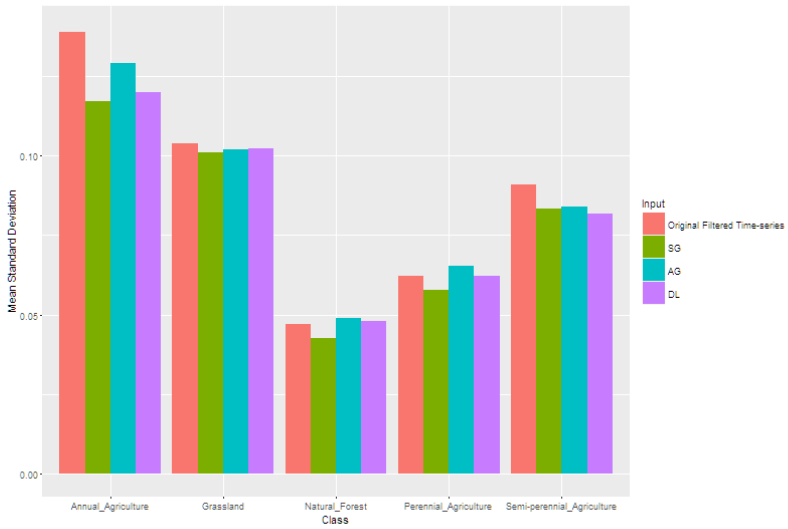

Figure 5. Within-class mean standard deviation for the different classes using the different smoothing algorithms with the filtered EVI time series.

We can observe in Figure 4 and 5 that the EVI time-series previously filtered presents bigger mean standard deviations within-class for all the classes for both smoothing methods, for example, for the Annual Agriculture class, the mean standard deviation measures increases from 0.124 to 0.139, 0.098 to 0.117, 0.120 to 0.129 and 0.106 to 0.120 for the original time-series, SG, AG and DL, respectively. Among the three smoothing algorithms, SG algorithm generated the smallest mean standard deviation for most classes. Similar results were observed by Shao et al. (2016) using MODIS data. Considering the raw EVI time-series, the smaller the mean standard deviation within-class the higher the producer´s accuracy, except for the Semi-perennial Agriculture class. But this relation cannot be observed in the previously filtered time-series.

4. Final considerations

In this work we evaluated three different time series smoothing methods, Savitzky-Golay, asymmetric Gaussian function and Double-logistic function, combined or not with filtering techniques. Besides, we evaluated the impact of their use in the agricultural use classification, in a region of the Brazilian Cerrado, using phenological parameters extracted from one year Enhanced Vegetation Index (EVI) Landsat-8 image time series. The smoothing method that provided the highest classification accuracy was the Savitzky-Golay applied to the raw time series (86% and kappa=0.82), followed by the asymmetric Gaussian applied to the filtered time series (81% and kappa=0.754). Further analyses are needed to evaluate this approach for large areas and also considering a bigger time span. We also suggest that an evaluation of inter-class separability could help to better understand the results.

Acknowledgments

This work was partially supported by FAPESP e-science program (grant 2014-08398-6) and by the Science without Borders Program of CNPq/CAPES (project 402597/2012-5).

References

Atzberger, C.; Eilers, P. H. 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, 2011.

Arvor, D.; Jonathan, M.; Meirelles, M. S. O. P.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time-series for crop mapping in the state of Mato Grosso, Brazil. **International Journal of Remote Sensing**, v. 32, n. 22, pp. 7847 – 7871, 2011.

Bendini, H. N.; Sanches, I. D.; Körting, T. S.; Fonseca, L. M. G.; Luiz, A. J. B.; Formaggio, A. R. Using Landsat 8 Image Time Series For Crop Mapping In A Region Of Cerrado, Brazil, **Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci**., XLI-B8, pp. 845 – 850, 2016a.

Bendini, H. N.; Fonseca, L. M. G.; Körting, T. S.; Sanches, I. D.; Marujo, R. F. B.; Arcanjo, J. S. Assessment of a Multi-Sensor Approach for Noise Removal on Landsat-8 OLI Time Series Using CBERS-4 MUX Data to Improve Crop Classification Based on Phenological Features. In: Brazilian Symposium on GeoInformatics (GEOINFO), 17., 2016, Campos do Jordão. **Proceedings**... Campos do Jordão: INPE, 2016b. pp. (in press).

Borges, E.F.; Sano, E. E. Séries temporais de EVI do MODIS para o mapeamento de uso e cobertura vegetal do oeste da Bahia. **Boletim de CiênciasGeodésicas**, v. 20, n. 3, pp. 526 – 547, 2014.

Breiman, L. Random Forests. **Machine Learning**, v. 45, pp. 5 – 32, 2001.

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, pp. 332 – 344, 2004.

DeVries, B.; Verbesselt, J.; Kooistra, L.; Herold, M. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. **Remote Sensing of Environment**, V. 161, pp. 107 – 121, 2015a.

DeVries, B.; Decuyper, M.; Verbesselt, J.; Zeileis, A.; Herold, M.; Joseph, S. Tracking disturbance-regrowth dynamics in tropical forests using structural change detection and Landsat time series. **Remote Sensing of Environment**, v. 169, pp. 320 – 334, 2015b.

Hall, M. A.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, pp. 10 – 18, 2009.

Hamunyela, E.; Verbesselt, J.; Roerink, G.; Herold, M. Trends in spring phenology of Western European deciduous forests. **Remote Sensing**, v. 5, n. 12, pp. 6159 – 6179, 2013.

Jönsson, P.; Eklundh, L. TIMESAT – a program for analyzing time-series of satellite sensor data. **Computers and Geosciences**, v. 30, n. 8, pp. 833 – 845, 2004.

Jönsson, P.; Eklundh, L. **TIMESAT 3.2 with Parallel Processing Software Manual**. Lund University, Sweden. pp. 22 – 24, 2015.

Müller, H.; Rufin, P.; Griffiths, P.; Siqueira, A. J. B.; Hostert, P. Mining dense Landsat time-series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. **Remote Sensing of Environment**, v. 156, pp. 490 – 499, 2015.

Peña, M.A.; Brenning, A. Assessing fruit-tree crop classification from Landsat-8 time-series for the Maipo Valley, Chile. **Remote Sensing of Environment**, v. 171, pp. 234 – 244, 2015.

Rodriguez-Galiano, V. F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 67, pp. 93 – 104, 2012.

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, pp. 258 – 265, 2016.

U.S. Geological Survey. **PRODUCT GUIDE: Landsat Surface Reflectance products courtesy of the U.S. Geological Survey**, pp. 1 – 27, 2015.

Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. **Remote Sensing of Environment**, v. 185, pp. 46 – 56, 2016.

Verbesselt, J.; Hyndman, R.; Zeileis, A.; Culvenor, D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. **Remote Sensing of Environment**, 114 (2010), pp. 2970–2980

Witten, I. H.; Frank, E.; Hall, M. A. **Data mining: practical machine learning tools and techniques**. 3ed. San Francisco: Morgan Kaufmann, 2011.

Zhang, X.; Friedl, M. A.; Schaaf, C. B. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, v. 84, pp. 471 – 475, 2003.

Zheng, B.; Myint, S. W. Thenkabail, P. S.; Aggarwal, R. M. A support vector machine to identify irrigated crop types usingtime-series Landsat NDVI data. **International Journal of Applied Earth Observation and Geoinformation**, v. 34, pp. 103 – 112, 2015.