USING LANDSAT 8 IMAGE TIME SERIES FOR CROP MAPPING IN A REGION OF CERRADO, BRAZIL

H. do N. Bendini [a], I. D. Sanches [a], T. S. Körting [b], L. M. G. Fonseca [b], A. J. B. Luiz [c], A. R. Formaggio [a]

[a] DSR/INPE, Remote Sensing Division of the National Institute for Space Research – (hbendini, formag)@dsr.inpe.br, ieda.sanches@inpe.br
[b] DPI/INPE, Image Processing Division of the National Institute for Space Research – tkorting@dpi.inpe.br, leila.fonseca@inpe.br
[c] Embrapa Environment, Brazilian Agricultural Research Corporation- alfredo.luiz@embrapa.br

Commission VI, WG VI/4

KEY WORDS: Crop mapping, Landsat Time series, Random Forest algorithm

ABSTRACT:

The objective of this research is to classify agricultural land use in a region of the Cerrado (Brazilian Savanna) biome using a time series of Enhanced Vegetation Index (EVI) from Landsat 8 OLI. Phenological metrics extracted from EVI time series, a Random Forest algorithm and data mining techniques are used in the process of classification. The area of study is a region in the Cerrado in a region of the municipality of Casa Branca, São Paulo state, Brazil. The results are encouraging and demonstrate the potential of phenological parameters obtained from time series of OLI vegetation indices for agricultural land use classification.

1. INTRODUCTION

1.1 Overview

Agriculture has significant participation in the Brazilian economy; it is the main responsible for the positive trade balance of the country. Given the high availability of arable land, and taking into account the growing demand for food in the world, Brazil has been consolidate as a big player in the world agricultural scenario. On the other hand, sustainable agriculture expansion is a key concern in Brazilian Cerrado biome, which suffers a great land conversion pressure. In this way, crop mapping is strategic to estimate acreage and production, as well as to better understanding the distribution of croplands, and its impact on the environment. In this context, remote sensing is an important tool due to its ability to generate information on large scale in a cost-effective way. With advances in data processing and storage technologies, and the availability of long-term image series, remote sensing is undergoing a paradigm shift, in which time series techniques allow to consider seasonal variations of the analysed target. This approach is useful for vegetation studies, especially in agricultural areas, since vegetation cover is quite dynamic in time, and the ability to capture these variations is essential to discriminate different types of crops, through its phenological characteristics. For this purpose, most studies have explored time series of vegetation indices like NDVI or EVI from MODIS data (Jakubauskas et al., 2002; Sakamoto et al., 2005; Wardlow et al., 2007; Esquerdo et al., 2011; Arvor et al., 2011; Körting, 2012; Risso et al., 2012; Coutinho et al., 2013; Borges & Sano, 2014; Tomás et al., 2015; Neves et al., 2016). However, there is a demand to produce more detailed maps, which can be obtained from higher spatial resolution satellites such as Landsat-like (Zheng et al., 2015; Peña et al., 2015; Pan et al., 2015). Therefore, the objective of this study is to employ phenological metrics obtained from time series of Landsat 8/OLI vegetation indices to classify agricultural land use in the municipality of Casa Branca, located in the Cerrado (Brazilian

Savanna) biome in the state of São Paulo, Southeast region from Brazil.

1.2 Study area

We conducted our study in the south of Casa Branca municipality, in state of São Paulo, Brazil (Figure 1). Such region is located in an overlapping area of two adjacent Landsat Worldwide Reference System 2 (WRS – 2) scenes (path/row 219/75 and 220/75), providing a temporal resolution of 8 days (Luiz et al., 2015). Casa Branca city has a tropical wet and dry climate (Aw, according to the Köppen Climate Classification System) with average annual temperature of 21.5° C and a seasonal rainfall pattern with most rainfall occurring from October to March. The average annual precipitation is 109.18 mm (CEPAGRI/UNICAMP, 2016).
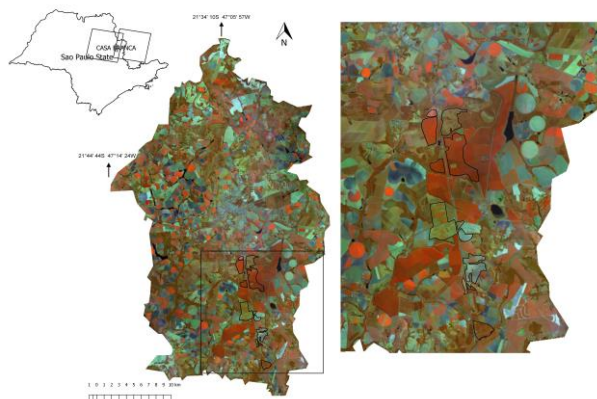

Figure 1. False color (bands 5, 6 and 4 in red, green and blue respectively) OLI Landsat imagery of the study area

In this region, farmers grow a variety of crops along all the year. Major field crops in this area are sugarcane, corn, bean, potato, soybean, peanuts, sorghum and cassava. There are also a

significant production of citrus, mango, coffee, rubber and eucalyptus (IBGE, 2016). Farmers grow crops in double cropping systems and even in triple cropping systems, mainly on the irrigated areas. The usual planting summer crop dates is from October to December and harvesting from February to April, but planting crops in late fall (May – July) and harvest in the next spring are also observed, especially in the irrigated areas.

2. MATERIALS AND METHODS

2.1 Remotely sensed data

A total of 46 scenes of Landsat 8 OLI (Path 219 and 220/ Row 75) between July 2014 (Day of year – DOY 206) and July 2015 (DOY 209) were obtained at the US Geological Survey platform (USGS) Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). These data are provided with geometric correction level 1 (L1T) and converted to surface reflectance by the algorithm of atmospheric correction LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing) (Masek et al., 2006). We then generate EVI layers for each image in according to this equation (Huete et al., 2002):

$$EVI = \frac{NIR - R}{NIR + (6*R - 7.5*B) + 1} , \quad (1)$$

where NIR = near-infrared reflectance band (OLI band 5)
R = red reflectance band (OLI band 4)
B = blue reflectance band (OLI band 2).

EVI was chosen because its capacity for highlighting the spectral response related to green vegetation canopy, taking into account the background and atmospheric effects.

2.2 Removal of outliers and null values and filtering

No cloud free images were identified in November, February and March. Therefore, we applied the following simple linear interpolation to remove outliers and null values:

$$x_t = \frac{x_{t-1} + x_{t+1}}{2} \quad \left\{ if \ x_t < 0.01 \right\}, \quad (2)$$

where x_t = an observation of the time series at time t
x_{t-1} = an observation at time t-1
x_{t+1} = at time t+1.

x_t observation is replaced by the average of x_{t-1} and x_{t+1} if x_t is less than 0.01. This method, however, is not capable of removing consecutive outliers. Afterwards, the time series were smoothed through the double logistic filter (Zhang et al., 2003; Jönsson; Eklundh, 2004). This function is recommended for smoothing image time series on cropland areas in the Brazilian Cerrado (Borges & Sano, 2014).

2.3 Seasonal data extraction

Phenological metrics extracted from EVI time series were obtained by the TIMESAT software (3.2 version) (Jönsson; Eklundh, 2004) (web.nateko.lu.se/timesat/), in which seasonal data were extracted for each of growing seasons of the central year (Figure 2). During a period of n years there may be $n – 1$

full seasons together with two fractions of a season in the beginning and end of the time series. Then, to extract seasonality parameters from one year of data the time series was duplicated to span three years, as recommended by Jönsson and Eklundh (2015). Figure 2 shows the schema of the seasonality parameters generated by TIMESAT. In this work we assume that the seasonality parameters are the same of phenological metrics. The time for the beginning of season (a), or start of the season (sos), and the end of season (eos) (b) is the time for which the left and right edge, respectively, has increased to a defined level (often a certain fraction of the seasonal amplitude) measured from the minimum level on the corresponding side. The length of the season (c) is the time from the start to the end of the season. Base value (d) is given as the average of the left and right minimum values. The middle of season (e) is computed as the mean value of the times for which, respectively, the left edge has increased to the 80 % level and the right edge has decreased to the 80 % level.
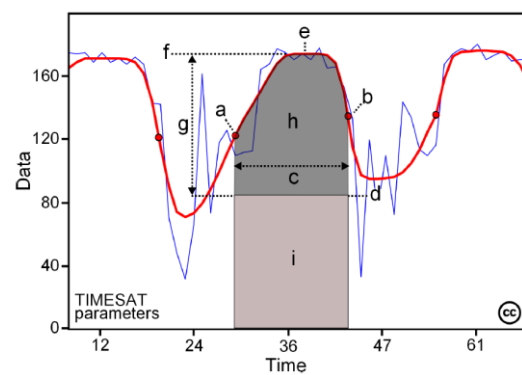

Figure 2. Examples of seasonality parameters generated by TIMESAT: (a) beginning of season, (b) end of season, (c) length of season, (d) base value, (e) time of middle of season, (f) maximum value, (g) amplitude, (h) small integrated value, (h+i) large integrated value. The red line represents the filtered data, and blue the original data

The maximum value (f), or the peak of phenological cycle, is the largest data value for the fitted function during the season. The seasonal amplitude (g) is the difference between the maximum value and the base level. The left derivative is calculated as the ratio of the difference between the left 20% and 80% levels and the corresponding time difference. The right derivative (i.e. the rate of decrease at the end of the season) is the absolute value of the ratio of the difference between the right 20% and 80% levels and the corresponding time difference. The rate of decrease is thus given as a positive quantity. Large seasonal integral (h+i) is integral of the function describing the season from the season start to the season end. And the small seasonal integral (h) is the integral of the difference between the function describing the season and the base level from season start to season end (Jönsson and Eklundh, 2015). For more details see Jönsson and Eklundh (2002; 2004).

2.4 Training and validation data

Field campaigns were carried out for collecting training samples, totalling 16,243 pixels. The classes considered were Annual Crop (potato and corn, on a double crop system), Semi-Perennial Crop (sugarcane and cassava), Perennial Crop (citrus,

mango and rubber) and Planted Forestry (eucalyptus). The polygons that represent each class were directly digitized over Landsat 8 images.

2.5 Random Forest classification

The variables obtained by TIMESAT were subjected to data mining using the Random Forest classification (Breiman, 2001). In this classification, data set is randomly divided in several subsets of smaller size by applying bootstrap (Han et al., 2011), and for each subset a decision tree is built. All trees contribute to the classification of each object under study, by voting on which class the target attribute must belong. Random Forest algorithm has been widely used in remote sensing applications (Clark et al, 2010; Müller et al, 2015; Peña et al, 2015) due to its advantages in efficiently handle large databases, providing estimates on most relevant variables, and allowing the identification of outliers (Rodriguez-Galiano et al., 2012).

3. RESULTS AND DISCUSSION

3.1 Crop phenological curves and TIMESAT features

Initially, we evaluated the time series spectral profiles of different types of targets. The phenological features were characterized by median values observed in all samples for each class. Figures 3 to 8 shows the time series spectral profiles for all targets and the results based on two approaches for noise removal, in which dotted line represents the time series outliers and null values removed (Equation 2) and the thick line represents the double logistic filtered time series. The samples of annual crop were located in two different farms. However, both were on pivot-center irrigation systems. We observed during fieldwork that in one of them there was a triple crop system (potato, bean and corn), where potatoes were harvested between July and mid-August. This was followed by corn planting in September, which was harvested between January and February. We also observed a short-cycle bean between February and March, followed by corn planting from May. The majority of time series samples of this farm, only one season was modelled by TIMESAT, corresponding to corn planting in late September (DOY 265) and harvested in early February (DOY 41). The middle of season was observed in mid-November (DOY 325) (Figure 3).
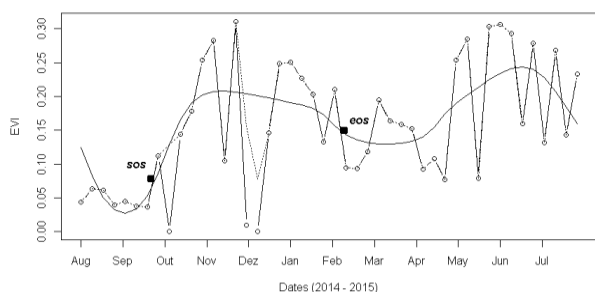

Figure 3. Time series EVI spectral profile for an annual crop sample located into a triple cropping system of potato, bean and corn. Dotted line represents the time series outliers and null values removed and the thick line represents the double logistic filtered time series.

It was not possible to identify potatoes seasons because it was not completely covered by the time series span. It was also not possible to capture the bean season, as well the late fall corn

season. We hypothesize that this was due to the short cycle of bean and mainly for the high intensity of cloud contaminated pixels on this period. Although potatoes were harvested later in the other farm, it was not possible to detect their season. Only the corn season was detected, starting between late November and mid-December (DOY 347) and the end of season it was on mid-April (DOY 102). The middle of season was observed in late January (DOY 26). On the semi-perennial crop samples corresponding to cassava farms it was observed that in 77% of pixels were detected only one season. The start of season was in late September (DOY 270) and the middle of season was detected in mid-February (DOY 45). Based on field observations, we hypothesize that this season can be associated to the end of the cassava senescence, characterized by the start of a new greenness period. The end of season coincided with the end of harvest, when the area becomes completely covered by weeds. In the other 23% of pixels it was possible to detect two seasons, where the second season began between early December (DOY 347) and early April (DOY 60) ending on late June (DOY 179). This indicates a predominant phenological behavior of weed, which has a great biomass gain after a rainfall period (Figure 4).
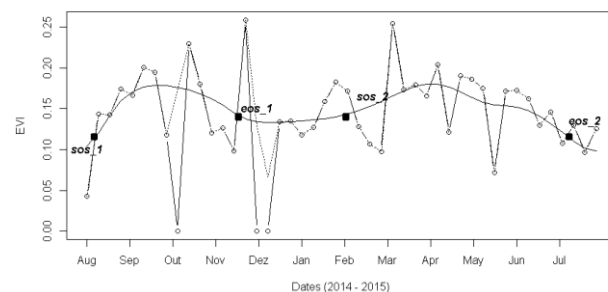

Figure 4. Time series EVI spectral profile for a semi-perennial crop sample of cassava.

We observed that sugarcane (semi-perennial crop class) was planted in June. However, the complete ground cover occurred only in November, when it is very difficult to have cloud-free images. Therefore, the start of season was identified in early December (DOY 338) and the end of season in mid-May (DOY 135), corresponding to the harvest (Figure 5).
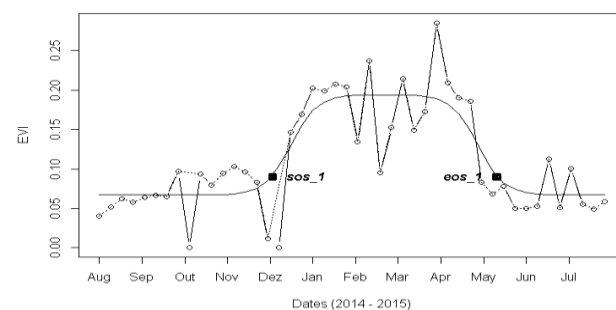

Figure 5. Time series EVI spectral profile for a semi-perennial crop sample of sugarcane.

Figure 6 shows the time series spectral profile of a perennial crop sample, corresponding to a rubber tree farm. We can observe two season's detection. In the first one, the start of season was in the middle of August (DOY 229) and the end in December (DOY 352). The second season began between December and February (DOY 29) and the end of season occurred in late July (DOY 190). This occurred in 61% of the

pixels inside this farm. This behavior was not expected for rubber trees, considered a perennial crop. However, this can be explained by the high frequency of cloud-contaminated images between November and February. This noise could not be corrected by the smoothing and noise removal techniques employed in this current study. Additionally, the rubber trees phenology in this region has a senescence period between July and August, followed by a greenness period after September. The 49% remaining pixels located in this farm produces time series with a single season between August (sos = DOY 231) and late July (eos = DOY 173).
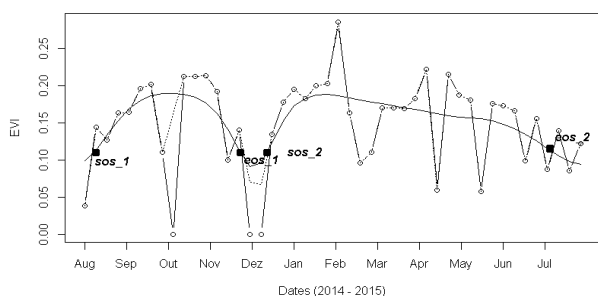

Figure 6. Time series EVI spectral profile for a perennial crop sample of rubber trees.

Time series of pixels located in one of the two citrus farms that represent the perennial crops class presents a subtle difference on phenological features. TIMESAT detected one short season starting in early December (DOY 343) and ending in mid-March (DOY 73). The other citrus farm, as well in the mango farm, the model detected a longer season, between December (DOY 338) and June (DOY 171). Among the other differences it was also observed bigger amplitudes in the first citrus farm (median = 0.073 EVI) than in the other perennial crops farms (median = 0.059 and 0.055 for citrus and mango, respectively). Furthermore, the middle of season in the first citrus farm occurred in January (DOY 18) and the EVI peak found corresponded to 0.22 EVI, while in the other areas the middle occurred in March (DOY 61 and 85, for citrus and mango), with a 0.18 EVI peak.
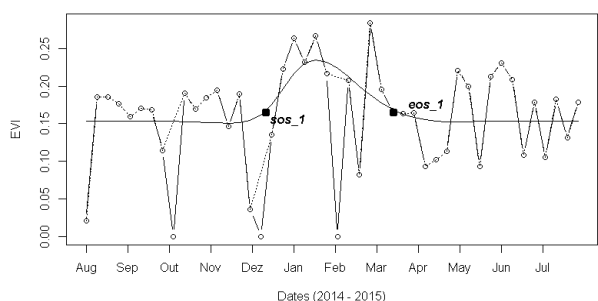

Figure 7. Time series EVI spectral profile for perennial crop sample of one of the citrus farm (similar to mango).

Finally, in the eucalyptus planted forests, around 60% of pixels presented time series in which it was not possible to detect any season, so the values of phenological features were zero. In the other 40% pixels time series TIMESAT detected one season that began on early February (DOY 41) and finished on July (DOY 185), as showed in Figure 8.
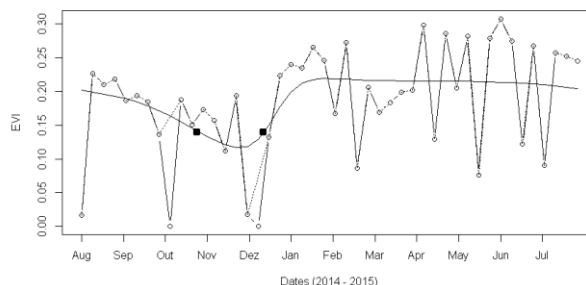

Figure 8. Time series EVI spectral profile for a planted forest sample of eucalyptus.

3.1 Classification performance

A 10 fold cross-validation technique was applied in the training set and 15,499 of the 16,243 pixels were classified correctly, resulting in an accuracy of 95.42% and Kappa 0.9284. The confusion matrix with the errors of commission and omission is presented in Table 1.

		Reference					
		a	**b**	**c**	**d**	**Total**	**EC%**
Classification	**a**	**6563**	40	184	127	6914	5.1
	b	3	**717**	7	10	737	2.7
	c	265	17	**6303**	40	6625	4.9
	d	26	21	4	**1916**	1967	2.6
	Total	6857	795	6498	2093	16243	
	EO%	4.3	9.8	3.0	8.5		

Table 1. Confusion matrix: (a) Perennial crop, (b) Annual crop, (c) Planted Forest and (d) Semi-perennial crop. EC = errors of commission; EO = errors of omission

Most misclassification happened between perennial crop and annual crop, and between perennial crop and semi-perennial crop, due to the rubber trees spectral similarity with the two other classes. The majority of omission errors were observed on annual crop and semi-perennial crop classes. We noticed that 5.03% of pixels in the annual crop class were omitted, being wrongly classified as perennial crop. As we discussed earlier, a significant number of samples of annual crop class presented time-series where only one season was detected (between December and March) with a 0.22 EVI peak. This is similar to some annual crop class pixels, what probably contributed to the omission of these samples. With respect to the semi-perennial crop class, major confusion was observed on the perennial crop, in which around 6.1% of pixels were wrongly classified. The majority of commission errors was founded for the perennial crop (5.1%) and planted forest (4.9%) classes. An inclusion of 2.6% of pixels of the planted forest and 1.84% of semi-perennial classes was also observed in the perennial crop class. Around 4% of pixels of perennial agriculture were included in the planted forest class. However, we considered that the explored features configured a good separability among the classes. Figure 9 shows the 3D-scatterplot with length of season, base value and left derivative features of the first season. The left derivative is the rate of increase at the beginning of the season. Higher values of left derivative indicate that there is a fast rise of the greenness.
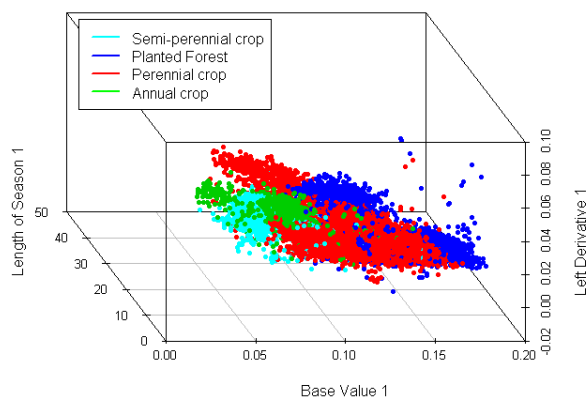
Figure 9. 3D-scatterplot of length of season, base value and left derivative features of the first seasons. The zero values was removed for best visualization

We can see in Figure 9 that lower values of length of season and of base value were found for annual crop samples. High values of left derivative were also observed for this class (> 0.035 EVI). The low rates of base value (< 0.08) indicate that the EVI values before and right after the season were also low. This is expected for cropland areas. Besides the omitted zero values of planted forest class samples, there are medium and high base values (0.13 − 0.17 EVI), medium values for length of season and left derivative (0.013 − 0.029 EVI). Pixels of the semi-perennial class presented low rates of base value and left derivative, and medium values for length of season. In the perennial crop class it is possible to visualize two clusters on the 3D-scatterplot, determined by length of season values. Both left derivative and base values were medium. By investigating the errors spatial distribution, it also seems that it may be associated with heterogeneity of the monitored areas that results in spectral mixing. In addition, for the extraction of the phenological parameters using TIMESAT, a generic model was considered for all classes, and the small number of samples may also have contributed to errors. The definition of time series span is also a very important criteria that must be considered.

3.1 Final considerations

These preliminary results are encouraging and demonstrate the potential of phenological parameters obtained from time series of OLI vegetation indices for agricultural land use classification. We show the potential use of higher temporal resolution Landsat-like images for crop mapping, what will soon become reality, once Landsat, Sentinel-2A and CBERS 4 data can be combined to generate consistent time series to produce land use maps. Further analysis are needed to apply this approach in large areas and to test different time span, different vegetation indices and the complete spectral resolution of Landsat-like image time series.

REFERENCES

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

Borges, E.F., Sano, E. E., 2014. Séries temporais de EVI do MODIS para o mapeamento de uso e cobertura vegetal do oeste da Bahia. *Boletim de Ciências Geodésicas*, 20 (3), pp. 526-547.

Breiman, L., 20*01*. Random Forests. *Machine Learning*. Vol 45, pp. 5-32.

CEPAGRI/UNICAMP, 2016. http://www.cpa.unicamp.br/ (accessed 15 March 2016)

Clark, M. L., Aide, T. M., Grau, H. R., Riner, G., 2010. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: a case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment*, 114 (1), pp. 2816-2832

Coutinho, A. C., Esquerdo, J. C. D. M., Oliveira, L. S., Lanza, D. A., 2013. Methodology for Systematical Mapping of Annual Crops in Mato Grosso do Sul State (Brazil). *Geografia*, 8 (special number), pp. 45-54.

Esquerdo, J. C. D. M., Zullo Júnior, J., Antunes, J. F. G., 2011. Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. *International Journal of Remote Sensing*, 32 (13), pp. 3711-3727.

Huete, A. R., Liu, H. Q., Batchily, K., Leeuwen, W. van., 2002. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sensing of Environment*, 59 (3), pp. 440-451.

IBGE, 2016. http://www.sidra.ibge.gov.br/ (accessed 15 March 2016)

Jakubauskas, M. E., Legates, D. R., Kastens, J. H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, 37 (1-3), pp. 127-139.

Jönsson, P., Eklundh, L. 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*. 40 (8), pp. 1824-1831.

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

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

Korting, T. S., 2012. *Geodma: A Toolbox Integrating Data Mining with Object-Based and Multi-Temporal Analysis of Satellite Remotely Sensed Imagery*. PhD Thesis, National Institute for Space Research (INPE). 97p.

Luiz, A. J. B., Sanches, I. D., Trabaquini, K., Eberhardt, I. D. R., Formaggio, A. R., 2015. Dinâmica agrícola em área de sobreposição de órbitas adjacentes dos satélites Landsat. In: *Anais do Simpósio Brasileiro De Sensoriamento Remoto*, João Pessoa, Brazil, Vol. XVII, pp. 1308-1315.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., 2006. A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, 3 (1), pp. 68-72.

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

Neves, A. K., Bendini, H. N., Körting, T. S., Fonseca, L. M. G., (in press) 2016. Combining Time Series Features and Data Mining to Detect Land Cover Patterns: A Case Study in Northern Mato Grosso State, Brazil. *Revista Brasileira de Cartografia.*

Pan, Z.; Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G. A.; Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1A/B data. *International Journal of Applied Earth Observation Geoinformation*, 34, pp. 188-197.

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

Risso, J., Rudorff, B. F. T., Adami, M., Aguiar, A. P. D., Freitas, R. M, 2012. MODIS Time Series for Land Use Change Detection in Fields of the Amazon Soy Moratorium. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. 23, pp. 339-344.

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

Sakamoto, T., Yokozawa, M., Toritani, T., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, 96 (3-4), pp. 366-374.

Tomás, J. C., Faria, F. A., Esquerdo, J. C. D. M., Coutinho, A. C., Medeiros, C. B. 2015. SiRCub – Brazilian Agricultural Crop Recognition System. In: *Anais do Simpósio Brasileiro De Sensoriamento Remoto*, João Pessoa, Brazil, Vol. XVII, pp. 6273-6280.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250m vegetation index data for crop classification in the US central Great Plains. *Remote Sensing of Environment*, 108 (3), pp. 290-310.

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

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