Annual cropland mapping using data mining and OLI Landsat-8

Lucas V. Oldoni[1], Carlos E. V. Cattani[2], Erivelto Mercante[2], Jerry A. Johann[2], João F. G. Antunes[3] & Luiz Almeida[1]

[1] Instituto Nacional de Pesquisas Espaciais/Divisão de Sensoriamento Remoto. São José dos Campos, SP, Brasil. E-mail: lucasoldoni@outlook.com (Corresponding author) - ORCID: 0000-0002-4434-3465; almeidalz@hotmail.com - ORCID: 0000-0002-4286-7900

[2] Universidade Estadual do Oeste do Paraná/Pós-Graduação em Engenharia Agrícola/Centro de Ciências Exatas e Tecnológicas. Cascavel, PR, Brasil. E-mail: carloscattani@outlook.com.br - ORCID: 0000-0001-9273-3823; erivelto.mercante@unioeste.br - ORCID: 0000-0002-0744-5286; jerry.johann@hotmail.com - ORCID: 0000-0001-6184-8011

[3] Empresa Brasileira de Pesquisa Agropecuária/Embrapa Informática e Agropecuária. Campinas, SP, Brasil. E-mail: joaoantunes@embrapa.com.br - ORCID: 0000-0001-8114-9971

ABSTRACT: In the state of Paraná, Brazil, there are no major changes in areas cultivated with annual crops, mainly due to environmental laws that do not allow expansions to new areas. There is a great contribution of the annual crops to the domestic demand of food and economic demand in the exports. Thus, the area and distribution of annual crops are information of great importance. New methodologies, such as data mining, are being tested with the objective of analyzing and improving their potential use for classification of land use and land cover. This study used the classifiers decision tree and random forest with Normalized Difference Vegetation Index (NDVI) temporal metrics on images from Operational Land Imager (OLI)/Landsat-8. The results were compared with traditional methods spectral images and Maximum Likelihood Classifier (MLC). At first, seven classes were mapped (water bodies, sugarcane, urban area, annual crops, forest, pasture and reforestation areas); then, only two classes were considered (annual crops and other targets). When classifying the seven targets, both methods had corresponding results, showing global accuracy near 84%. NDVI temporal metrics showed producer's and user's accuracy for the annual crop class of 86 and 100%, respectively. However, if considering only two classes, the NDVI temporal metrics reached global accuracy of near 98% and producer's and user's accuracy above 94%.

Key words: decision tree, random forest, NDVI temporal metrics

Mapeamento de áreas agrícolas anuais utilizando mineração de dados e séries temporais do OLI/Landsat-8

RESUMO: No Estado do Paraná, Brasil, não há grandes mudanças nas áreas cultivadas com culturas anuais, principalmente devido a leis ambientais que não permitem expansões para novas áreas. Há grande contribuição das culturas anuais para a demanda doméstica de alimentos e econômica nas exportações. Assim, a área e distribuição das culturas anuais são informações de grande importância. Novas metodologias, como data mining, estão sendo testadas com o objetivo de analisar e melhorar seu potencial de uso para classificação do uso e cobertura da terra. Neste estudo, foram utilizados os classificadores decision tree e random forest com métricas temporais de Normalized Difference Vegetation Index (NDVI) em imagens do Operational Land Imager (OLI)/ Landsat-8. Os resultados foram comparados com os métodos tradicionais (imagens espectrais e classificador Maximum Likelihood Classifier - MLC). Inicialmente, foram mapeadas sete classes (corpos d'água, cana-de-açúcar, área urbana, culturas anuais, floresta, pastagem e áreas de reflorestamento) e posteriormente apenas duas classes foram consideradas (culturas anuais e outras classes). Ao classificar os sete alvos, ambos os métodos tiveram resultados correspondentes, mostrando exatidão global próxima a 84%. As métricas temporais de NDVI mostraram a acurácia do produtor e do usuário para a classe de cultura de 86 e 100%, respectivamente. No entanto, considerando-se apenas duas classes, as métricas temporais do NDVI alcançaram exatidão global próxima a 98% e a acurácia do produtor e do usuário acima de 94%.

Palavras-chave: árvore de decisão, random forest, métricas temporais de NDVI

INTRODUCTION

Remote sensing, given its synoptic character and data acquisition promptness, stands out as a technique able to monitor the crops throughout their lifecycle. Even though there are several orbital remote sensors with different configurations and resolutions (Toth & Jóźków, 2016), most of the current ones are unable to distinguish different agricultural crops in terms of spectral characteristics (Yao et al., 2015).

To overcome this issue, new approaches such as Data Mining (DM) have been tested to assess and improve spectral differentiation (Grande et al., 2016). DM approach has tools to analyze large amounts of data, allowing the development of a learning mechanism (Vintrou et al., 2013). Another procedure to assist in the multispectral classification of images is the multi-temporal analysis of Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974) since spectral-temporal profiles are strongly tied to agriculture dynamics (Cattani et al., 2017). This type of approach has been used to classify crop types (Chen et al., 2018) and land cover (Jia et al., 2014).

Among the orbital image classifiers, MLC (Maximum Likelihood Classifier) is one of the most used (Silva et al., 2013). Chen et al. (2018) used MLC to generate a crop/non-crop map on OLI/Landsat-8 images in the state of Mato Grosso, Brazil, with overall accuracy greater than 95% and producer's and user's accuracy over 90%. Jia et al. (2014) when classifying land cover in China obtained overall accuracy of up to 94.6% using the MLC; however, MLC may present limitations, such as incorrect identification of targets with similar spectral classes (Amaral et al., 2009).

Algorithms based on Machine Learning (ML) have been an alternative which achieved extremely efficient results in terms of agricultural target classifications (Valero et al., 2016). The majorly used algorithms are Decision Trees (DT) and Random Forests (RF), or even combinations of them (Lary et al., 2016).

Against this background, this study aimed to compare two orbital image classification approaches. One of them consisted of using data mining techniques to classify a NDVI time series data from OLI/Landsat-8 images. The other was to classify using only spectral information from four image dates.

MATERIAL AND METHODS

The study was conducted according to the steps of Knowledge Discovery in Databases (KDD) process (Fayyad et al., 1996), which is divided into five steps: 1) data selection, 2) preprocessing, 3) transformation, 4) data mining, and 5) interpretation.

The imagery was acquired from the Operational Land Imager (OLI) sensor, onboard the Landsat 8 satellite (WRS-2 Path: 223; WRS-2 Row: 077). This is a region of great agricultural output in the West of Paraná state (Brazil), mainly soybeans and corn crops (Souza et al., 2015).

The OLI sensor has nine spectral bands, 12-bit radiometric resolution, and 16-day revisit cycle (U.S. Geological Survey 2019). Landsat 8 images of high-level surface reflectance (Level 2) made available on demand by the USGS (https://earthexplorer.usgs.gov/) were downloaded. These images are processed by the Landsat Surface Reflectance Code (LaSRC) (Vermote et al., 2016). Eleven images with less than 3% of clouds at different days of year (DOY) were selected from 2015 (294), 2016 (73, 89, 121, 185, 265, 281, 297 and 329) and 2017 (27 and 59). Pixels with cloud and cloud shadow were eliminated in the selected images using the quality band (Zhu et al., 2015), which is distributed along with the Landsat 8 images of surface reflectance. Bands comprising the blue (0.452 - 0.512 µm, Band 2), green (0.533 - 0.590 µm, Band 3), red (0.636-0.673 µm, Band 4), NIR - Near Infrared (0.851 - 0.879 µm, Band 5), the Shortwave Infrared 1 - SWIR 1 (1.566-1.651 µm Band 6) and SWIR 2 (2.107-2.294 µm, Band 7) (USGS, 2019).

First, images were reprojected to the Universal Transverse Mercator (UTM) zone 22 South. Afterward, the NDVI was calculated as the ratio of the difference by the sum between the reflectance in the red and the NIR (Rouse et al., 1974). The NDVI is widely used in the agricultural monitoring and mapping since it exploits the vegetation contrast in relation to other targets.

The NDVI of annual agricultural crops range from values close to zero (beginning of lifecycle) to one - maximum vegetative development (flowering, fruiting and grain-filling); then, they decrease to values near zero again (senescence, remains and bare soil), being followed by a new annual crop cycle with the same trend (Cattani et al., 2017). There is little spectral-temporal variation in targets such as cities, reforestation areas and forests, which show mean NDVI values near 1.0 for reforestation and forest, and values close to 0.5 for urban areas. Yet sugarcane fields and pastures have lower spectral-temporal variations compared to other annual crops. As for the water, for lowly reflecting in near infrared, it has NDVI values near or below zero.

The NDVI differences ($NDVI_{SD}$) of the Landsat-8 images were summed (Eq. 1) to quantify the spectral-temporal variation of NDVI for annual crops, creating a new variable able to differentiate these surfaces from the other targets. The expression for the $NDVI_{SD}$ is

$$NDVI_{SD} = \sum_{i=1}^{n} \left| NDVI_i - NDVI_{i+1} \right| \qquad (1)$$

where:

$NDVI_{SD}$ - NDVI differences;
n   - the number of images of the temporal-series;
$NDVI_i$ - the i image from the temporal series; and,
$NDVI_{i+1}$ - the i+1 image from the temporal series.

Then, the mean, minimum, maximum, standard deviation, coefficient of variation, amplitude, median and sum were calculated for the NDVI time series. These measurements were used as input data for classification along with $NDVI_{SD}$ ones, which from here were called NDVI temporal metrics.

In preprocessing, a cube was created with the mentioned temporal metrics (NDVI cube -NC), and another with the spectral bands 2, 3, 4, 5, 6, 7 (Multispectral Cube - MC) from 03/07/2016, 21/09/2016, 24/11/2016 and 11/01/2017. This was used in the classifications for comparison with the NC. The false-color composition (RGB-564) was generated for sample collection.

Data mining was performed using the supervised classifiers Decision Tree (DT) and Random Forest (RF) in both image cubes (NC and MC). For comparison, a classification was performed using a Maximum Likelihood Classification algorithm (MLC) in MC.

The DT and RF classification algorithms used here derived from the python scikit-learn library (Pedregosa et al., 2011) for machine learning. This library uses an optimized version of the Classification and Regression Tree Algorithm (CART) (Breiman et al., 1984), which supports meta-variables, also allowing regression. RF is a method that combines k decision trees from the CART; it matches predictors from the trees in such a way that each of them depends on the values of a random vector sampled independently and with the same distribution for all the trees within a forest (Breiman, 2001).

A priori, the Overall Accuracy (OA), which is the percentage of correctly labeled pixels in a dataset, was assessed, in addition to the Kappa coefficient (K) (Cohen, 1960). Both were generated by the classification algorithms to verify the best used.

Accuracies of the produced maps were determined by error matrices. For that, a technique known as sample panel was used; it is characterized by a random distribution of sampling points within the area, with the purpose of surveying the land-use and cover classes of each point (Luiz et al., 2002). Three hundred fifty randomly distributed sample points were used in the mappings, 50 of them per class. Evaluations were carried out visually by Google Earth high-resolution images, with the aid of MC, generating the error matrices for each mapping. From the error matrix, OA and K were calculated.

Other accuracy indices were also determined. One is based on the Producer's Accuracy (PA), which stands for the probability of a given pixel value being a member of a particular class. Another is the User's Accuracy (UA), which is the probability of a pixel classified on the map actually representing that category on the field (Congalton, 1991). To check for significant differences in precision measurements among different classification results, the Z test (Foody, 2009) was used as follows:

$$Z = \frac{P_1 - P_2}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (2)$$

where:

$\overline{p}$ - $(x_1 + x_2)/(x_1 - x_2)$

P_1 and P_2 - Kappa indices of each method compared;

x_1 - number of cases allocated correctly in data classifications with size n_1; and,

x_2 - number of cases allocated correctly in data classifications with size n_2.

In this test, it is assumed that if $|Z| > 1.96$, both classifications are significantly different at $p \leq 0.05$ (Foody, 2009).

Results and Discussion

The classifiers showed different performances regarding the mapping of the seven classes of land-use and cover (Figure 1)
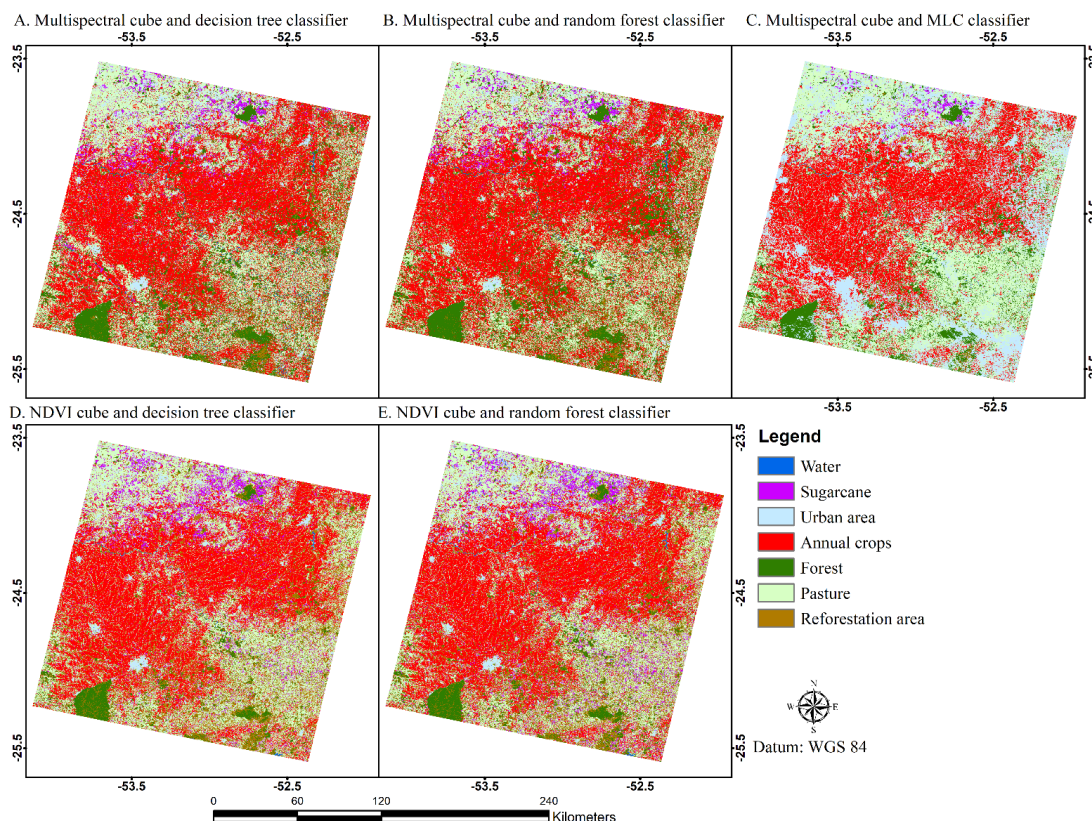

Figure 1. Classification of Landsat-8 images spectral metrics (MC), using as classifiers Decision Tree (A), Random Forest (B), and Maximum Likelihood Classifier - MLC (C), and by Normalized Difference Vegetation Index (NDVI) temporal metrics (NC), using as classifiers Decision Tree (D) and Random Forest (E) for urban areas, forests, sugarcane field, reforestation areas, annual crops, pasture, and water bodies

with two databases (MC and NC). The DT and RF classifier in the MC confused the pasture class with the annual crops (Figure 1A and B). In turn, the classifiers using the NC (Figure 1D and E), generated confusion between the sugarcane and the pasture classes. All classifiers were able to identify the Iguaçu National Park in the southeast region of the scene (Figure 1), which represents a large homogeneous and preserved area of Atlantic Forest (Ribeiro et al., 2009).

In the classifications that used the NC, larger amounts of areas classified as the sugarcane class were observed, mainly in the northern region of the study area (Figures 1D and E). For Adami et al. (2012a,b), these sugarcane areas were only in the northern region, during the 2010/2011 agricultural year. Likewise, when mapping crops in the state of Paraná, Brazil, between 2010 and 2014, Cechim Junior et al. (2017) identified areas under sugarcane solely to the northern; these authors used MLC classifier on images from Landsat 5 TM, Landsat 8 OLI, and IRS LISS-3, with OA above 93%.

Classifications using NC also identified the largest areas under annual crops, which were concentrated more from west to north of the study area, corroborating the results of other studies (Souza et al., 2015; Zhong et al., 2016). The best mapping accuracy was achieved when the algorithm RF was used, for both MC and NC images. Using a sample panel to classify the entire satellite scene, the mappings with NDVI DT (OA: 84% and K: 0.81) showed the best results, followed by the NDVI RF and MLC (OA: 82% and K: 0.79) (Table 1).

The MC maps (DT, RF and MLC) obtained low user's accuracy when classifying other targets as city and mainly annual agricultural crops. This was because some agricultural areas were in fallow period, or with the soil turned over; therefore, they are spectrally like urban areas. The RGB MLC achieved the best results for the class Water, showing PA and UA of 100 and 96%, respectively.

Land use classifications using NDVI temporal data had low PA for the forest (DT: 72% and RF: 75%) and UA for reforestation area (DT: 66% and RF: 70%), classifying forest as reforestation areas. Pasture also showed a low value of PA (DT: 73% and RF: 68%). This was mainly due to the misclassification errors between the classes pasture and sugarcane. This issue

was also reported by other authors (Xavier et al., 2006; Adami et al., 2012a). For Xavier et al. (2006), this is due to a similarity in temporal behavior of NDVI for both classes.

Regarding the annual crops, the best results were seen when using NDVI temporal metrics both for DT (PA: 86%; UA: 100) and RF (PA: 77%; UA: 100). Likewise, Jia et al. (2014a) observed the best results using NDVI metrics (maximum, minimum and mean values and standard deviation) when compared to phenological metrics (start and end of the growing season, duration, seasonal amplitude and maximum adjusted NDVI) and to spectral data of a single date using images from OLI sensor. According to these authors, this outcome arises from a lack of sensitivity of the NDVI temporal metrics to planting and harvesting periods. For a single image (RGB), the date has relevant influence on results (Senf et al., 2015), as in some areas crops are under development, whereas in others, they have already been harvested. By using NDVI temporal metrics, fewer misclassification errors were found for annual crops, but with misleading interpretations in other classes (mainly between sugarcane with pasture). Therefore, rankings were also evaluated separating only the annual crops from a general class representing the other targets.

The NDVI RF, NDVI DT and RGB MLC classifications showed no statistical difference by the Z test ($|Z| < 1.96$) with higher accuracy than the others. The same trend was seen for RGB RF and RGB DT, but with the lowest accuracy. Yet the classifiers using NDVI temporal metrics had statistically the same results (Table 2).

Table 2. Comparison of Kappa indices by Z test obtained by random distribution of sampling points for classification of urban areas, forest, sugarcane, reforestation, annual crops, pasture and water bodies using Decision Tree (DT), Random Forest (RF) and Maximum Likelihood Classifier (MLC) on Normalized Difference Vegetation Index (NDVI) and temporal metrics (NC) and spectral metrics (MC)

Methods	MC MLC	MC DT	NC DT	MC RF
MC DT	3.16*	-		
NC DT	0.61ns	3.75*	-	
MC RF	2.13*	1.03ns	2.73*	-
NC RF	0.10ns	3.06*	0.70ns	2.04*

* - Significant at $p \leq 0.05$ by z test; ns - Not significant

Table 1. Accuracy indices generated from algorithms and random distribution of points in classifications using Normalized Difference Vegetation Index (NDVI) temporal metrics (NC) and spectral metrics (MC) and the classifiers Decision Tree (DT), Random Forest (RF), and Maximum Likelihood Classifier (MLC)

		NC DT		NC RF		MC DT		MC RF		MC MLC	
Algorithms	OA (%)	98.34		99.10		98.90		99.93		-	
	K	0.9807		0.9930		0.9872		0.9992		-	
Sample panel	OA (%)	84		82		72		76		82	
	K	0.81		0.79		0.68		0.72		0.79	
		NC DT		**NC RF**		**MC DT**		**MC RF**		**MC MLC**	
		PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
		(%)									
Urban area		82	80	90	72	100	60	97	58	95	42
Forest		72	96	75	98	75	98	75	82	68	98
Sugarcane		100	60	89	68	96	52	89	62	93	82
Reforestation		94	66	95	70	79	60	95	76	93	84
Annual crops		86	100	77	100	54	90	49	80	75	94
Pasture		73	90	68	92	56	80	65	80	71	80
Water bodies		98	96	100	74	97	66	98	92	100	96

OA - Overall Accuracy; K - Kappa coefficient; PA - Producer's Accuracy; UA - User's Accuracy

Classifications using spectral information (MC) had more classification noise and misleading between annual crops and other targets compared to those using NC metrics (Figure 2).

The classification accuracy estimated by algorithms showed good results (OA: 94.4 to 100%, K: 0.98 to 1.0). Nonetheless, the same is not true for accuracy evaluation by means of sample panel. Superior results were achieved by NC classifications both using DT (OA: 98%; K: 0.96) and RF (OA: 96%; K: 0.92) when compared to MC (the best result was with RGB RF; OA: 88%; K: 0.76) (Table 3).

In the literature there are other authors reporting equivalent results. Using NDVI spectral-temporal metrics, Müller et al. (2015) obtained an OA of 93% while identifying grazing areas on the Cerrado biome. Similarly, Jia et al. (2014b) came to close results with an OA of 93% and K of 0.87 for classification of forest cover by means of NDVI spectral-temporal metrics.

NDVI temporal metrics improved OA by nearly 11% and K by 16%. Thus, statistical values extracted from the NDVI profile showed to be able to improve land-use and cover characterization (Jia et al., 2014a; Valero et al., 2016).

The joining of urban area, forest, sugarcane, reforestation, pasture and water bodies into a single class improved classification results (OA: from 84 to 98%; K from 0.80 to 0.96). This is because there is an increase in misclassifications while trying to differentiate such classes. Thus, by reducing the number of classes, a better classification accuracy can be achieved (Senf et al. 2015).

The NC classification with DT reached high PA and UA (above 96%) for both classes (Table 3). This classification reached a PA of 100% for other targets (i.e. all the points from other targets were correctly sorted) and UA of 100% for annual crops (all points classified as a crop are true). This classification obtained 3.8% error of omission for crops and 4% error of inclusion for other targets.

MC classifications had no statistical differences between each other by Z test (| Z |<1.96) (Table 4). The same is true for the NC analysis method. Therefore, the differentiation between annual crops and other targets was more influenced by NDVI metrics than the use of classification algorithm.
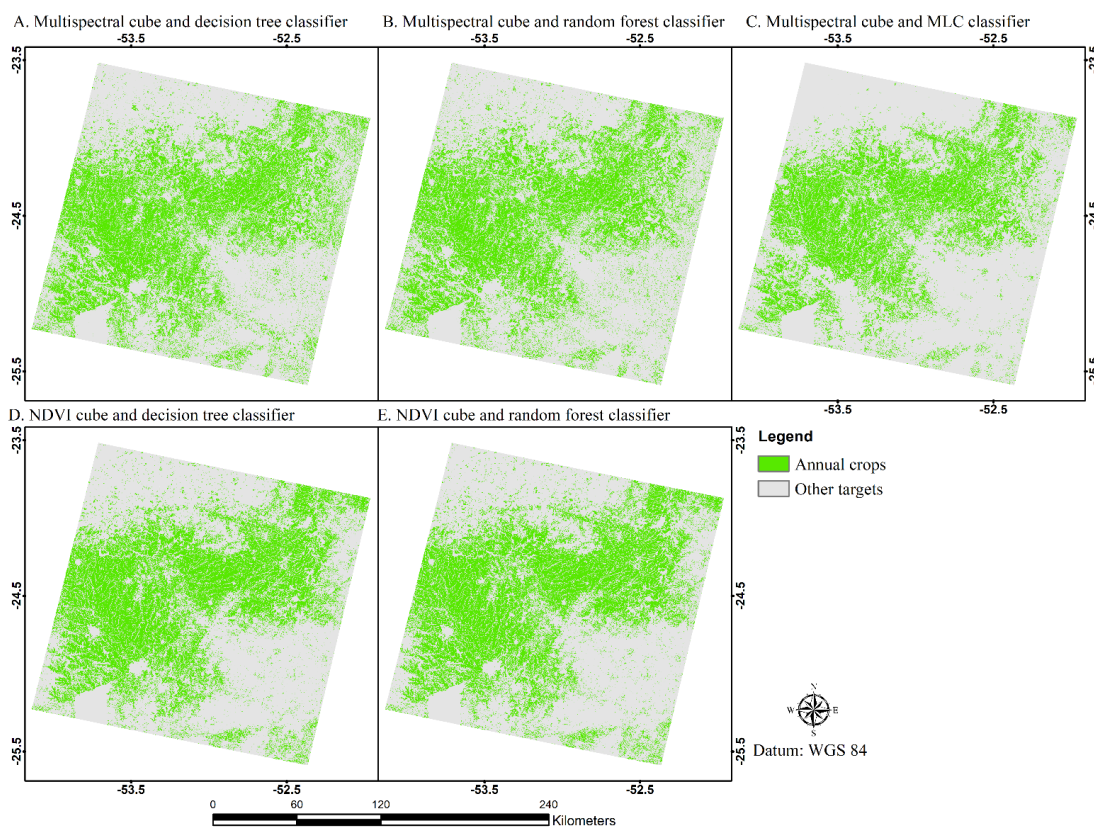

Figure 2. Classification of Landsat-8 images spectral metrics (MC), using as classifiers Decision Tree (A), Random Forest (B) and Maximum Likelihood Classifier - MLC (C), and by Normalized Difference Vegetation Index (NDVI) temporal metrics (NC), using as classifiers Decision Tree (D) and Random Forest (E), for the classes annual crops and other targets

Table 3. Accuracy indices generated from algorithms and random distribution of points in Normalized Difference Vegetation Index (NDVI) temporal metrics (NC) and spectral metrics (MC) using as classifiers Decision Tree (DT), Random Forest (RF) and Maximum Likelihood Classifier (MLC) for annual crops and other targets

Methods	Algorithms		Sample panel		Producer's Accuracy - PA (%)		User's Accuracy - UA (%)	
	Overall accuracy (%)	Kappa index	Overall accuracy (%)	Kappa index	Annual crops	Other targets	Annual crops	Other targets
NC DT	99.4	0.99	98.0	0.96	96.2	100.0	100	96
NC RF	100.0	1.00	96.0	0.92	94.2	97.9	98	94
MC DT	99.6	0.98	85.0	0.70	85.7	84.3	84	86
MC RF	99.9	0.99	88.0	0.76	86.5	89.6	90	86
MC MLC	-	-	87.0	0.74	82.5	93.0	94	80

Table 4. Comparison of Kappa indices by Z test obtained by random distribution of sampling points for classification of annual crops and other targets using Decision Tree (DT), Random Forest (RF) and Maximum Likelihood Classifier (MLC) on temporal metrics (NC) and espectral metrics (MC) images

Methods	MC MLC	MC DT	NC DT	MC RF
MC DT	0.41^{ns}	-		
NC DT	2.95^{*}	3.30^{*}	-	
MC RF	0.21^{ns}	0.62^{ns}	2.77^{*}	-
NC RF	2.28^{*}	2.65^{*}	0.8^{ns}	2.09^{*}

* - Significant at $p \leq 0.05$; ns - Not significant

Conclusions

1. The temporal metrics (NC) obtained good producer's and user's accuracies with the annual crop class, while for this class with the espectral metrics (MC) there were more confusions for all the classification algorithms used.

2. Considering only two classes (annual crops/other targets), the classifications using the temporal metrics (NC) obtained higher accuracy than classifications that used the spectral attributes.

3. The classification result depends more on the attribute used than on the classification algorithms.

4. The use of Normalized Difference Vegetation Index (NDVI) metrics information, which shows the phenological variations of the crops, together with data mining techniques, proved to be effective in the differentiation of annual crops from the other targets, generating a precise mapping.

Acknowledgements

Literature Cited

Adami, M.; Mello, M. P.; Aguiar, D. A.; Rudorff, B. F. T.; Souza, A. F. A web platform development to perform thematic accuracy assessment of sugarcane mapping in South-Central Brazil. Remote Sensing, v.4, p.3201-3214, 2012a. https://doi.org/10.3390/rs4103201

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, p.574-585, 2012b. https://doi.org/10.3390/su4040574

Amaral, M. V. F.; Souza, A. L. de; Soares, V. P.; Soares, C. P. B.; Leite, H. G.; Martins, S. V.; Fernandes Filho, E. I.; Lana, J. M. de. Avaliação e compação de métodos de classificação de imagens de satélites para o mapeamento de estádios de sucessão florestal. Revista Árvore, v.33, p.575-582, 2009. https://doi.org/10.1590/S0100-67622009000300019

Breiman, L. Random forests. Machine Learning, v. 45, p. 5-32, 2001. https://doi.org/10.1023/A:1010933404324

Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. Classification and regression trees. 1. ed. Belmont: Wadsworth, 1984. 368p.

Cattani, C. E. V.; Garcia, M. R.; Mercante, E.; Johann, A. J.; Correa, M. M.; Oldoni, L. V. Spectral-temporal characterization of wheat cultivars through NDVI obtained by terrestrial sensors. Revista Brasileira de Engenharia Agrícola e Ambiental, v.21, p.769-773, 2017. https://doi.org/10.1590/1807-1929/agriambi.v21n11p769-773

Cechim Junior, C.; Johann, J. A.; Antunes, J. F G. Mapping of sugarcane crop area in the Paraná state using Landsat / TM / OLI and IRS / LISS-3 images. Revista Brasileira de Engenharia Agrícola e Ambiental, v.21, p.427-432, 2017. https://doi.org/10.1590/1807-1929/agriambi.v21n6p427-432

Chen, Y.; Lu, D.; Moran, E.; Batistella, M.; Dutra, L. V.; Sanches, I. D.; da Silva, R. F. B.; Huang, J.; Luiz, A. J. B.; de Oliveira, M. A. F. Mapping croplands, cropping patterns, and crop types using modis time-series data. International Journal of Applied Earth Observation and Geoinformation, v. 69, p. 133-147, 2018. https://doi.org/10.1016/j.jag.2018.03.005

Cohen, J. A Coefficient of agreement for nominal scales. Educational and Psychological Measurement, v. 20, p. 37-46, 1960. https://doi.org/10.1177/001316446002000104

Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, v.37, p.35-46, 1991. https://doi.org/10.1016/0034-4257(91)90048-B

Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. AI Magazine, v. 17, p. 37-54, 1996.

Foody, G. M. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. Remote Sensing of Environment, v.113, p.1658-1663, 2009. https://doi.org/10.1016/j.rse.2009.03.014

Grande, T. O. de; Almeida, T. de; Cicerelli, R. E. Classificação orientada a objeto em associação às ferramentas reflectância acumulada e mineração de dados. Pesquisa Agropecuária Brasileira, v.51, p.1983-1991, 2016. https://doi.org/10.1590/s0100-204x2016001200009

Jia, K.; Liang, S.; Wei, X.; Yao, Y.; Su, Y.; Jiang, B.; Wang, X. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sensing, v.6, p.11518-11532, 2014a. https://doi.org/10.3390/rs61111518

Jia, K.; Liang, S.; Zhang, L.; Wei, X.; Yao, Y.; Xie, X. Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data. International Journal of Applied Earth Observation and Geoinformation, v.33, p.32-38, 2014b. https://doi.org/10.1016/j.jag.2014.04.015

Lary, D. J.; Alavi, A. H.; Gandomi, A. H.; Walker, A. L. Geoscience Frontiers Machine learning in geosciences and remote sensing. Geoscience Frontiers, v. 7, p. 3-10, 2016. https://doi.org/10.1016/j.gsf.2015.07.003

Luiz, A. J. B.; Oliveira, J. C.; Epiphanio, J. C. N.; Formaggio, A. R. Auxílio das imagens de satélite aos levantamentos por amostragem em agricultura. Agricultura em São Paulo, v.49, p.41-54, 2002.

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, p.490-499, 2015. https://doi.org/10.1016/j.rse.2014.10.014

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, v.12, p.2825-2830, 2011.

Ribeiro, M. C.; Metzger, J. P.; Martensen, A. C.; Ponzoni, F. J.; Hirota, M. M.; The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biological Conservation, v.142, p.1141-1153, 2009. https://doi.org/10.1016/j.biocon.2009.02.021

Rouse, J. W.; Hass, R. H.; Schell, J. A.; Deering, D. W.; Harlan, J. C. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. Greenbelt: NASA/GSFC, p.1-137, 1974. Final Report, n. September 1972

Senf, C.; Leitão, P. J.; Pflugmacher, D.; Linden, S. van der; Hostert, P. Mapping Land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sensing of Environment, v.156, p.527-236, 2015. https://doi.org/10.1016/j.rse.2014.10.018

Silva, C. R. da; Souza, K. B. de; Furtado, W. F. Evaluation of the progress of intensive agriculture in the Cerrado Piauiense - Brazil. IERI Procedia, v.5, p.51-58, 2013. https://doi.org/10.1016/j.ieri.2013.11.069

Souza, C. H. W.; Mercante, E.; Johann, J. A.; Lamparelli, R. A. C.; Uribe-Opazo, M. A. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. International Journal of Remote Sensing, v.36, p.1809-1824, 2015. https://doi.org/10.1080/01431161.2015.1026956

Toth, C.; Jóźków, G. Remote sensing platforms and sensors: A survey. ISPRS Journal of Photogrammetry and Remote Sensing, v.115, p.22-36, 2016. https://doi.org/10.1016/j.isprsjprs.2015.10.004

USGS - United States Geological Survey. Landsat 8 (L8) Data users handbook. 4.ed. Sioux Falls: USGS, 2019. 115p.

Valero, S.; Morin, D.; Inglada, J.; Sepulcre, G.; Arias, M.; Hagolle, O.; Dedieu, G.; Bontemps, S.; Defourny, P.; Koetz, B. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Remote Sensing, v.8, p.1-21, 2016. https://doi.org/10.3390/rs8010055

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, p.46-56, 2016. https://doi.org/10.1016/j.rse.2016.04.008

Vintrou, E.; Ienco, D.; Begue, A.; Teisseire, M. Data mining, a promising tool for large-area cropland mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v.6, p.2132-2138, 2013. https://doi.org/10.1109/JSTARS.2013.2238507

Xavier, A. C.; Bernardo F. T. R.; Shimabukuro, Y. E.; Berka, L. M. S.; Moreira, M. A. Multi-temporal analysis of MODIS data to classify sugarcane crop. International Journal of Remote Sensing, v.27, p.755-768, 2006. https://doi.org/10.1080/01431160500296735

Yao, F.; Tang, Y.; Wang, P.; Zhang, J. Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China Plain. Physics and Chemistry of the Earth, v.87-88, p.142-152, 2015. https://doi.org/10.1016/j.pce.2015.08.010

Zhong, L.; Hu, L.; Yu, L.; Gong, P. G. S. Biging. Automated mapping of soybean and corn using phenology. ISPRS Journal of Photogrammetry and Remote Sensing, v.119, p.151-164, 2016. https://doi.org/10.1016/j.isprsjprs.2016.05.014

Zhu, Z.; Wang, S.; Woodcock, C. E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. Remote Sensing of Environment, v.159, p.269-277, 2015. https://doi.org/10.1016/j.rse.2014.12.014