Remote Sensing/ Original Article

Multitemporal variables for the mapping of coffee cultivation areas

Abstract – The objective of this work was to propose a new methodology for mapping coffee cropping areas that includes multitemporal data as input parameters in the classification process, by using the Landsat TM NDVI time series, together with an object-oriented classification approach. The algorithm BFAST was used to analyze coffee, pasture, and native vegetation temporal profiles, allied to a geographic object-based image analysis (GEOBIA) for mapping. The following multitemporal variables derived from the R package greenbrown were used for classification: mean, trend, and seasonality. The results showed that coffee, pasture, and native vegetation have different temporal behaviors, which corroborates the use of these data as input variables for mapping. The classifications using temporal variables, associated with spectral data, achieved high-global accuracy rates with 93% hit. When using only temporal data, ratings also showed a hit percentage above 80% accuracy. Data derived from Landsat TM time series are efficient for mapping coffee cropping areas, reducing confusion between targets and making the classification process more accurate, contributing to a correct characterization and mapping of objects derived from a RapidEye image, with a high spatial solution.

Index terms: BFAST, classification, MODIS, NDVI, remote sensing, R package greenbrown.

Carolina Gusmão Souza[1] ✉,
Tássia Borges Arantes[2],
Luis Marcelo Tavares de Carvalho[2] and
Polyanne Aguiar[2]

[1] Universidade Estadual do Sudoeste da Bahia, Departamento de Ciências Exatas e Naturais, CEP 45700-000, Itapetinga, BA, Brazil.
E-mail: carolinagusmao@uesb.edu.br

[2] Universidade Federal de Lavras, Departamento de Ciências Florestais, Caixa Postal 3037, CEP 37200-000 Lavras, MG, Brazil.
E-mail: tassiabarantes@gmail.com.br, passarinho@ufla.br, polyanneaguiar@gmail.com

✉ Corresponding author

Variáveis multitemporais para o mapeamento de áreas de cultivo de café

Resumo – O objetivo deste trabalho foi propor uma nova metodologia para o mapeamento de áreas cafeeiras que inclui dados multitemporais como parâmetros de entrada no processo de classificação, por meio de uma série temporal NDVI do Landsat TM, juntamente com uma abordagem de classificação orientada a objeto. O algoritmo BFAST foi utilizado para a análise dos perfis temporais de café, pastagem e vegetação nativa, aliada à análise da imagem baseada em objetos geográficos. Para a classificação, utilizaram-se as seguintes variáveis multitemporais derivadas do pacote greenbrown R: média, tendência e sazonalidade. Os resultados mostraram que o café, a pastagem e a vegetação nativa têm comportamentos temporais distintos, o que corrobora o uso destes dados como variáveis de entrada para o mapeamento. As classificações com uso das variáveis temporais, associadas a dados espectrais, obtiveram altos índices de acurácia global com 93% de acerto. Quando utilizados somente os dados temporais, as classificações ainda mostraram um percentual de acerto acima de 80%. Dados oriundos de séries temporais do Landsat TM são eficientes para o mapeamento de áreas de cultivo cafeeiro, diminuindo a confusão entre os alvos e tornando o processo de classificação mais preciso, o que contribui para a caracterização e o mapeamento de objetos derivados de uma imagem RapidEye, com alta resolução espacial.

Termos para indexação: BFAST, classificação, MODIS, NDVI, sensoriamento remoto, pacote greenbrown R.

Introduction

Coffee (*Coffea* spp.) is one of the most valuable basic products in the international scenario, surpassed in monetary value only by petroleum, which makes it an activity of great importance in the international market. Brazil stands out as the largest coffee producer in the world, as it is responsible for 31.9% of world production (OIC, 2018). Considering the dimension of coffee cultivations in the country, and coffee as an important commodity, the mapping and monitoring of coffee cropping areas are essential to the management of the resources. They are also used to obtain information about spatial distribution of coffee cultivations over various scales.

The mapping of agricultural areas has become essential because these areas are crucial to the economic development of many regions. Several techniques for mapping coffee cultivation areas were previously used by some authors (Cordero-Sancho & Sader, 2007; Martínez-Verduzco et al., 2012; Santos et al., 2012; Sarmiento et al., 2014), including the visual classification (Machado et al., 2010), the supervised pixel-based classification approach (Cordero-Sancho & Sader, 2007; Martínez-Verduzco et al., 2012), the object-based approach (Santos et al., 2012; Sarmiento et al., 2014; Souza et al., 2016), the machine-learning algorithms (Santos et al., 2012; Sarmiento et al., 2014; Souza et al., 2016), the use of different variables (Santos et al., 2012; Souza et al., 2016), and physical data (Prado et al., 2016). However, most of these works did not achieve satisfactory results for accuracy, once coffee cultivations are commonly misinterpreted with other vegetation, such as pasture and native vegetation, when automatic classification techniques are used.

An alternative to improve mapping of coffee cropping areas may consist in the use of data from time series. Many orbital satellites can provide data images with a regular temporal frequency, which is ideal for time series studies. These temporal data can help with the characterization of land cover (Huang et al., 2010). However, for the classification of coffee cropping areas, these variables have not yet been used as data input for classification.

Remote sensing images such as those of the Landsat are important for the monitoring of land cover, once they provide a large land-surface data and, thereby, they are able to support the analysis of temporal or phenological information of certain features, which makes them useful as tools for mapping, especially for purposes of classifying the vegetation cover (Jia et al., 2013).

Methodologies have been developed to extract temporal trajectories of different land cover by vegetation indexes such as BFAST (breaks for additive seasonal and trend), which is a package available in the software RStudio (R Core Team, 2014). This algorithm is capable of decomposing the time series into three different components: trend, seasonality, and error, using methods to detect and characterize changes in time series (Verbesselt et al., 2010). Moreover, greenbrown algorithm is another tool that evaluates the performance of different methods to estimate the trend, as well as to infer the time series properties (Forkel et al., 2013).

Some authors have used multitemporal variables during the classification process, obtaining successful results when using these data for land cover mapping. Mello et al. (2013) showed a new approach in sugarcane mapping, by using temporal variables, and obtained excellent results with 99% hit indexes. Jia et al. (2014) used temporal variables of fused data in order to improve land cover mapping. Temporal features were extracted and added to improve the accuracy of classification. The results showed that temporal features extracted from image data with coarse resolution have a significant effect on the improvement of classification accuracy of images with better resolution, especially for different types of vegetation. Santos (2014) evaluated the performance of different methods to detect trends and breakpoints (points of interruption in the time series) in normalized difference vegetation index (NDVI) time series. The use of such methods allowed of a critical assessment of trends, as well as the understanding of the seasonal component, providing a significant improvement in classifications, with accuracy indexes of more than 90%.

The use of the spectral profiles over time, along with an object-oriented classification approach, integrating temporal and spectral data in the classification process, may be a valuable alternative to obtain successful results for the mapping of coffee cropping areas. Developing a more accurate methodology is important, as faster and more precise results of the status and size of coffee cropping areas are essential as a basis to monitor them.

For the mapping of the coffee cultivation areas, most works that used temporal data also use bitemporal analysis, or just analyze the behavior of this culture over the years (Bernardes et al., 2012). However, the use of multitemporal data as input variables in the classification process of coffee cropping areas is still absent in the literature.

It is expected to answer the following questions: if the temporal variables are efficient for coffee separability; what is the adequate quantity of variables needed to improve the accuracy of the classification; and, if the temporal data from Landsat TM are compatible for mapping RapidEye images.

The objective of this work was to propose a new methodology for mapping coffee cropping areas that includes multitemporal data as input parameters in the classification process, by using the Landsat TM NDVI time series, together with an object-oriented classification approach.

Materials and Methods

The study area was located in the southern Minas Gerais state, Brazil, and corresponds to a subset of the Landsat TM path/row 218/75 scene (Figure 1). The vegetation is characteristic to the Brazilian Atlantic Forest biome. The region's climate is a transition between Cwa (humid temperate climate with dry winter and hot summer) and Cwb (humid temperate climate with dry winter and moderately hot summer), according to the Köppen-Geiger's classification. The local is a transitional region between the previously mentioned biomes, and it is characterized by a great diversity of environmental configurations.

The prevailing native vegetation in the region is the montane semideciduous seasonal forest (*Floresta Estacional Semicidual Montana*), according to Veloso et al. (1991). The remaining native vegetation is very fragmented, and the anthropic impacted portion is constituted mainly by areas of pastures and coffee cultivations.

Images of the RapidEye sensor represent a data set of high-spatial resolution (5 m), radiometric resolution of 16 bits, and 5-band spectral resolution. These images were acquired previously with the geometric and radiometric corrections (level 3A). The image used (tile 2328914, from September 19th 2012) was provided

by the Ministério do Meio Ambiente (MMA), Brazil, and served as a basis for the classification.

To obtain the temporal data used as classification parameters, a time series from Landsat TM images were used. For this purpose, all the available images between the years 2000 and 2011, path/row 218/75 (corresponding to the RapidEye tile), were downloaded from the U.S. Geological Survey (USGS, 2019a), totalling 195 images (Landsat Surface Reflectance Climate Data Records - CDRs), with 30 m spatial resolution, 8-bit-radiometric resolution, and 7-band spectral resolution.

The CDR product is generated from a specialized software called Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS). The LEDAPS produce data about surface reflectance and temperature, cloud and cloud shadow location, and soil and water, from information such as water vapor, ozone, geopotential height, optical thickness of aerosols, and terrain elevation. These datasets are inputs for the 6S algorithm (Satellite Signal in the Solar Spectrum), resulting in CDR products that are orthorectificated (Masek et al., 2006).

However, because the images were acquired from a tropical region, Landsat TM images may show a significant cloud cover. In addition, many Landsat TM images are not available for download in some specific dates, thus hindering its temporal analysis. In order to solve this problem, a fusing method between orbital images was used, using data from the Moderate-Resolution Imaging Spectroradiometer (MODIS) sensor MOD13Q1, from the Modis Reproject Tools Web (USGS, 2019b), corresponding to the same period of the Landsat TM images. The product MOD13Q1 has a spatial resolution of 250 m for NIR and RED bands, which were used for fusion with the respective Landsat bands. NIR and RED bands from MODIS images were reprojected to Universal Transverse Mercator (UTM), datum WGS-84 zone 23S, using the software MODIS Re-projection Tools (MRT). In addition, the pixel was resampled for 30 m, following the same settings of the Landsat TM images. Subsets of the size of the Landsat TM images were prepared for all MODIS images.

For the prediction of Landsat images, the algorithm Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), proposed by Gao et al. (2006), was used for fusing the images Landsat TM and MODIS. MODIS images, despite its lower-spatial resolution

and more recently production of orbital data, has a better temporal resolution (1 or 2 days), significantly improving the composition of time series.

The algorithm STARFM predicts surface-reflectance values from pairs of images Landsat (fine-resolution) and MODIS (time T_1), preferable from the same date and MODIS (coarse-resolution) images, corresponding to the image date to be predicted (time T_2). The final result is an image of high-temporal and spatial resolutions (Gao et al., 2006).

The time series Landsat TM was constituted of original and synthetic images, where all images were transformed to NDVI. From this time series,

the temporal parameters used in the classification of RapidEye images were obtained, as below described.

With the purpose to verify if temporal data were efficient to separate classes, the behavior of temporal signatures of different types of land cover was evaluated. To complete this task, the algorithm BFAST (Verbesselt et al., 2010) was used. Modifications were detected by means of identifying resulting breakpoints from the BFAST decomposition process. With the establishment of the decomposed time series, the trend component will have high relevance for the detection of a real disturbance, once it allows of the determination of the direction of a modification by analyzing the signal of the slope.
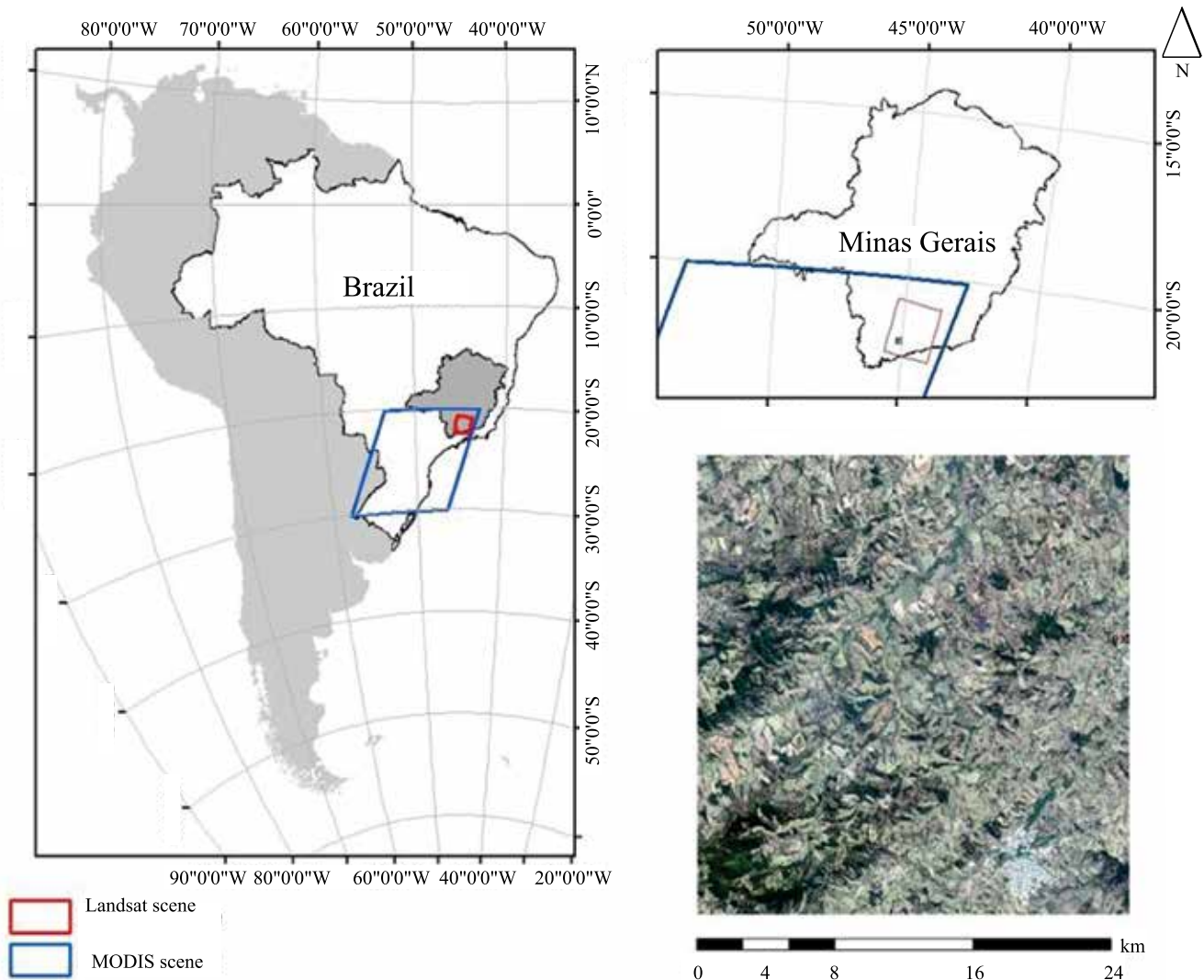
Figure 1. Location of the study area, in the southern Minas Gerais state, Brazil, showing the location of scenes Landsat TM 218/75 and MODIS h13v11. The RapidEye subset in which the study was performed is showed in the detail.

In order to extract the temporal variables for the classification process, the method proposed by Forkel et al. (2013) was used, which allows to evaluate different methods to estimate the trend. This methodology was implemented by using the package named "greenbrown" (Forkel & Wutzler, 2015), available for the R software that shows a variety of functions to analyze trends in time series. In the present paper, statistical properties of time series from Landsat TM images were evaluated. This procedure renders seven layers of information (denominated GetStatistic): mean, trend, inter-annual variability (IAV), standard deviation of IAV, seasonal variation, intra-annual variation (STV), and the standard deviation of STV. In addition to these statistical properties, other trend methods were also analyzed, such as the annual aggregated time series (AAT), season-trend model (STM), mean annual cycle (MAC), and annual cycle based on singular spectrum analysis (SSA). These methods differ for their temporal resolution from the NDVI time series, presence, or absence of the seasonal component, and how the trend is estimated (Forkel et al., 2013).

In all methods, the same algorithm for detecting breakpoints was used, as described by Bai & Perron (2003). At first, the test of ordinary-least squares moving sum, or MOSUM, is applied to verify the existence of significant breakpoints in the time series. Then, the optimal number of breakpoints and their location in the series are interactively tested (Forkel et al., 2013).

To verify the effectiveness of different variables for classification, two sets of variables – spectral and temporal – were used. Spectral variables were used according the methodology proposed by Souza et al. (2016), which shows the importance of these variables for the classification of remote sensing images in coffee cropping areas. Temporal variables were chosen according the methodology proposed by Forkel et al. (2013), as previously described.

The following input variables were used for classification: spectral ones – brightness, maximum difference, mean, NDVI, and standard deviation – from RapidEye images (Souza et al., 2016); and temporal ones – GetStatistic (mean, trend, IAV, standard deviation IAV, seasonal variation, STV, standard deviation of STV), AAT, STM, MAC, SSA, from the Landsat TM time series analysis using greenbrown (Forkel et al., 2013).

Different sets of temporal variables were tested, in order to evaluate the prominence of these variables in the process of class separation, as follows: 1, GetStatistic; 2, GetStatistic+STM; 3, GetStatistic+STM+AAT; 4, GetStatistic+STM+AAT+MAC; 5, GetStatistic+STM+AAT+MAC+SSA; and 6, best variables. These data were tested in two forms: the first one, using spectral RapidEye data together with temporal data, and the second one, using only temporal data.

The classification was performed in the RapidEye image, using an object-based approach, in which the image was segmented by gathering pixels according to similarities among their spectral values, thus forming homogeneous units within the landscape – the so called "objects". Variables (spectral and temporal ones) associated with the objects were subsequently used as input data for the classification algorithm.

The landscape was divided into five classes of land cover – native vegetation, coffee cultivation, pastures, "other-use areas", and water bodies – according to their spectral characteristics, described as follows. Native vegetation, which designates dense forest areas and gallery forests, as well as Cerrado formations. Coffee, which are the nonproducing croppings, at the beginning of their growing stage, and producing croppings. Pastures, which are areas with natural, or planted pastures. And areas designated as of "other use", which were subdivided into two classes for the training sample collection, as follows: "other-use 1", which are areas with annual crops at different development stages, and production forests; and "other-use 2", which are urban and constructions, areas with bare soil, and burned areas. And water bodies, which are areas of rivers, streams, and natural and artificial lakes.

All classifications were performed using the machine learning algorithm support vector machine (SVM). This algorithm was selected because it showed consistent results in previous researches (Li et al., 2014; Souza et al., 2016). The parameters of classification used in the algorithm SVM were the kernel function RBF, with 0.02 length, and the sensibility of decision margin of wrongly classified supporting vectors, defined as 1. Algorithms as random forest and decision tree build a ranking of the most important variables for the classification process. Consequently, these

rankings were used to classify the best variables for separation of the map classes.

Different quantities of samples were tested, with the aim to estimate the ideal number of training samples for classification. Four different sets were used: A, 500 samples; B, 700 samples; C, 1,000 samples; and D, 1,500 samples. These numbers were selected according to the suggestion by Van Niel et al. (2005), who recommend that the size of training samples should not be smaller than 10 to 30 times the number of bands.

A total of 48 classifications was performed, using four different numbers of training samples (A, B, C, and D), six sets of different variables (1, 2, 3, 4, 5, and 6) and two different approaches: one with the spectral variables of RapidEye images (CVR), and another without the use of these variables (SVR).

Finally, to check the accuracy of classifications, 300 points of test data were collected. The points were divided as follows: 60 as pastures; 90 as vegetation; 70 as other uses; 65 as coffee cultivations; and 15 as water bodies. The points were obtained using a stratified random sampling, by which each stratum was represented by a category of classification (Martínez-Verduzco et al., 2012). The verification of points was performed by means of field visits, and from data obtained in Google Earth. The Kappa coefficient and overall accuracy were used as evaluation criteria.

Results of accuracy were compared using the test of McNemar (Foody, 2004), a parametric test that considers the significance between the differences of two classes, based on bidimensional confusion matrix.

The test evaluation was based in the chi-square distribution. Therefore, the calculated value of chi-square is compared with the tabulated chi-square value, showing its statistical significance (Foody, 2004).

Results and Discussion

Different behaviors from different types of land cover over time (Figure 2) were observed through the analysis of time series decomposition. Breakpoints were detected for the trend component in coffee and pasture classes, while the seasonal component showed no alteration over time in the three evaluated classes (natural vegetation, coffee areas, and pasture). According to Verbesselt et al. (2010), changes in the trend component indicate gradual or sudden

modifications, however, they do not evidence necessarily a change in the type of land cover. Besides, modifications in the seasonal component indicate phenological changes (Verbesselt et al., 2010), which may evidence changes in the type of land cover. Land cover classes were separated by the algorithm BFAST.

Concerning the seasonal component, the interval between the curves generated for each cover type showed a threshold, and it was possible to observe differences of amplitude between classes and differences between native vegetation, coffee areas, and pasture. For pastures, the seasonality ranged from -0.15 to 0.10, while for native vegetation this interval was narrower, varying from -0.04 to 0.08, and for coffee areas, from -0.05 to 0.06. The shape of the growth profiles from the different land cover types, represented by the analysis of time series, hold information about the behavior of each analyzed vegetation type (Brown et al., 2013).

The trend component also showed variations over time, showing breakpoints for pastures in 2008, and for coffee areas in 2003, while the native vegetation showed no breakpoint. Maybe, the native vegetation has undergone often a minor anthropic intervention than other types of land cover, thus showing no modifications over time, in contrast to coffee and pasture areas, which were likely subjected to management, resulting in more anthropic modifications over time. Galford et al. (2008) emphasize that agricultural areas have a more complex phenology than native vegetation, showing therefore more breakpoints. As to pastures, the breakpoints in the trend component may have occurred due to the lack of management in these areas and to the fact that they are strongly degraded areas.

According to Gómez et al. (2016), the temporal signatures of vegetation indexes provide complementary information on the targets on the land surface and the changes in their biophysical attributes. These results express the significance of these variables in the differentiation of land cover, since the behavior of each type of cover is different over time. Therefore, the use of these variables in the process of classification may be favorable to the process of separation between classes.

Some temporal variables were tested with the objective to separate coffee cropping areas from other types of land cover, mainly the native vegetation and pastures. The temporal variables showed a significant

effect on the classification of land cover (Table 1). The results obtained were efficient, showing high indexes of accuracy in classification for all sets of the tested variables, with minimum accuracy of 80.50% hits (Table 1 - SVR, A, 3), even when synthetic images to compose time series were used, which favored the derivations of the temporal parameters used in classification. Jia et al. (2014) also evidenced the efficiency of such images for classification while using temporal data.

Variable sets 1 and 2 had the best performance in automatic classifications, with maximum accuracy of 92.75% hits, using also RapidEye variables (CVR). It is important to stress that all classifications using these sets of variables (CVR) showed accuracy results of more than 90% hits. All other sets of variables (3, 4, 5, and 6) showed alternated results, all with good hit indexes. The set of variables 1 used GetStatistic, which is composed by different temporal variables, such

as seasonality, trend, intra-annual and interannual differences, unlike the other variables which work with trend issues (Forkel et al., 2013). It is possible that the better performance of classifications was due to the fact that this set had a more heterogeneous quantity of parameters, with more temporal analyses. Besides, Forkel et al. (2013) emphasize that greenbrown has the capacity to estimate undergone changes and to characterize the different land covers existing in the ecosystems by using time series. This way, it is possible to sustain that data originated from their analyses were able to compose the input variables in the process of land cover classification.

The variable set 6 was composed by the best variables of separability in the classification (Figure 3); however, even using this set of data, the classifications did not achieve superior results. This shows that the most significant variables for separability of classes were those from set 1.
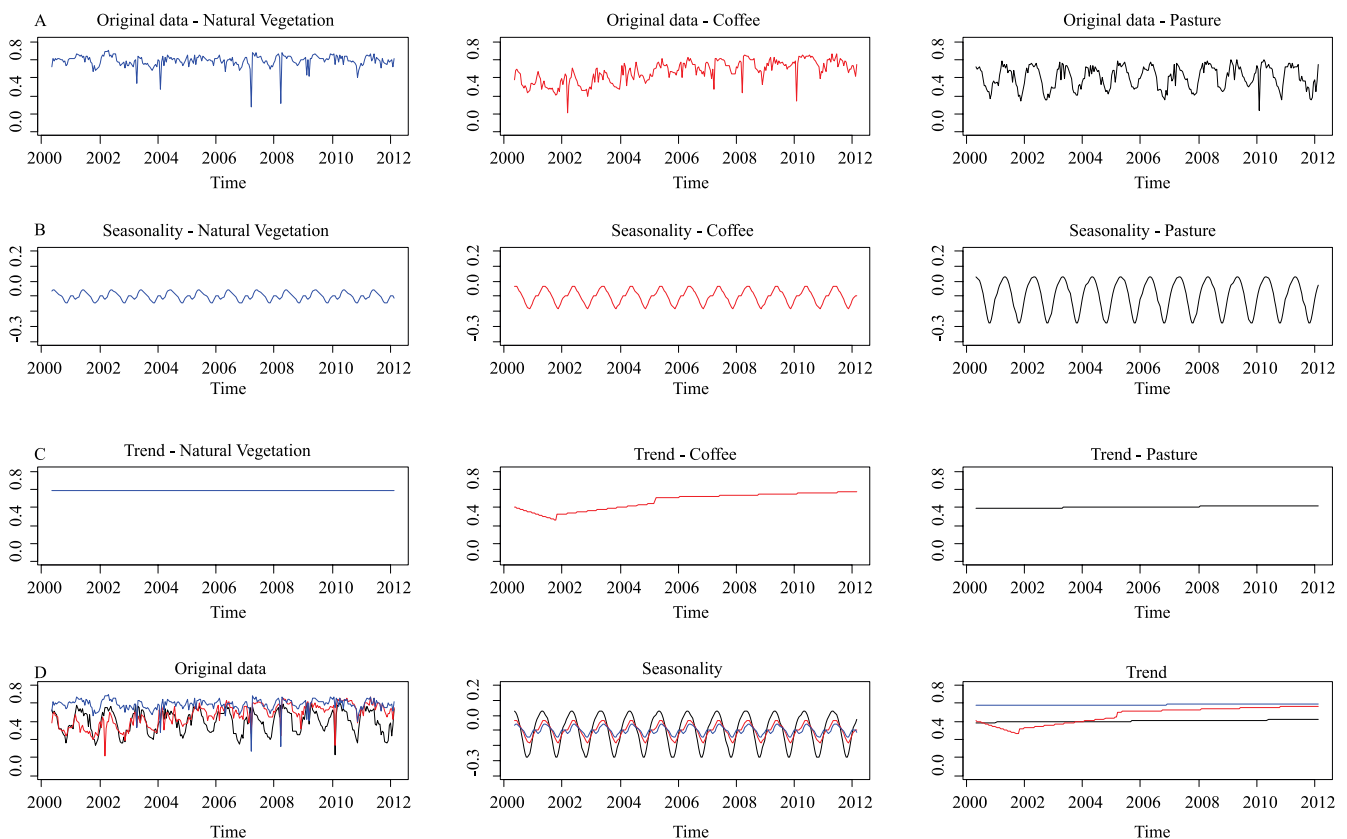
Figure 2. BFAST processing for natural vegetation, coffee cultivations, and pastures. A, original data from time series; B, seasonality of the three types of use; C, trend; D, comparison between different types of land use. Line colors: blue, vegetation; red, coffee cultivations; black, pastures.

Table 1. Classification accuracy for different sets of temporal variables[1].

Sets	Sample	Variable	Global (%)	Kappa	Accuracy (%)									
					Vegetation		Pasture		Coffee		Other Uses		Water	
					Producer	User	Producer	User	Producer	User	Producer	User	Producer	User
CVR	A	1	92.00	0.89	89.47	94.44	97.04	90.97	89.11	88.17	86.00	93.48	90.00	100.00
		2	90.50	0.87	88.42	93.33	96.30	87.84	90.00	88.04	80.00	93.02	90.00	100.00
		3	87.50	0.83	85.26	95.29	94.07	86.39	85.56	81.91	76.00	80.85	90.00	100.00
		4	88.25	0.84	86.32	95.35	94.07	87.59	88.89	83.33	76.00	80.85	86.67	100.00
		5	88.50	0.84	85.26	94.19	94.07	88.81	91.11	79.61	76.00	90.48	86.67	100.00
		6	89.75	0.86	86.32	93.18	97.04	87.92	88.89	84.21	78.00	95.12	90.00	100.00
	B	1	91.50	0.88	90.53	93.48	96.30	90.91	91.11	88.17	78.00	97.50	96.67	90.63
		2	90.75	0.87	90.53	92.47	96.30	88.44	90.00	90.00	74.00	97.37	96.67	90.63
		3	88.75	0.85	85.26	95.29	95.56	88.36	88.89	82.47	74.00	90.24	93.33	90.32
		4	88.50	0.84	85.26	94.19	95.56	87.16	91.11	81.19	68.00	100.00	93.33	90.32
		5	89.25	0.85	84.21	93.02	96.30	90.28	92.22	81.37	72.00	100.00	93.33	87.50
		6	89.00	0.85	85.26	93.10	97.04	87.33	90.00	83.51	68.00	100.00	96.67	90.63
	C	1	92.50	0.90	90.53	92.47	96.30	92.86	91.11	89.13	90.00	93.75	90.00	100.00
		2	92.75	0.90	90.53	93.48	96.30	92.86	91.11	89.13	92.00	93.88	90.00	100.00
		3	90.50	0.87	85.26	93.10	95.56	90.85	92.22	83.00	84.00	95.45	90.00	100.00
		4	89.75	0.86	84.21	94.12	94.81	90.14	92.22	80.58	82.00	95.35	90.00	100.00
		5	90.50	0.87	84.21	93.02	96.30	91.55	93.33	81.55	82.00	97.62	90.00	100.00
		6	90.00	0.86	85.26	93.10	96.30	90.28	91.11	82.83	78.00	97.50	93.33	93.33
	D	1	91.25	0.88	88.42	94.38	95.56	93.48	92.22	86.46	92.00	85.19	76.67	100.00
		2	91.50	0.88	89.47	93.41	95.56	93.48	92.22	87.37	92.00	86.79	76.67	100.00
		3	89.50	0.86	86.32	94.25	94.07	92.70	94.44	80.95	84.00	85.71	73.33	100.00
		4	89.25	0.85	85.26	94.19	94.07	92.70	94.44	80.19	84.00	85.71	73.33	100.00
		5	89.50	0.86	85.26	93.10	93.33	92.65	93.33	79.25	84.00	91.30	83.33	100.00
		6	90.50	0.87	85.26	93.10	95.56	91.49	91.11	82.83	86.00	91.49	86.67	100.00
SVR	A	1	81.50	0.74	86.32	93.18	94.81	70.33	90.00	86.17	40.00	100.00	50.00	93.75
		2	82.50	0.75	87.37	92.22	94.81	71.91	87.78	85.87	48.00	100.00	50.00	93.75
		3	80.50	0.73	85.26	94.19	94.07	72.57	87.78	79.00	42.00	87.50	46.67	93.33
		4	81.75	0.75	83.16	95.18	94.07	73.84	91.11	78.10	50.00	96.15	46.67	100.00
		5	82.50	0.76	84.21	94.12	94.81	74.85	92.22	79.81	50.00	96..15	46.67	100.00
		6	82.50	0.76	83.16	95.18	95.56	73.30	93.33	82.35	48.00	96.00	46.67	100.00
	B	1	83.25	0.77	84.21	95.24	94.07	73.84	93.33	84.85	44.00	100.00	66.67	86.96
		2	83.50	0.77	86.32	92.13	94.07	74.71	88.89	85.11	50.00	100.00	66.67	90.91
		3	82.50	0.76	85.26	94.19	94.07	74.27	87.78	80.61	46.00	100.00	66.67	90.91
		4	83.00	0.76	83.16	95.18	94.07	75.15	92.22	79.81	46.00	100.00	66.67	95.24
		5	83.75	0.77	83.16	95.18	95.56	75.88	93.33	81.55	46.00	100.00	66.67	95.24
		6	83.50	0.77	82.11	95.12	95.56	75.44	93.33	81.55	46.00	95.83	66.67	100.00
	C	1	83.25	0.77	87.37	94.32	94.07	73.41	92.22	85.57	42.00	100.00	63.33	90.48
		2	83.25	0.77	85.26	93.10	94.07	74.27	91.11	84.54	48.00	100.00	63.33	90.48
		3	82.75	0.76	83.16	94.05	93.33	76.83	90.00	78.64	50.00	92.59	66.67	90.91
		4	83.75	0.78	81.05	95.06	93.33	76.83	94.44	79.44	54.00	100.00	66.67	95.24
		5	84.25	0.78	84.21	94.12	94.07	77.91	93.33	80.00	52.00	100.00	66.67	95.24
		6	84.25	0.78	83.16	95.18	96.30	76.02	93.33	82.35	48.00	100.00	66.67	100.00
	D	1	83.00	0.77	82.11	96.30	94.07	75.15	93.33	80.77	48.00	96.00	63.33	90.48
		2	84.25	0.78	85.26	94.19	94.07	76.51	92.22	83.00	54.00	100.00	63.33	90.48
		3	84.00	0.78	83.16	95.18	93.33	78.26	94.44	79.44	52.00	96.30	66.67	90.91
		4	84.00	0.78	83.16	95.18	93.33	78.75	94.44	78.70	52.00	92.86	66.67	95.24
		5	83.75	0.78	82.11	95.12	94.81	78.05	94.44	80.19	52.00	86.67	60.00	100.00
		6	84.75	0.79	82.11	95.12	95.56	78.66	93.33	80.77	56.00	93.33	66.67	100.00

[1]CVR, with RapidEye variables; SVR, without RapidEye variables; A, 500 samples; B, 700 samples; C, 1,000 samples; D, 1,500 samples; 1, GetStatistic; 2, GetStatistic+AAT; 3, GetStatistic + AAT + STM; 4, GetStatistic + AAT + STM + MAC; 5, GetStatistic + AAT + STM + MAC + SSA; 6, best variables. AAT, annual aggregated time series; STM, season-Trend model; MAC, mean annual cycle; SSA, annual cycle based on singular spectrum analysis.

An extensive processing time is necessary to obtain all the analyses available in the greenbrown package, using Landsat images (approximately three weeks for the methods MAC and SSA). The set using only GetStatistic as temporal variables consumed the shorter processing time (four hours), and this is the data set that showed also the best results in classifications. Therefore, this fact evidences that other variables are not necessary in the classification process, which can reduce the processing time, allowing this methodology to be used for larger areas.

The quantitative increment of the temporal variables used showed that the classification performances were slightly reduced when CVR data were used. This reduction did not happen when using SVR data, whose accuracies displayed alternate results for the different sets of variables. However, the best results were obtained when using the variables of the set 6, which contains a number of minor variables.

The use of temporal variables is an alternative to improve the classification accuracy, and it was important to efficiently distinguish coffee cropping areas from other types of land cover (Figure 4). Coffee croppig areas were well differentiated from

native vegetation areas. Temporal data have a great potential to improve land cover classification, mainly if the characteristics of images with better resolutions are involved in the classification process (Jia et al., 2014). The present results are in agreement with those obtained by Jia et al. (2014), once the best results were obtained using CVR.

Zhu & Woodcock (2014) also used some variables (mean, frequency, intra-annual and interannual differences, seasonality) derived from time series for the classification of land use, and obtained successful results of accuracy, with 90.48% hit index. Vieira et al. (2012) also obtained successful results while mapping sugar-cane, using time series data, in conjunction with object-based classification, obtaining excellent accuracy results with 94% hits. Such results are corroborated by present study, showing the importance of these variables for classification of areas with high-spectral similarity.

There was no significant difference between the classifications, when comparing the sets of variables used (Table 2). However, the greater differences were observed in the results obtained with the CVR, by which the classifications showed a greater difference
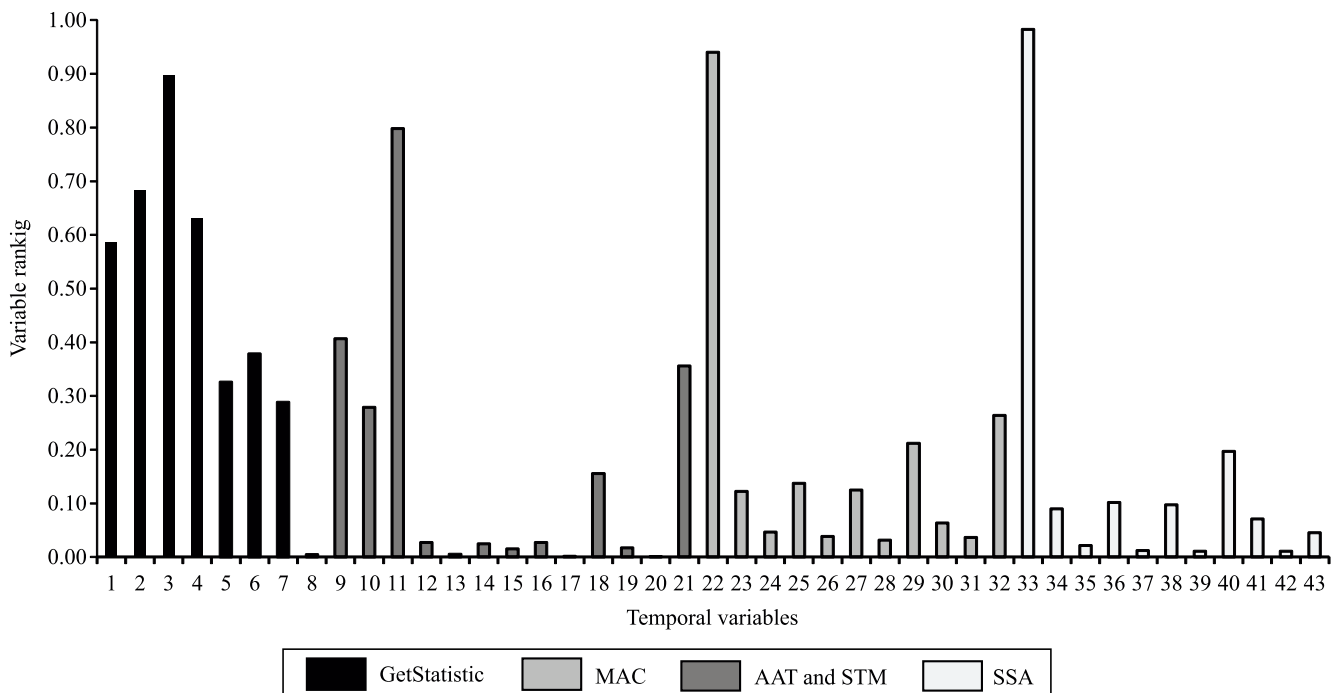
Figure 3. Ranking of variables used in the classification process obtained with the algorithms random forest and decision tree. AAT, annual aggregated time series; STM, season-trend model; MAC, mean annual cycle; SSA, annual cycle based on singular spectrum analysis.

between hit indexes; this dataset showed better hit indexes than the classifications using SVR.

The number of training samples had no influence over the index of hits. The best results were obtained using the set C, with 1,000 training samples, with a total of 20 times the number of variables used. Van Niel et al. (2005) sustain that such number may be up to 30 times the quantity of characteristics. Therefore, it is possible to affirm that the set of training samples used was efficient for attaining successful results in the classification process of the present study.

The use of temporal variables takes into account the characteristics of temporal dynamics (Vieira et al., 2012). In this study, the area has vegetation with different temporal variations; in that sense such parameters were relevant to maximize better classifications and separability between classes. The

results obtained showed the importance of these variables for a successful coffee classification, with indexes higher than the previously reported ones (Martínez-Verduzco et al., 2012; Santos et al., 2012; Sarmiento et al., 2014), even when using methodologies with higher accuracy; however, they require more time for mapping (Cordero-Sancho & Sader, 2007).

While researching in the Central-West region in Brazil, Mello et al. (2013) showed that variables derived from temporal data showed significant results in the mapping of agricultural areas. Variables resulting from temporal data have a considerable potential to represent modifications of spectral characteristics in the earth's surface over time, making them useful not only for classification purposes, but also for other applications such as understanding of the modifications in land cover (Mello et al., 2013).
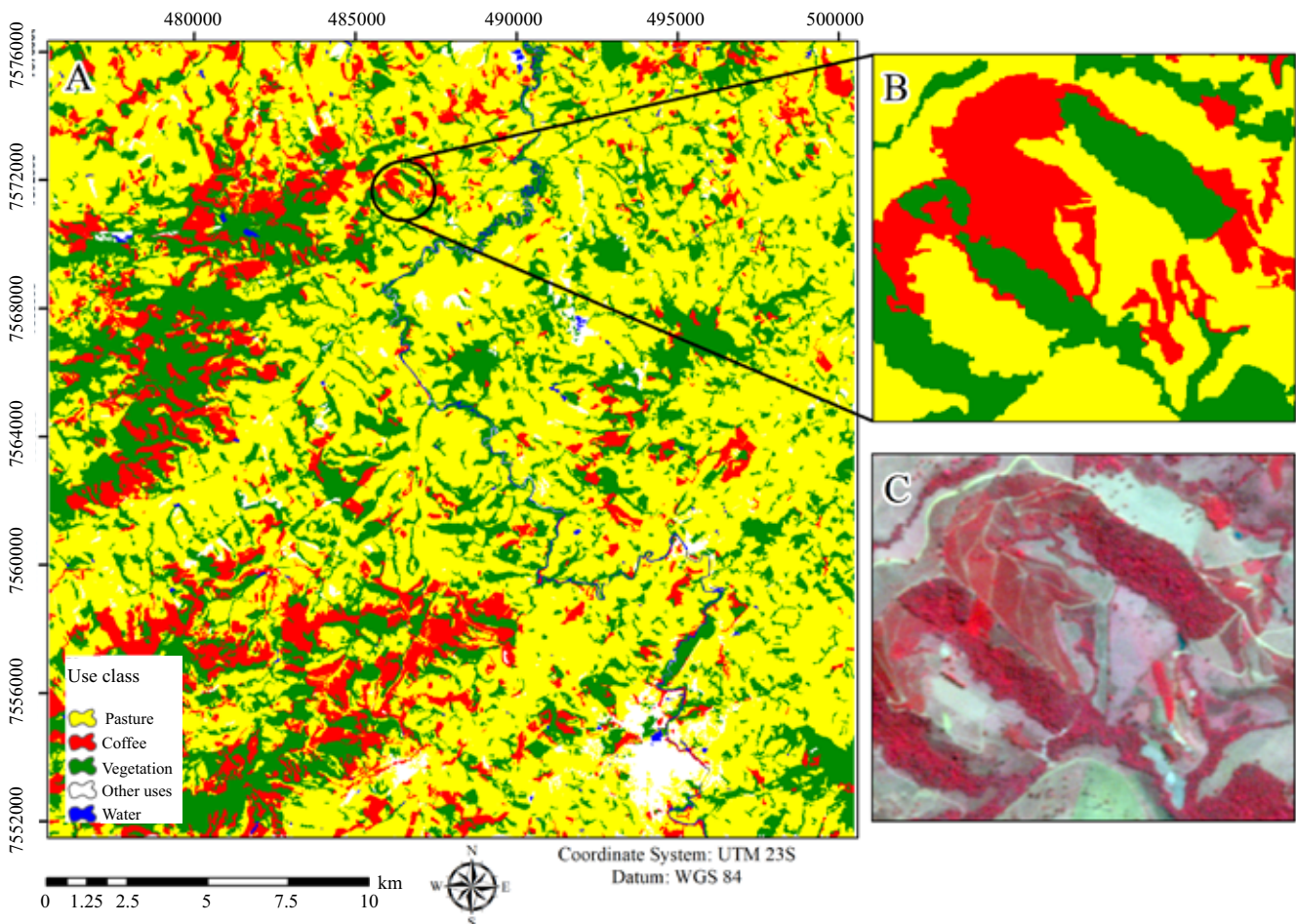
Figure 4. Classification: A, CVR-C1; B, detail of the classification map from an amplified area; and C, the corresponding detail in the RapidEye scene.

Very heterogeneous areas, such as the study area, are complex to be mapped (Li et al., 2014); however, this new method using temporal variables, displayed a feasible procedure that is easy to be implemented, reducing the confusion between coffee cropping areas and native vegetation. Moreover, this method is suitable for the mapping of different areas, and has a great potential for mapping coffee cropping areas and for separating land cover and land use classes.

To evaluate the performance of spectral variables from RapidEye, tests were performed after the exclusion of these data from the analysis of the variables. The results showed that the classifications that did not use such data obtained inferior performances than those that used these spectral data. However, it is important to emphasize that, even when not using these data, the results were superior to previous works mapping coffee cropping areas, as those by Martínez-Verduzco et al. (2012) and Cordero-Sancho & Sader (2007).

A similarity analysis was also performed between the two groups of classifications – CVR and SVR –, and the results (Table 3) showed that both groups had a significant difference between them (p<0.05 and chi-square> 3.81); therefore, these classifications may be considered as different.

Temporal data from the greenbrown package were significant for a successful classification of land cover. Similar results were found by Santos (2014), whose work showed that temporal data were significant for mapping reminiscent forests. Souza et al. (2016) used only spectral data for the mapping of coffee plantation areas in the same region of the present study, and those authors obtained 82% accuracy indexes, which is inferior to those of the present study. This indicates that even without the use of RapidEye variables, using only temporal variables, the results in the present study were superior. Nevertheless, when the variables RapidEye were inserted, the improvement of hit indexes was significant (Table 3), which shows the importance of combining spectral and temporal data to obtain a successful performance in the classification process.

Santos (2014) showed the efficiency of a temporal series Landsat TM for mapping data from RapidEye images, as verified in the present study. Carvalho & Carvalho (2015) evaluated the agreement of multisensor image use in an object-oriented approach. They identified a high degree of agreement from Rapideye and Landsat TM images, with 5 and 30 m spatial resolution, respectively. From these results, it

Table 2. Comparative matrices of classifications (A and B, lower diagonal; C and D, upper diagonal), showing the difference between the sets of variables, performed by the McNemar test with chi-square tabulated =3.81[1].

A and C	1	2	3	4	5	6	B and D	1	2	3	4	5	6
						CVR							
1	-	0.00	0.32	0.60	0.32	0.50	1	-	0.00	0.24	0.32	0.24	0.08
2	0.18	-	0.40	0.72	0.40	0.60	2	0.04	-	0.32	0.40	0.32	0.12
3	1.62	0.72	-	0.04	0.00	0.02	3	0.60	0.32	-	0.00	0.00	0.04
4	1.12	0.40	0.04	-	0.04	0.00	4	0.72	0.40	0.00	-	0.00	0.08
5	0.98	0.32	0.08	0.00	-	0.02	5	0.40	0.18	0.02	0.04	-	0.04
6	0.40	0.04	0.40	0.18	0.12	-	6	0.50	0.24	0.00	0.02	0.00	-
A and C	1	2	3	4	5	6	B and D	1	2	3	4	5	6
						SVR							
1	-	0.00	0.02	0.02	0.08	0.08	1	-	0.12	0.08	0.08	0.04	0.24
2	0.04	-	0.02	0.02	0.08	0.08	2	0.00	-	0.00	0.00	0.02	0.02
3	0.08	0.24	-	0.08	0.18	0.18	3	0.04	0.08	-	0.00	0.00	0.04
4	0.00	0.02	0.12	-	0.02	0.02	4	0.00	0.02	0.02	-	0.00	0.04
5	0.08	0.00	0.32	0.04	-	0.00	5	0.02	0.00	0.12	0.04	-	0.08
6	0.08	0.00	0.32	0.04	0.00	-	6	0.00	0.00	0.08	0.02	0.00	-

[1]CVR, with RapidEye variables; SVR, without RapidEye variables; A, 500 samples; B, 700 samples; C, 1,000 samples; D, 1,500 samples; 1, GetStatistic; 2, GetStatistic+AAT; 3, GetStatistic + AAT + STM; 4, GetStatistic + AAT + STM + MAC; 5, GetStatistic + AAT + STM + MAC + SSA; 6, best variables. AAT, annual aggregated time series; STM, season-Trend model; MAC, mean annual cycle; SSA, annual cycle based on singular spectrum analysis.

is possible to claim that the use of multisensor images with different spatial resolutions could be used to work on large-scale projects. Moreover, it is an alternative to work with a significant data number, like time series, which may be limited by the available images, thus bringing a great subsidy for the land cover monitoring and characterization. The results obtained in the present study evidenced that data derived from Landsat TM temporal series, composed by original and predicted images, were efficient to characterize objects derived from a RapidEye image with high-spatial resolution, obtaining excellent accuracy indexes and being able to separate, in an adequate manner, coffee cropping areas from other types of land cover.

A new strategy of coffee mapping, based in spectral and temporal data from remote sensing data, was proposed and showed great potential.

Table 3. Comparison of classifications showing the differences between sets using RapidEye variables (CVR), and without RapidEye variables (SVR), performed by the McNemar test with tabulated Z and tabulated chi-square (χ^2) =3.81; bold numbers show the significant results, at 5% probability.

CVR x SVR	Z	χ^2	Significant
CVR-A1 x SVR-A1	**2.96**	**8.82**	Yes
CVR-A2 x SVR-A2	**2.33**	**5.44**	Yes
CVR-A3 x SVR-A3	**2.64**	**7.00**	Yes
CVR-A4 x SVR-A4	1.83	3.38	Not
CVR-A5 x SVR-A5	1.69	2.88	Not
CVR-A6 x SVR-A6	**2.05**	**4.20**	Yes
CVR-B1 x SVR-B1	**2.33**	**5.44**	Yes
CVR-B2x SVR-B2	**2.05**	**4.20**	Yes
CVR-B3x SVR-B3	**2.17**	**4.71**	Yes
CVR-B4x SVR-B4	1.55	2.42	Not
CVR-B5 x SVR-B5	1.55	2.42	Not
CVR-B6 x SVR-B6	1.55	2.42	Not
CVR-C1 x SVR-C1	**2.61**	**6.84**	Yes
CVR-C2 x SVR-C2	**2.68**	**7.22**	Yes
CVR-C3 x SVR-C3	**2.48**	**6.18**	Yes
CVR-C4 x SVR-C4	1.69	2.88	Not
CVR-C5 x SVR-C5	1.76	3.12	Not
CVR-C6 x SVR-C6	2.64	2.64	Not
CVR-D1 x SVR-D1	**2.33**	**5.44**	Yes
CVR-D2 x SVR-D2	**2.05**	**4.20**	Yes
CVR-D3 x SVR-D3	1.81	3.27	Not
CVR-D4 x SVR-D4	1.48	2.20	Not
CVR-D5 x SVR-D5	1.62	2.64	Not
CVR-D6 x SVR-D6	1.55	2.42	Not

This methodology can be recommended for mapping other coffee cropping areas, and to obtain information on productive systems and improve the estimation of plantation areas, crop areas, production, spatial distribution, and the surrounding environment. This way, we can have the opportunity to improve environmental project strategies and management practices, supporting the planning of municipal, state, and national government.

Further studies should be completed using this methodology for different areas. Different variables may be incorporated into the classification process, as for instance, physical data, like topography (Prado et al., 2016), hypsometry, and others.

Conclusions

1. Temporal data, associated with an object-oriented approach, is efficient for classifying coffee cropping areas, and separating land cover classes, reducing the existing confusion between targets.

2. The variables derived from the greenbrown package, used as input data in the classification process, are efficient for the mapping of coffee cropping areas, as they reduce the confusion level between native vegetation and coffee, as well as pasture and coffee, and favors the mapping of these areas with greater precision.

3. Statistical variables (Getstatistic) perform better for mapping, increasing the success index of classifications; and the increment of the number of variables reduces the percentage of success of classifications, showing that best results can be obtained by the Getstatistic, using also spectral variables.

4. Data derived from a Landsat TM time series are efficient, contributing to a correct characterization and mapping of objects derived from a RapidEye image, with high-spatial resolution.

Acknowledgments

References

BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. **Journal of Applied Econometrics**, v.18, p.1-22, 2003. DOI: https://doi.org/10.1002/jae.659.

BERNARDES, T.; MOREIRA, M.A.; ADAMI, M.; GIAROLLA, A.; RUDORFF, B.F.T. Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. **Remote Sensing**, v.4, p.2492-2509, 2012. DOI: https://doi.org/10.3390/rs4092492.

BROWN, J.C.; KASTENS, J.H.; COUTINHO, A.C.; VICTORIA, D. de C.; BISHOP, C.R. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. **Remote Sensing of Environment**, v.130, p.39-50, 2013. DOI: https://doi.org/10.1016/j.rse.2012.11.009.

CARVALHO, N.S. de; CARVALHO, L.M.T. de. Viabilidade do uso de imagens multi-sensor: Grau de concordância enter os sensores Rapideye, Landsat/TM e MODIS. In: CONGRESSO MINEIRO DE ENGENHARIA E TECNOLOGIA, 1., Lavras, 2015. **Engenharia e Tecnologia para o Desenvolvimento Nacional**: anais. Lavras: UFLA, 2015. p.1227-1235. I COMET. Editores: Ednilton Tavares de Andrade e Marcelo de Carvalho Alves.

CORDERO-SANCHO, S.; SADER, S.A. Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. **International Journal of Remote Sensing**, v.28, p.1577-1593, 2007. DOI: https://doi.org/10.1080/01431160600887680.

FOODY, G.M. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. **Photogrammetric Engineering & Remote Sensing**, v.70, p.627-633, 2004. DOI: https://doi.org/10.14358/PERS.70.5.627.

FORKEL, M.; CARVALHAIS, N.; VERBESSELT, J.; MAHECHA, M.D.; NEIGH, C.S.R.; REICHSTEIN, M. Trend change detection in NDVI time series: effects of inter-annual variability and methodology. **Remote Sensing**, v.5, p.2113-2144, 2013. DOI: https://doi.org/10.3390/rs5052113.

FORKEL, M.; WUTZLER, T. **greenbrown**: land surface phenology and trend analysis. A package for the R software. Version 2.2. 2015. Available at: <http://greenbrown.r-forge.r-project.org/>. Accessed on: Aug. 28 2019.

GALFORD, G.L.; MUSTARD, J.F.; MELILLO, J.; GENDRIN, A.; CERRI, C.C.; CERRI, C.E.P. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. **Remote Sensing of Environment**, v.112, p.576-587, 2008. DOI: https://doi.org/10.1016/j.rse.2007.05.017.

GAO, F.; MASEK, J.; SCHUWALLER, M.; HALL, F. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. **IEEE Transactions on Geoscience and Remote Sensing**, v.44, p.2207-2218, 2006. DOI: https://doi.org/10.1109/TGRS.2006.872081.

GÓMEZ, C.; WHITE, J.C.; WULDER, M.A. Optical remotely sensed time series data for land cover classification: a review.

ISPRS Journal of Photogrammetry and Remote Sensing, v.116, p.55-72, 2016. DOI: https://doi.org/10.1016/j.isprsjprs.2016.03.008.

HUANG, C.; GOWARD, S.N.; MASEK, J.G.; THOMAS, N.; ZHU, Z.; VOGELMANN, J.E. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. **Remote Sensing of Environment**, v.114, p.183-198, 2010. DOI: https://doi.org/10.1016/j.rse.2009.08.017.

JIA, K.; LIANG, S.; ZHANG, N.; WEI, X.; GU, X.; ZHAO, X.; YAO, Y.; XIE, X. Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v.93, p.49-55, 2014. DOI: https://doi.org/10.1016/j.isprsjprs.2014.04.004.

JIA, K.; WU, B.; LI, Q. Crop classification using HJ satellite multispectral data in the North China Plain. **Journal of Applied Remote Sensing**, v.7, art.073576, 2013. DOI: https://doi.org/10.1117/1.JRS.7.073576.

LI, C.; WANG, J.; WANG, L.; HU, L.; GONG, P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. **Remote Sensing**, v.6, p.964-983, 2014. DOI: https://doi.org/10.3390/rs6020964.

MACHADO, M.L.; ALVES, H.M.R.; VIEIRA, T.G.C.; FERNADES-FILHO, E.I.; LACERDA, M.P.C. Mapeamento de áreas cafeeiras (*Coffea arabica* L.) da Zona da Mata Mineira usando sensoriamento remoto. **Coffee Science**, v.5, p.113-122, 2010.

MARTÍNEZ-VERDUZCO, G.C.; GALEANA-PIZAÑA, J.M.; CRUZ-BELLO, G.M. Coupling community mapping and supervised classification to discriminate shade coffee from natural vegetation. **Applied Geography**, v.34, p.1-9, 2012. DOI: https://doi.org/10.1016/j.apgeog.2011.10.001.

MASEK, J.G.; VERMOTE, E.F.; SALEOUS, N.E.; WOLFE, R.; HALL, F.G.; HUEMMRICH, K.F.; GAO, F.; KUTLER, J.; LIM, T.-K. A Landsat surface reflectance dataset for North America, 1990–2000. **IEEE Geoscience and Remote Sensing Letters**, v.3, p.68-72, 2006. DOI: https://doi.org/10.1109/LGRS.2005.857030.

MELLO, M.P.; VIEIRA, C.A.O.; RUDORFF, B.F.T.; APLIN, P.; SANTOS, R.D.C.; AGUIAR, D.A. STARS: A new method for multitemporal remote sensing. **IEEE Transactions on Geoscience and Remote Sensing**, v.51, p.1897-1913, 2013. DOI: https://doi.org/10.1109/TGRS.2012.2215332.

OIC. ORGANIZAÇÃO INTERNACIONAL DO CAFÉ. **Relatório sobre o mercado do café**. 2018. Available at: <http://www.ico.org/documents/cy2018-19/cmr-1218-p.pdf>. Accessed on: May 5 2019.

PRADO, D.F.C. do; HAYAKAWA, E.H.; KAWAKUBO, F.S. Mapeamento do cultivo de café no Sul de Minas Gerais utilizando imagens Landsat-5 TM e variáveis topográficas. **Revista do Departamento de Geografia**, v.31, p.43-57, 2016. DOI: https://doi.org/10.11606/rdg.v31i0.103040.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014. Available at: <http://www.R-project.org/>. Accessed on: August 5, 2017.

SANTOS, J.A. dos; GOSSELIN, P.-H.; PHILIPP-FOLIGUET, S.; TORRES, R. da S.; FALCÃO, A.X. Multiscale classification of remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, v.50, p.3764-3775, 2012. DOI: https://doi.org/10.1109/TGRS.2012.2186582.

SANTOS, P.A. dos. **Avaliação do uso de variáveis temporais na classificação da cobertura da terra**. 2014. 80p. Dissertação (Mestrado) - Universidade Federal de Lavras, Lavras.

SARMIENTO, C.M.; RAMIREZ, G.M.; COLTRI, P.P.; SILVA, L.F.L. e; NASSUR, O.A.C.; SOARES, J.F. Comparação de classificadores supervisionados na discriminação de áreas cafeeiras em Campos Gerais - Minas Gerais. **Coffee Science**, v.9, p.546-557, 2014.

SOUZA, C.G.; CARVALHO, L.; AGUIAR, P.; ARANTES, T.B. Algoritmos de aprendizagem de máquina e variáveis de sensoriamento remoto para o mapeamento da cafeicultura. **Boletim de Ciências Geodésicas**, v.22, p.751-773, 2016. DOI: https://doi.org/10.1590/S1982-21702016000400043.

USGS. United States Geological Survey. **EarthExplorer**. Available at: <http://earthexplorer.usgs.gov/>. Accessed on: Aug. 28 2019a.

USGS. United States Geological Survey. **Modis Reproject Tools Web**. Available at: <https://mrtweb.cr.usgs.gov/>. Accessed on: Aug. 28 2019b.

VAN NIEL, T.G.; MCVICAR, T.R.; DATT, B. On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. **Remote Sensing of Environment**, v.98, p.468-480, 2005. DOI: https://doi.org/10.1016/j.rse.2005.08.011.

VELOSO, H.P.; RANGEL FILHO, A.L.R.; LIMA, J.C.A. **Classificação da vegetação brasileira, adaptada a um sistema universal**. Rio de janeiro: IBGE, 1991. 124p.

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, v.114, p.106-115, 2010. DOI: https://doi.org/10.1016/j.rse.2009.08.014.

VIEIRA, M.A.; FORMAGGIO, A.R.; RENNÓ, C.D.; ATZBERGER, C.; AGUIAR, D.A.; MELLO, M.P. Object based image analysis and data mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. **Remote Sensing of Environment**, v.123, p.553-562, 2012. DOI: https://doi.org/10.1016/j.rse.2012.04.011.

ZHU, Z.; WOODCOCK, C.E. Continuous change detection and classification of land cover using all available Landsat data. **Remote Sensing of Environment**, v.144, p.152-171, 2014. DOI: https://doi.org/10.1016/j.rse.2014.01.011.