EFFECTIVENESS OF BFAST ALGORITHM TO CHARACTERIZE TIME SERIES OF DENSE FOREST, AGRICULTURE AND PASTURE IN THE AMAZON REGION

Tássia Borges Arantes
Universidade Federal de Lavras/UFLA, Departamento de Ciências Florestais/DCF, Lavras, MG, Brazil
Corresponding author: E-mail: tassiabarantes@gmail.com

Michel Eustáquio Dantas Chaves
Universidade Federal de Lavras/UFLA, Departamento de Engenharia/DEG, Lavras, MG, Brazil
E-mail: micheldchaves@gmail.com

Rafael Lemos Bastos
Universidade Federal de Lavras/UFLA, Departamento de Estatística e Experimentação Agropecuária, Lavras, MG, Brazil
E-mail: iel2702@yahoo.com.br

Luis Marcelo Tavares de Carvalho
Universidade Federal de Lavras/UFLA, Departamento de Ciências Florestais/DCF, Lavras, MG, Brazil
E-mail: passarinho@dcf.ufla.br

Marcelo Silva de Oliveira
Universidade Federal de Lavras/UFLA, Departamento de Estatística e Experimentação Agropecuária, Lavras, MG, Brazil
E-mail: marcelo.oliveira@dex.ufla.br

Abstract: Vegetation is one of the most important components of ecosystems, attracting attention and interest of the scientific community due to its undergoing constant transformation. The remote sensing systems provide data to detect, identify, map and monitor these changes. This study aimed at (1) evaluating the effectiveness of the BFAST algorithm to characterize time series of dense forest, agriculture and pasture in the Amazon region; (2) performing statistical tests in order to compare these series, and (3) fitting models to predict future values. By using the cumulative sums test, the time series of the three classes of land use were statistically different from each other, when comparing in pairs. As the series were different, the time series analysis of remote sensing data was useful in the identification and classification of different types of land use. The use of adjusted models to predict future values of the time series has proven effective for the use of Agriculture and Pasture, but not for the Forest class. It is concluded that the BFAST algorithm characterization of time series for the subsequent adjustment of models was useful for predicting harvests, considering the Agriculture use class.

Keywords: Vegetation dynamics; MODIS; Land use land cover.

Received: January, 30, 2017 - Accepted: August, 11, 2017

INTRODUCTION

The land cover has been undergoing a continuous process of change over time and space. In Mato Grosso State, in the central-west region of Brazil, the conversion of forests to agricultural use has occurred intensively in recent decades (Jasinski et al., 2005). Vegetation attracts attention and interest of the scientific community since it is considered one of the most important components of ecosystems, going through constant changes. According to Jensen (2009), much effort has been giving to the development of sensors and digital signal processing algorithms and modeling images to extract important information of the biophysical vegetation from orbital data.

In order to reduce the impact of this problem and improve policies planning, the scientific community has developed important technological projects related to the detection and indication of gradual changes in the phenological development in the field. The result has been the development of tools directly related to monitoring the productivity of crops that helps with decisionmaking in management. Many of these efforts are related to the application of remote sensing of the earth's surface, the temporal analysis of vegetation indexes in monitoring and changes detection.

The remote sensing systems provide data to detect, identify, map and monitor changes in ecosystems through multi-temporal and multi-spectral techniques (Martínez & Gilabert, 2009). Some changes occur suddenly and are caused by human activities such as forest cutting for monoculture or urban expansion, which can be easily characterized by a pair of images acquired before and after the change. On the other hand, gradual changes in the earth's surface are difficult to detect using only bi-temporal data (Coppin et al., 2004). Therefore, long and dense time series are required for an adequate characterization of incremental dynamic processes, such as the development of vegetation, seasonality, degradation and regeneration.

Furthermore, within this context, the MODIS (Moderate Resolution Imaging Spectroradiometer) has excelled in the studies of vegetation due to its daily frequency imaging of the entire globe, high geometric quality of the images and the presence of a sophisticated procedure for atmospheric correction (Justice et al., 1998). The MODIS application is directly related to vegetation through indexes, which represents advances in monitoring and generating information about the development of crops, improving the estimates by monitoring the phenological cycle and the history of changes on the Earth's surface.

The Enhanced Vegetation Index (EVI) is generated through the MODIS sensor. This vegetation index was developed to minimize constraints resulting from the saturation of an orbital image, observed mainly in vegetated areas, influenced by atmospheric effects, the substrate and the acquisition geometry (Huete et al., 2002). The index is obtained by combining the bands of red (0.62-0.67μm), near infrared (0.841 to 0.876μm), and also blue (0.459-0.479μm) (Huete et al., 1999), ranging from -1 and 1. Values close to -1 correspond to flooded areas or clouds, close to zero represent the non-existent or very sparse vegetation and close to 1 represent well-developed vegetation (Ponzoni & Shimabukuro, 2010).

Daily values of vegetative vigor are obtained due to the frequency of passage of the sensor, which enables time series analysis.

Recently, a new approach of time series processing was proposed, the BFAST (Breaks For Additive Seasonal and Trend). This method decomposes the time signature on three components: trend, seasonality and noise (Verbesselt et al., 2010a). It uses the STL procedure - Seasonal-Trend decomposition, which gives an accurate estimative, robust trend and seasonal components because of its ability to cope with extreme values or lack of values within the historical series (Verbesselt et al., 2010a, 2010b).

The BFAST was chosen for this study because of its easy implementation, through the package "bfast" free R software (R Core Team, 2016). Furthermore, it is a robust algorithm that can be applied to any type of sensor, which, in addition to decompose the series in the components of seasonality and trend, detects gradual and abrupt changes in the series.

Nevertheless, for a more detailed analysis of the dynamics of land use, it is important to know if the time signatures of different classes of land use are generated by the same stochastic process or not; meaning that these time series are statistically equal or different, assisting the identification and characterization of different land uses.

The hypothesis that guides this study is that it is possible to develop a precise model for the identification of different land uses by using time series analysis of products that are composed of vegetation indexes through the MODIS sensor, which directly assess the spectral behavior of vegetation during the course of the phenological cycle.

Therefore, this study aimed at:

a) evaluating the effectiveness of the BFAST algorithm (Breaks For Additive Season and Trend) to characterize the time signatures, based on objects from three different land use classes in the Amazon region: dense forest, agriculture and grazing, using time series of MODIS images;

b) carrying out statistical tests to compare the time series of the three land use classes;

c) fitting models to predict future values of the series.

MATERIAL AND METHODS

The study area is located in the north of Mato Grosso State, Brazil, presented in the Figure 1:

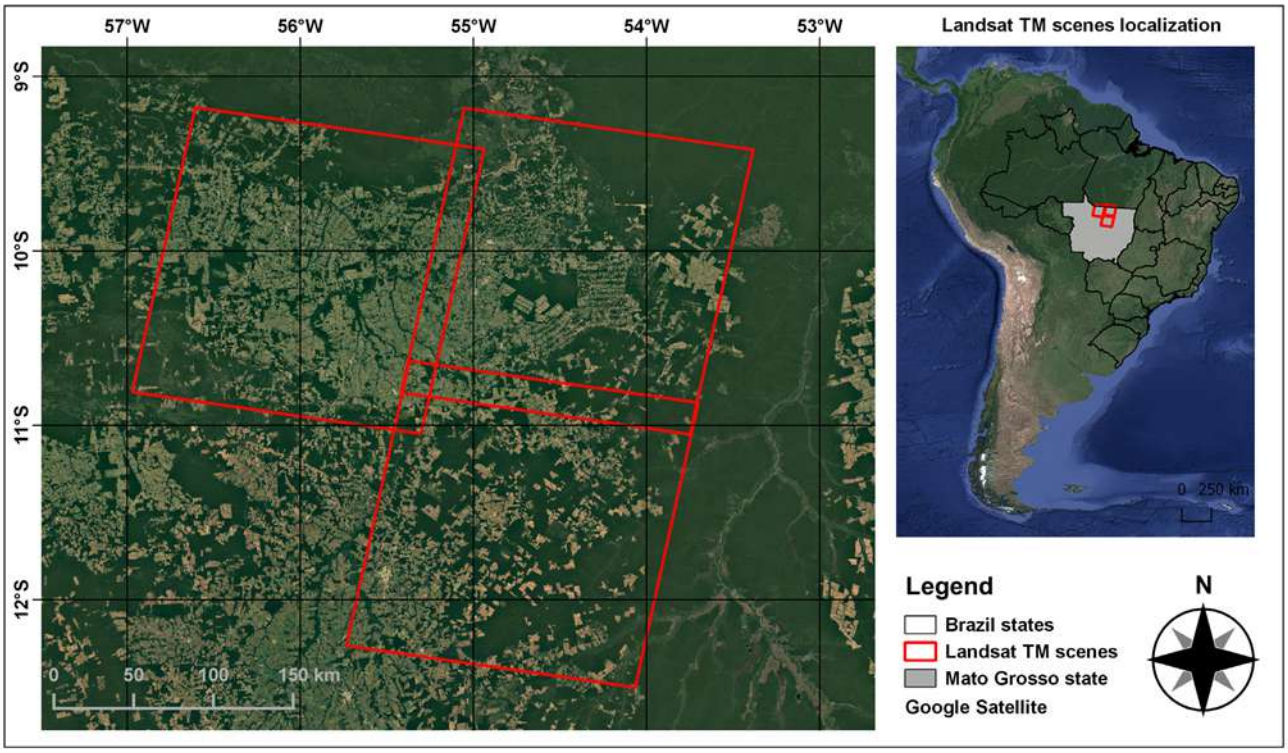

Figure 1: Study area at north of Mato Grosso, Brazil. Landsat scenes' detail, where: A) path/row 227/67, 07-31-2010; B) path/row 226/67, 07-24-2010; C) path/row 226/68, 07-31-2010.

The images from MODIS EVI index (Enhanced Vegetation Index) were acquired between September 2000 and January 2011. There was a 16-day interval between each observation (23 observations per year), with a total of 237 images. Equation 1 calculates the EVI:

$$EVI = G\frac{IVP - V}{IVP + C1xV - C2xA + L'} \quad (1)$$

where IVP represents the reflectance in near infrared, V is the red band, A is the blue band, C1and C2 are coefficients of the aerosol resistance term for red and blue respectively, L is the soil-adjustment factor and G is a gain factor.

The values adopted for the EVI were L = 1, C1 = 6, C2 = 7.5 e G = 2.5, according to recommendations proposed by Justice et al. (1998).

For the time series processing, the averages of the objects of the segmentation of Landsat TM images of the same area were used. It was selected an object for each of the following cover classes: dense forest (herein named "Forest"), Agriculture and Pasture.

The time series were analyzed using the BFAST algorithm with a package available for R software: the Gretl software (GNU General Public License, 2007) and the Statistica software (Statsoft, 2004).

The GRETL software requires monthly data. Thereby, the average of two observations of each month was considered and for the months with only one observation, 125 observations were included. From there, all other procedures were followed.

Statistical comparison between time series

For a statistical comparison between time series of the selected objects, two statistical tests were used, comparing the series in pairs, in order to determine whether they were generated by the same stochastic process or not. The tests herein used were the cumulative sums test and a series different procedures, as described below:

Cumulative sum test

Proposed by Coates & Diggle (1986), the cumulative sum test consists in using the Kolmogorov-Smirnov (KS) statistics to test a statistic based on the periodogram of the series, testing the removal of uniform distribution U(0,1). In order to perform the test, it is necessary to calculate the periodogram of the series and its reason (J). Thereby, the values of zi, cj e oj, are calculated with i=1,…,m e j=1,…,m-1 and are presented in Equations 2, 3 and 4:

$$z = \ln\left(1 + J\right), \quad (2)$$

$$c_j = \sum_{i=1}^{j} z_i, \quad (3)$$

$$o_j = \frac{c_j}{c_m}. \quad (4)$$

Lastly, it is necessary to compare the oj distribution with the distribution U(0,1) trough the Kolmogorov-Smirnov test.

If the p-value is greater than α, the hypothesis accepted is that the same stochastic process generates the series. This test is suitable for small samples without normal distribution, since it is considered a nonparametric test (Gibbons, 1985). In this test, if the distance between the distributions is small, with only random deviations, H0 is accepted. If the cumulative distributions are very distant from one another at any point, H0 is rejected, suggesting that they are originated from different populations.

Time series comparison method

The procedure is as follows: the difference is made between the two series, verifying the existence of any trend and the seasonality of the series-difference using the Cox-Stuart and Fisher's tests, respectively (Morettin & Toloi, 2006). The next step is to verify if the waste is white noise, meaning that it is independent and identically distributed, with null average and constant variance, using the Box & Pierce test (1970).

If after checking by the above tests, the series-difference does not present any trend or seasonality and is considered a white noise, it can be concluded that the two series are considered equal (Costa & Sáfadi, 2010).

Models adjustment and predictions

In order to properly prepare time series predictions, it is necessary to fit a model to the series studied. The first step is to adjust the identification of the model made on the basis of autocorrelations and partial autocorrelations estimated to determine the values of p, q and d of the ARIMA (Autoregressive Integrated Moving Averages) (Morais, 2012). Therefore, autocorrelation of the series will be used to determine the model to be adjusted (preliminary estimates of the parameters to be adjusted). After the model adjustment, it is necessary to examine whether their errors are white noise or not, which means that the model represents the data properly (Morais, 2012). In this study, the Box-Pierce test was used.

According to Morais (2012), the prediction is denoted by $\hat{Z}_t(h)$ for a series Z_{t+h}, for h=1,2,... and the prediction error is given by $e_t = Z_{t+h} - \hat{Z}_t(h)$, which Z_{t+h} is the real value and $\hat{Z}_t(h)$ is the predicted value.

RESULTS AND DISCUSSION

Time series analysis and characterization

Firstly, a visual analysis of graphs was made. In Figure 2, it is observed decomposed series in seasonal, trend and noise components.

The Figure 3 presented, in the same scale, the seasonal component of the classes:

It can be observed from Figure 2 that for the three land use categories (Forest, Agriculture and Pasture), seasonality is constant throughout the period studied, observed in the Figure 3. The same behavior occurs for the trend component of Forest class, with a slight negative slope. In Pasture class, the trend remains constant until 2009, where there is a drop in the EVI values followed by growth, which may have been caused by fire. For the class of Agriculture, there were three turning points in the slope over the years. These "breaks" in the trend line probably occurred because it is an annual crop, with tillage operations every year.

By verifying if the variance in the data is constant, using the amplitude graph in function of the average, it resulted that there is no correlation, indicating that there is no need to apply logarithmic transformation on the data to stabilize the variance.

For statistical comparison of the series, it is necessary to verify if they are stationary, which means that there is a significant trend in the data. However, if the data is not stationary, it could go through some simple processing, such as differentiation, in order to get stationary. In this verification, the Dickey-Fuller test (unit root test) was applied at a 5% significance level.

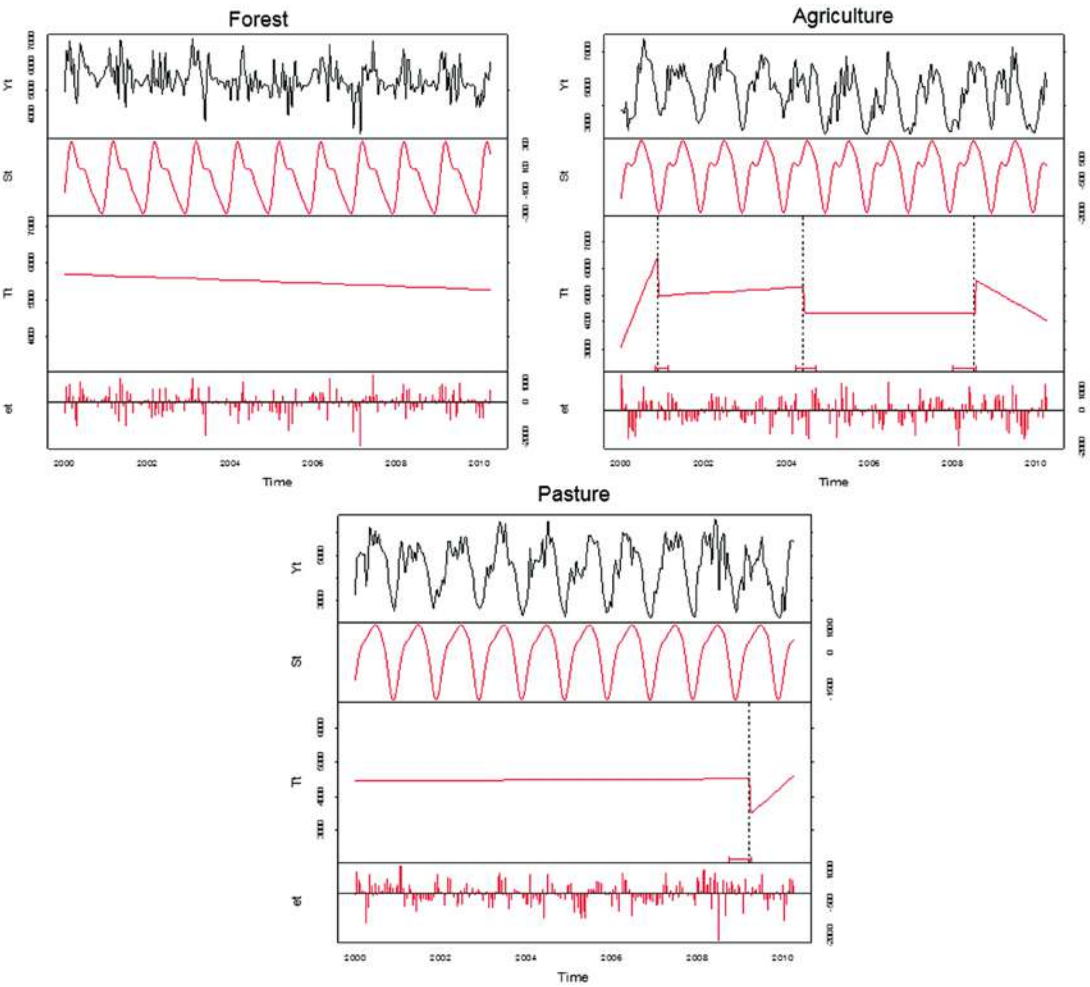

Figure 2: BFAST result for the three land use classes - Forest, Agriculture and Pasture, where Yt is the original data; St is the seasonal component; Tt is the trend component and et is the noise component.

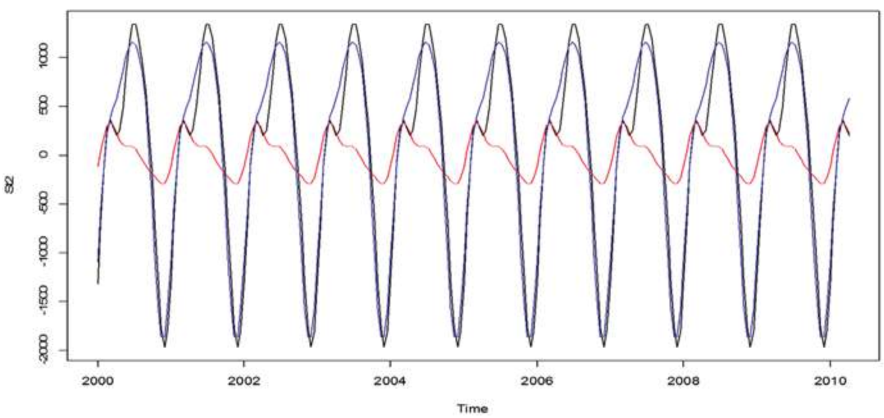

Figure 3: EVI values of the seasonal component of the three clases. Forest (red line), Agriculture (black line) and Pasture (blue line).

For the Forest class, the test showed that the series is stationary. However, for the classes of Agriculture and Pasture, the test showed that these series are not stationary. Thereby, the first-difference was applied for the two series and when the test was performed again, the data became stationary.

The periodogram of the series are shown in Figure 4. A peak is observed in period 12, approximately. Therefore, the Fisher test was applied in this period.

At 5% significance level, the test showed that there is seasonality only for the classes of Agriculture and Pasture. Therefore, a difference of 12 was applied to remove the seasonality of these series.

Comparison between series

Comparisons between the stationary series of the three different land use classes were: Forest vs. Agriculture; Forest vs. Pasture and Agriculture vs. Pasture. The results from each of the comparisons are shown next.

Forest vs. Agriculture

For the application of the cumulative sums test, the periodograms of the stationary series of Forest and Agriculture (Figure 5) were used.

In order to apply the test, the values of the periodograms of the stationary series were used, which was calculated as the ratio between the periodograms of the stationary series of Forest (numerator) and Agriculture (denominator). Thereafter, the values of de zi, cj e oj were calculated. The Kolmogorov-Smirnov test was used to compare the statistic oj of the cumulative sums test with distribution U(0,1). A p-value equals to $1.804e^{-13}$ were provided, rejecting the hypothesis of equality of the spectral density functions at 5% significance level. Therefore, the series of Forest and Agriculture are different and not generated by the same stochastic process.

The Cox-Stuart test was used, where the p-value was equal to 0.2522, at a 5% significance level, showing that the series difference presents

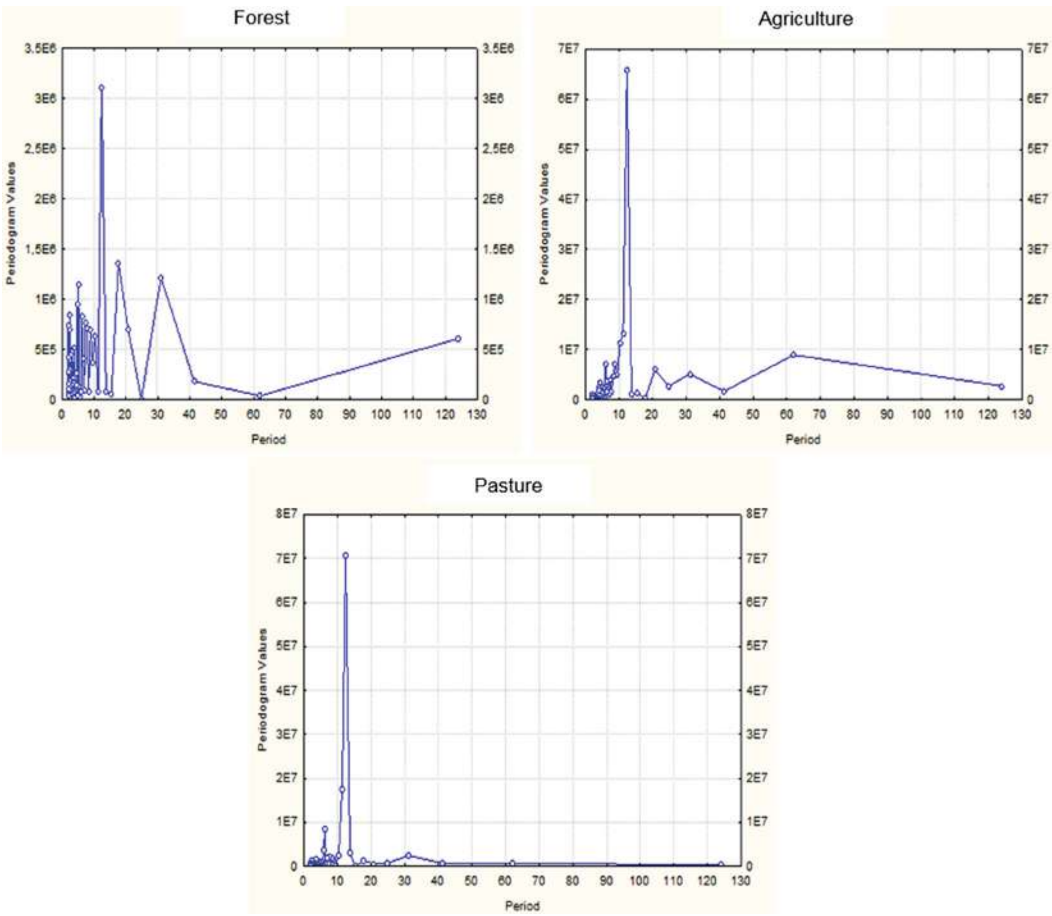

Figure 4: Periodogram of the monthly series of EVI values for the three land use classes – Forest, Agriculture and Pasture.

no trend. The Fisher's test was used to verify the seasonality of about 7 days, which did not occur.

The Box-Pierce test was used in order to verify if the difference series is considered a white noise and with $Q_{48} = 83.5710 > X^2_{48,0.05} = 62.8296$, the hypothesis that the difference series is a white noise is rejected, and therefore, the series are considered different.

Forest vs. Pasture

For the application of the cumulative sums test between Forest and Pasture, the periodogram of the stationary series of Pasture was acquired (Figure 6).

The values of de zi, cj e oj were calculated. The Kolmogorov-Smirnov test provided a p-value equals to $1.631e^{-13}$. Thereby, it rejects the hypothesis of equality of the spectral density functions at 5% significance level. Therefore, the Forest and Pasture series are different and are not generated by the same stochastic process.

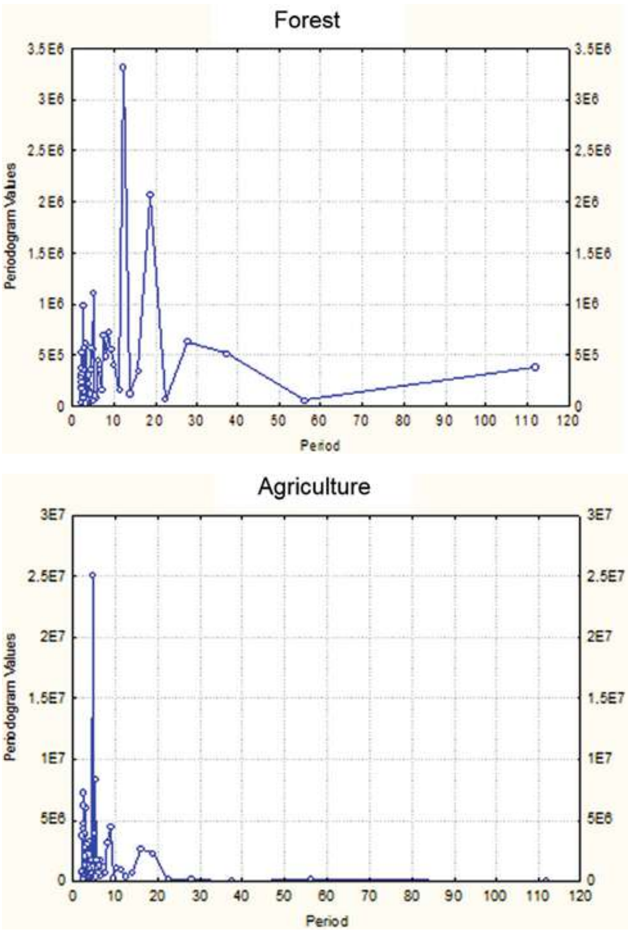

Figure 5: Stationary series periodogram of the of EVI values for the classes Forest and Agriculture.

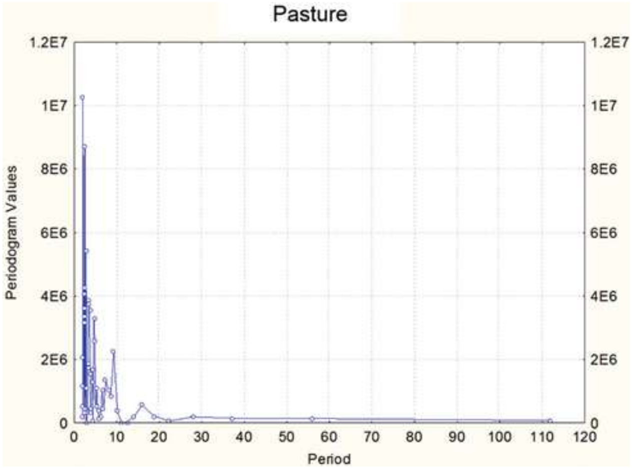

Figure 6: Stationary series periodogram of the of EVI values for the Pasture class.

Using the method proposed by Silva et al. (2000) for series comparison, it was obtained differences between the stationary series Forest and Pasture.

The Cox-Stuart test produced a p-value equal to 0.4469, indicating that there is no trend in the difference series at a 5% significance level. The Fisher's test was used to verify the seasonality of about 7 days, which did not occur.

The Box-Pierce test presented $Q_{48,0.05} = 55.6457 < X^2_{48,0.05} = 62.8296$, accepting the hypothesis that the series is considered white noise and concluding that the stationary series of Forest and Pasture are equal.

Agriculture vs. Pasture

For the cumulative sums test, the values of de zi, cj e oj were calculated. The Kolmogorov-Smirnov test provided a p-value equals to $1.887e^{-15}$. Thereby, it rejects the hypothesis of equality of the spectral density functions at 5% significance level. Therefore, the Agriculture and Pasture series are different and are not generated by the same stochastic process.

The Cox-Stuart test produced a p-value equivalent to 0.2522, indicating that there is no trend in the difference series at a 5% significance level. The Fisher's test was used to verify the seasonality of about 7 days, which did not occur.

The Box-Pierce test presented $Q_{48,0.05} = 117.4486 > X^2_{48} = 62.8296$, rejecting the hypothesis that the series is considered white noise and concluding that the stationary series of Agriculture and Pasture are different.

Model adjustment and prediction

Analyzing the correlogram and considering the orders of autoregressive operators and movable averages identified by FACP and FAC, respectively, the models were adjusted and selected according to their lower value of AIC (Akaike Information Criterion):

- Forest – ARIMA model (37, 0, 0) (AIC = 1870.889):

$$\left(1-\phi_1 B^1 - \phi_2 B^{36} - \phi_3 B^{37}\right) Z_t = a_t$$

- Agriculture – SARIMA model (3,1, 0) (1, 1, 0) (AIC = 1844.990):

$$(1 - \Phi B^{12})(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B^{12})(1 - B) Z_t = a_t$$

- Pasture – SARIMA model (0, 1, 1) (0, 1, 1) (AIC = 1755.026):

$$(1 - B^{12})(1 - B)Z_t = (1 - \Theta B^{12})(1 - \theta B)a_t$$

Tables 1, 2 and 3 shows the parameter estimates of the models proposed for the Forest, Agriculture and Pasture classes, respectively.

Table 1: ARIMA model parameters estimation proposed for the monthly series of EVI values for Forest class from September 2000 to January 2011.

Parameter	Estimate	Standard error
Constant	5504.790	77.839900
ϕ_1	0.184800	0.0846457
ϕ_2	0.232685	0.0909087
ϕ_3	0.234533	0.0954391

Table 2: ARIMA model parameters estimation proposed for the monthly series of EVI values for Agriculture class from September 2000 to January 2011.

Parameter	Estimate	Standard error
ϕ_1	-0.285053	0.0860709
ϕ_2	-0.292794	0.0851340
ϕ_3	-0.391918	0.0878760
Φ	-0.412260	0.0961857

Table 3: ARIMA model parameters estimation proposed for the monthly series of EVI values for Pasture class from September 2000 to January 2011.

Parameter	Estimate	Standard error
θ	-0.858831	0.0466562
Θ	-0.719185	0.0748421

Box-Pierce was applied to each model, giving the following results:

- Model (1): $Q(48) = 29.6525 < \mathcal{X}^2_{45,0.05} = 61.6562$
- Model (2): $Q(48) = 52.0206 < \mathcal{X}^2_{38,0.05} = 60.4809$
- Model (3): $Q(48) = 27.5562 < \mathcal{X}^2_{46,0.05} = 62.8296$

Therefore, all models were well adjusted. Figure 7 shows the residues autocorrelation.

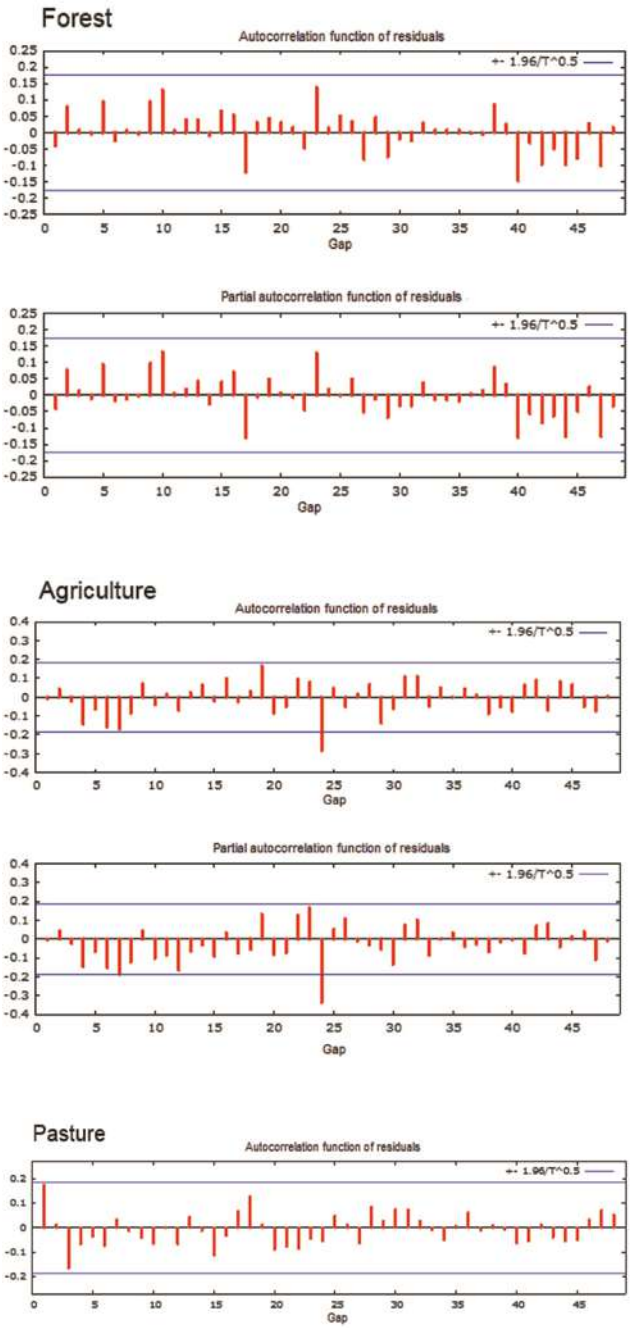

Figure 7: Autocorrelation (fac) and partial autocorrelation (facp) functions of the residuals of the models adjusted for the series of the three land use classes - Forest, Agriculture and Pasture.

The predictions were made from February 2011 to January 2012 for all land use classes, as shown in Figure 8.

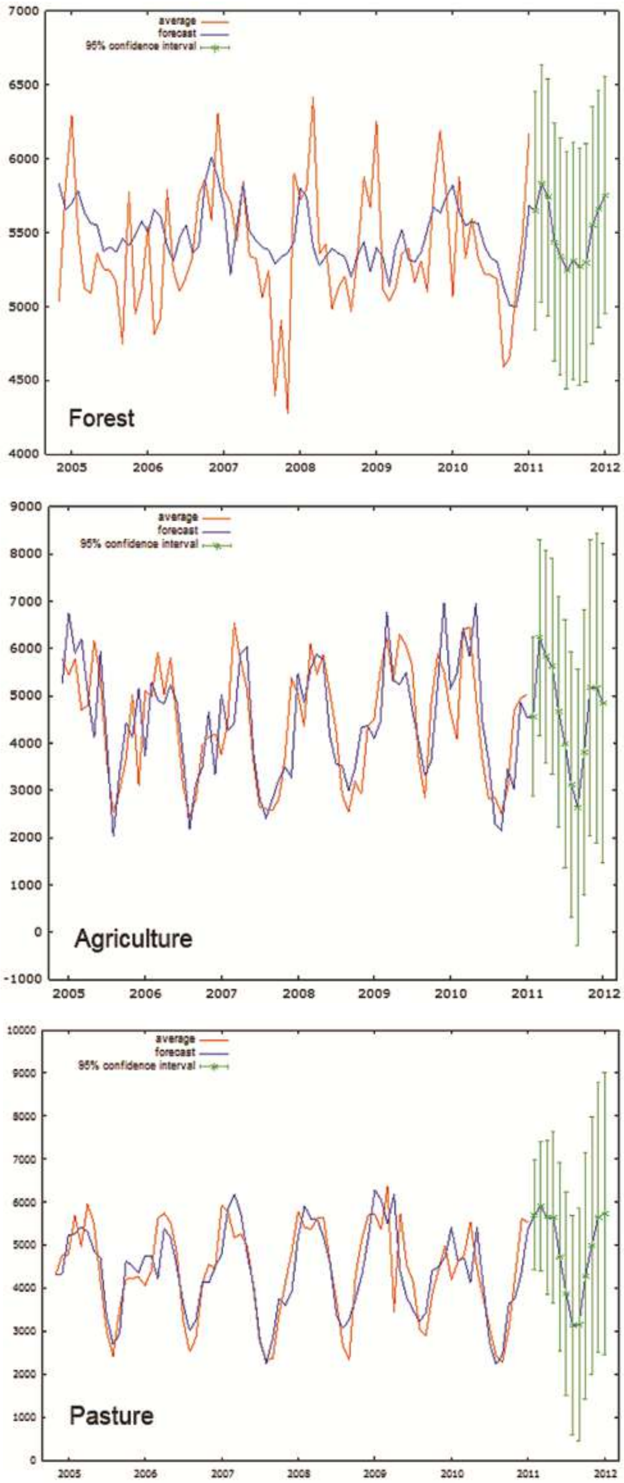

Figure 8: Observed and estimated series values according to models (1), (2) and (3) for the EVI values for the three land use classes - Forest, Agriculture and Pasture - respectively, until January 2011 and predicted values between February 2011 and January 2012.

From Figure 8, it was observed that Forest class model did not captured the real values of the series, but for the series of Agriculture and Pasture classes, the models were adequate.

CONCLUSIONS

The objectives of this study were achieved.

The BFAST algorithm was useful in characterizing the time series, allowing the observation of any changes (or not) both for the seasonal and the trend component. Using the procedure for comparing the series, only the Forest and Pasture classes showed the same behavior, statistically. However, using the cumulative sums test for comparison, the time series of the three land use classes were statistically different from each other, when compared in pairs. Therefore, as the series were different, the temporal analysis of remote sensing data was useful in the identification and classification of different types of land use.

The use of adjusted models to predict future values of the time series were effective for Agriculture and Pasture, but not for the Forest class. It is noteworthy that there is a saturation point of the values of the EVI index in areas of dense forest, which is a likely explanation for the failure of predicting future values of the time series.

It can be concluded that using the BFAST algorithm to characterize time series for subsequent adjustment models may be useful for predicting crop yields, considering the use of Agriculture class. Future research ought to be conducted with a larger number of data and locations, in order to better characterize and differentiate the types of land use.

ACKNOWLEDGEMENTS

This research project was supported by FAPEMIG and CAPES.

REFERENCES

BOX, G. E. P.; PIERCE, D. A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. **Journal of the American Statistical Association**, v. 65, n. 332, p. 1509–1526, 1970.

COATES, D. S.; DIGGLE, P. J. Tests for comparing two estimated spectral densities. **Journal of Time Series Analysis**, Clevedon, v. 7, n. 1, p.7-20, 1986.

COPPIN, P.; JONCKHEERE, I.; NACKAERTS, K.; MUYS, B. Review ArticleDigital change detection methods in ecosystem monitoring: a review. **International Journal of Remote Sensing**, v. 25, n. 9, p. 1565-1596, 2004.

COSTA, F. M.; SÁFADI, T. Statistical comparison of two series of particulate matter (PM10) in the city of São Paulo. **Rev. Bras. Biom**. São Paulo, v. 38, n. 3, p. 23-38, 2010.

GIBBONS, M. R.; FERSON, W. Testing asset pricing models with changing expectations an unobservable market portfolio. **Journal of Financial Economics**. v. 14, n. 2, p. 217-236, 1985.

GNU GENERAL PUBLIC LICENSE. Version 3, 29. Copyright (C) 2007 Free Software Foundation, Inc. <http://fsf.org/> June 2007.

HUETE, A., et al. A light aircraft radiometric package for MODL and Quick Airborne Looks (MQUALS). **Earth Observations**, p. 22–25, 1999 (NASA/GSFC).

HUETE, A., et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 195-213, 2002.

JASINSKI, E.; MORTON, D.; DEFRIES, R.; SHIMABUKURO, Y.; ANDERSON, L.; HANSEN, M. Physical Landscape Correlates of the Expansion of Mechanized Agriculture in Mato Grosso, Brazil. **Earth Interactions**, v. 9, n. 16, p. 1–18, 2005.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em Recursos Terrestres**. Traduação de J. C. N. Epiphanio. São José dos Campos, SP: Parênteses, 2009. 598 p. (Prentice Hall Series in Geographic Information Sciennce) Tradução de: Remote Sensing of the environment: na earth resource perspective.

Justice, C. O. et al. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 4, 1998.

MARTÍNEZ, B.; GILABERT, M. A. Vegetation dynamics from NDVI time series analysis using the wavelet transform. **Remote Sensing of Environment**, v. 113, n. 9, p. 1823–1842, 2009.

MORAIS, T. S. T. **Estudo temporal do nível médio do mar em diferentes oceanos**. 109 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária). Universidade Federal de Lavras, Lavras, MG, 2012.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2.ed. São Paulo: E. Brucher, 2006. 538p.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento Remoto no Estudo da Vegetação**. ed. São José dos Campos: Parêntese, 2010. 128 p.

R CORE TEAM R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2016.

SILVA, R. B. V.; FERREIRA, D. F.; SÁFADI, T. Modelos de séries temporais aplicados à série dos índices de preços ao consumidor na região de Lavras, MG, no período de 1992 a 1999. **Organizações Rurais e Agroindústriais**, Lavras, v. 2, n. 2, p. 44-55, 2000.

STATSOFT, Inc. **STATISTICA** (data analysis software system), version 7. www.statsoft.com. 2004.

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 1, p. 106–115, 2010a.

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 12, p. 2970-2980, 2010b.