

AUTOMATIC CLASSIFICATION OF LAND COVER CHANGE ASSOCIATED WITH THE BRAZILIAN SUGARCANE EXPANSION OVER THE LAST DECADE

Marcio Pupin Mello*, Marcos Adami, Daniel Alves Aguiar, Bernardo Friedrich Theodor Rudorff

National Institute for Space Research (INPE), Remote Sensing Division (DSR),
São José dos Campos, 12227-010, São Paulo State, Brazil.
*mello@ieee.org

ABSTRACT

MODIS time series have provided high quality data which allows retrieving land use and land cover change (LUCC) history related to sugarcane expansion in Brazil. Although visual interpretation is accurate, automatic classification methods for remotely sensed data should be tested mainly due to the extent of the Brazilian sugarcane area. Thus, this work aims at introducing automatic approaches to classify MODIS time series in order to retrieve LUCC history regarding sugarcane expansion over the last decade in Brazil. Results showed that about 65% and 20% of sugarcane expansion was over pasture and annual crop land, respectively. Although partial results indicated that automatic classification is a promising tool to analyze MODIS time series for LUCC studies, several improvements have to be considered before achieving reliable results.

Index Terms— Time series, MODIS, remote sensing, LUCC, Canasat Project.

1. INTRODUCTION

Remote sensing satellite imagery provides a valuable tool to monitor Brazil's environmental changes. MODIS sensors on board of Terra and Aqua satellites have provided high quality images that have been widely used in Land Use and Land Cover Change (LUCC) studies throughout the world [1]. MODIS sensors acquire images over the entire Earth on an almost daily basis generating a huge amount of data [2] that requires powerful computational tools to provide adequate statistical methods for information extraction and analyses.

Brazil's National Institute for Space Research (INPE) has recently developed a web tool integrating MODIS data for instant visualization for every pixel, over the South America continent, since the year 2000 [3]. Each curve of the time series represents the variation over time of the Two-Band Enhanced Vegetation Index (EVI-2) [4] for a user's selected pixel on the virtual globe of Google Maps. The time series were constructed based on filtered vegetation index

(EVI2) of the MOD13Q1 product (collection 5, 16 days composite at spatial resolution of 250m) available at NASA. Throughout MODIS time series, it is possible to retrieve LUCC history for a given pixel over the last decade, including those which represent areas where sugarcane has expanded, enabling researchers to evaluate the sustainability of this expansion in Brazil.

Due to its importance, sugarcane cultivation in the South-central region of Brazil has been monitored since 2003 through visual interpretation of Landsat type images by Canasat Project (www.dsr.inpe.br/laf/canasat/en/index.html), which provides sugarcane thematic maps with excellent accuracy yearly [5]. These annual maps have been used not only to estimate the cultivated sugarcane area but also as reference for several investigations such as the monitoring of sugarcane harvest practices [6] and crop yield analyses [7]. Based on Canasat Project thematic maps Adami et al. [8] have recently started the assessment of land use change in response to sugarcane expansion in the South-central of Brazil through visual interpretation of MODIS time series based on expert knowledge and also supported by Landsat type images (Fig 1). Although it is the most accurate classification method, visual interpretation is a laborious and time-consuming task, especially given the size of the Brazilian territory.

Within this context this work aims to introduce automatic approaches to classify MODIS time series in order to retrieve LUCC history related to sugarcane expansion over the last decade in Brazil.

2. METHODOLOGY

The South-central region of Brazil comprises the states of São Paulo, Minas Gerais, Paraná, Mato Grosso, Mato Grosso do Sul, Goiás, Rio de Janeiro, Espírito Santo, Santa Catarina, and Rio Grande do Sul. However, the states of Rio de Janeiro and Espírito Santo have a relatively small sugarcane area with low potential for expansion and the states of Santa Catarina and Rio Grande do Sul have even fewer sugarcane areas and, therefore, these states were not

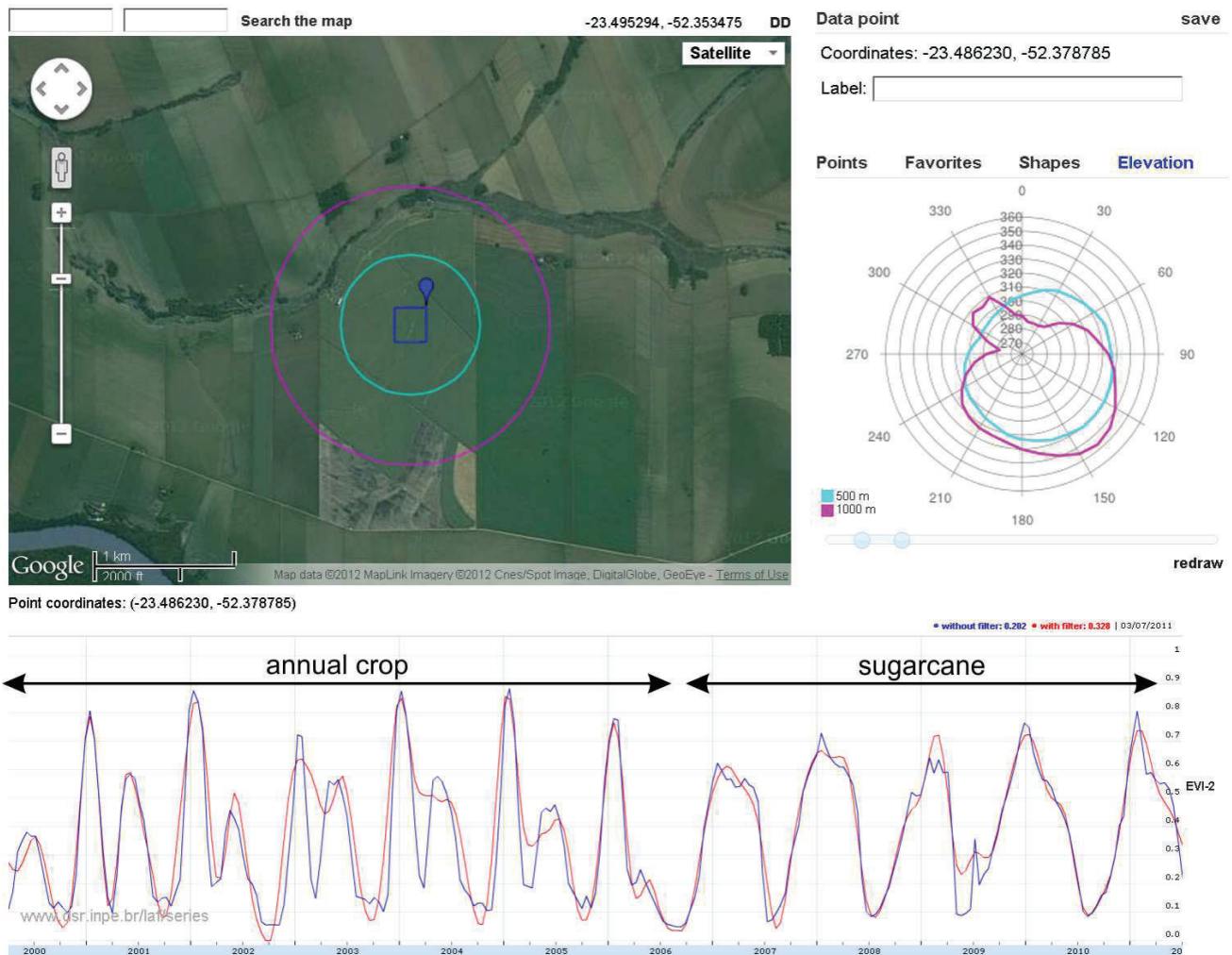


Fig. 1. The time series graph for the pixel (blue square indicated by the balloon) in this figure is from a field within an agricultural region in the north-west of Paraná State, Brazil. The visual analysis of the time series indicates that the land was originally annual crop with double crop cultivation, possibly soybean and corn. In 2006 the field was converted to sugarcane.

considered in the present study. The analysis of LUCC dynamics from 2000 to 2009 in the South-central region was carried out considering the sugarcane expansion observed for São Paulo State from crop years 2005/06 to 2010/11; and for the other states of the region from crop years 2007/08 to 2010/11 since the major expansions by state were observed during these crop years.

Firstly, 1035 MODIS time series were systematically sampled over recently sugarcane expansion fields based on Canasat Project thematic maps of sugarcane areas. Each time series was visually interpreted by assigning one of the following classes to each crop year: *pasture*, *annual crop*, *citrus*, *forest* and *sugarcane*. Multitemporal Landsat images were used to assist the visual interpretation procedure [8].

Several metrics extracted from the sampled MODIS time series were then extracted (minimum, maximum, mean, standard deviation, amplitude, etc) considering each crop year (i.e. from August to July) as a separate instance. These metrics were then used as input attributes for two tested classifiers: Instance Based (IB) and Decision Tree (DT). IB is based on the instances themselves, instead of on a model derived from labeled instances and, although computationally demanding, it is able to deal with practically any kind of data distribution [9]. IB classifiers consider a positive integer number (k) of nearest neighbors for labeling. We used *ipredknn* function of the package *ipred* [10] for R [11]. DT classifiers, in turn, can deal with categorical data together with numerical ones and sometimes can be easily interpretable. For DT, we used the *rpart* function of the *rpart* package [12]. The tuning of both classifiers was done as described by Sluiter and Pebesma [13], selecting $k=11$ for IB and $cp=0.1$ (*default*) for DT.

From the 1035 MODIS time series, two thirds were used for training the classifiers and the remaining third was used for accuracy assessment. The multinomial distribution defined the accuracy sample size, as recommended by Congalton and Green [14]. Eventually, the performances of classifiers were compared through a Z statistical test [14] using the estimates of kappa index and kappa variances [15].

3. PRELIMINARY RESULTS AND PARTIAL CONSIDERATIONS

Considering the crop year 2000/01, the classification results for both IB and DT classifiers showed that about 65% and 20% of sugarcane expansion was over pasture and annual crop land, respectively. The citrus and forest classes did not present a representative number of samples and were omitted by both classifiers. The remaining of the analyzed time series was sugarcane under crop rotation and overestimated sugarcane expansion areas.

The overall accuracy indices were 76.4 and 73.3% for IB and DT, respectively. According to the Z test, both classifiers presented similar performance ($Z_{\text{calculated}}=1.04$; $Z_{\text{tab}}=1.96$ for a confidence level of 95%). By comparing our results with those presented by Adami et al. [8] we stated that, although the overall accuracy indices for automatic classification indicated that this is a promising tool to analyze MODIS time series for land use and land cover change studies, there are still several improvements to be made.

Indeed other classifiers should be tested and more importantly, the classification should consider a transition matrix. For example, we know that an area covered by pasture in 2002 would not be forest in 2003, since forest regrowth takes more than one year. Thus a classification scheme which considers the probabilities of each transition would have a much better performance in studies involving

the use of satellite time series for land cover classification. Such improvements will be investigated in the future.

ACKNOWLEDGMENTS

The authors like to thank the financial support of the Brazilian Council CNPq (Conselho Nacional do Desenvolvimento Científico e Tecnológico – 153608/2010-2 and 142845/2011-6) and the team of the Laboratory of Remote Sensing in Agriculture and Forestry (LAF).

4. REFERENCES

- [1] B. Ramachandran, C.O. Justice, and M.J. Abrams, Eds., *Land remote sensing and global environmental change*. New York, NY, USA: Springer New York, 2011, p. 873.
- [2] C.O. Justice, J.R.G. Townshend, E.F. Vermote, E. Masuoka, R.E. Wolfe, N. Saleous, D.P. Roy and J.T. Morisette, “An overview of MODIS Land data processing and product status,” *Remote Sensing of Environment*, vol. 83, no. 1–2, pp. 3-15, 2002.
- [3] R. M. Freitas, E. Arai, M. Adami, A.S. Ferreira, F.Y. Sato, Y.E. Shimabukuro, R.R. Rosa, L.O. Anderson and B.F.T. Rudorff, “Virtual laboratory of remote sensing time series: Visualization of MODIS EVI2 data set over South America,” *Journal of Computational Interdisciplinary Sciences*, vol. 2, no. 1, pp. 57-68, 2011.
- [4] Z. Jiang, A.R. Huete, K. Didan, and T. Miura, “Development of a two-band enhanced vegetation index without a blue band,” *Remote Sensing of Environment*, vol. 112, no. 10, pp. 3833-3845, 2008.
- [5] M.P. Mello, M. Adami, D.A. Aguiar, and B.F.T. Rudorff, “Canasat Project accuracy assessment of sugarcane thematic maps,” in *Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2012)*. Florianópolis, SC, Brazil: ISASA, 2012. (in press).
- [6] D.A. Aguiar, B.F.T. Rudorff, W.F. Silva, M. Adami, and M. P. Mello, “Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in São Paulo State, Brazil,” *Remote Sensing*, vol. 3, no. 12, pp. 2682-2703, 2011.
- [7] L.M. Sugawara, “Variação interanual da produtividade agrícola da cana-de-açúcar por meio de um modelo agrônômico,” Ph.D. Thesis, National Institute for Space Research (INPE), São José dos Campos, SP, Brazil, 2010.
- [8] M. Adami, B.F.T. Rudorff, R.M. Freitas, D.A. Aguiar, and M.P. Mello, “Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in Brazil,” *Sustainability*, vol. 4, no. 4, pp. 574-585, 2012.

- [9] D.W. Aha, D. Kibler, and M.K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37-66, 1991.
- [10] A. Peters and T. Hothorn, *ipred: Improved Predictors*. 2012. [Online]. R package version 0.8-13. Available at: <http://CRAN.R-project.org/package=ipred> (assessed on Sep. 7th, 2011).
- [11] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012. [Online]. Available at: <http://www.R-project.org/> (assessed on May 1st, 2012).
- [12] T.M. Therneau and B. Atkinson, *rpart: Recursive Partitioning*. 2012. [Online]. R package version 3.1-51. Available at: <http://CRAN.R-project.org/package=rpart> (assessed on Sep. 7th, 2011).
- [13] R. Sluiter and E.J. Pebesma, "Comparing techniques for vegetation classification using multi- and hyperspectral images and ancillary environmental data," *International Journal of Remote Sensing*, vol. 31, no. 23, pp. 6143-6161, 2010.
- [14] R.G. Congalton and K. Green, *Assessing the accuracy of remotely sensed data: principles and practices*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2009, p. 183.
- [15] W.D. Hudson, "Correct formulation of the Kappa coefficient of agreement," *Photogrammetric Engineering & Remote Sensing*, vol. 53, no. 4, pp. 421-422, 1987.