Time series classification using features extraction to identification of use land and cover land: A case study in the municipality of Itaqui, South Region of Brazil

Adeline Marinho Maciel[1]
Lúbia Vinhas[1]

[1] National Institute for Space Research – INPE
Caixa Postal 515 – 12227-010 – São José dos Campos - SP, Brasil
{adeline.maciel, lubia.vinhas}@inpe.br

Abstract. One of the main applications in remote sensing is the analysis and classification of land cover and land use. Sensors, such as MODIS, are have been largely used for monitoring land cover change due to its high temporal resolution. Although several studies perform time series classification by features extracting or similarities measures to verify annual usage patterns and land cover, a little has been explored about the extraction of information by focal neighborhood operation and different sub-intervals of a time series whole. In order to explore the use of different features extracted of annual time series, for sub-intervals and focal neighborhood to identify patterns of land use and land cover, this work takes the use of statistical measures already extracted in the context of annual time series and presents an approach to information extraction for sub-intervals of year and focal neighborhood to characterise temporal patterns. To demonstrate the applicability of this study an experiments were conducted to classify of time series of land use and land cover using EVI MODIS sensor data, using Random Forest algorithm, where resulted in the creation of temporal maps identifying temporal patterns.

Keywords: time series, random forest, classification, features extraction, MODIS, zonal operation

1. Introduction

Earth Observation (EO) satellites provide a continuous and consistent set of information about the Earth's land and oceans. Using EO data sets, it is possible to detect long-term changes and trends in the environment, and measure the impacts of climate change, urban and ocean pollution and land expansion for food production. Vegetation indexes products such as NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index) are largely used to assess land cover specially to support land change studies. The MODIS sensor (Moderate Resolution Imaging Spectroradiometer) is the main source of NDVI and EVI data used in mappings of land change across large areas and long periods.

In the last decade, it is possible to observe an increase in the interest for EO time-series processing, resulting in new algorithms to execute classification, clustering or anomaly detection. Regarding classification of EO time series two approaches can be identified: based on features extraction and based on similarity. The first approach is based on the extract relevant features that characterise EO time-series of each unit of analysis, usually a pixel, and use a data mining algorithm to find groups of pixels with similar characteristics. The second approach uses temporal patterns that represent the classes of interest and use a distance function to decide to each class a pixel is more similar (RALANAMAHATANA et al., 2005).

To characterise the temporal pattern of land use/cover is not an easy task. The NDVI index is correlated with vegetation canopy greenness, a composite property of leaf area, chlorophyll and canopy structure. Plotting a time-series NDVI data produces a temporal curve that summarizes the various stages that green vegetation undergoes during a complete growing season.

It is possible to use the EVI/NDVI signal directly obtained from the input data set, or to extract other metrics or features from the time series to represent a temporal pattern. Several works use distinct features to classify time series (ZHANG et al., 2008; HüTTICH et al., 2009;

ISBN: 978-85-17-00088-1

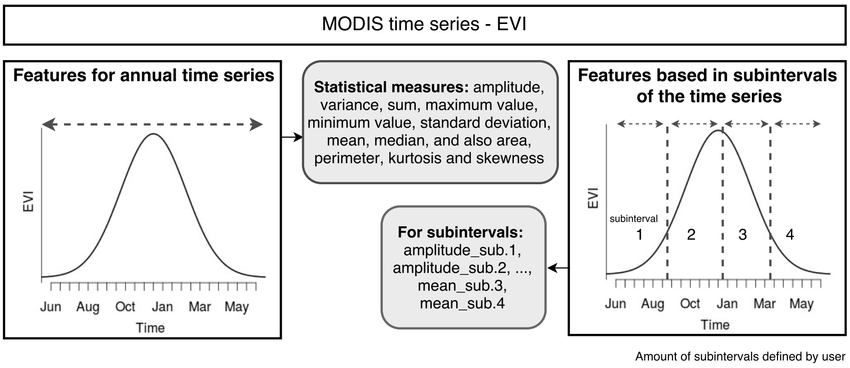
ARVOR et al., 2011; NITZE; BARRETT; CAWKWELL, 2015). This work addresses the problem of finding the set of features to characterise a temporal patterns, in order to extract more information from EO time series.

We discuss the use of different features extracted from time series to be used in land cover classification methods. We studied three sets of features: metrics from an annual time spam; metrics from sub-intervals in an annual time spam; and aggregated metrics in spatio-temporal neighborhood following the concept of focal operation from Tomlin's map algebra (TOMLIN, 1990). With this work we intend to contribute to the research in EO time series, regarding the characterisation of temporal patterns using different types of features.

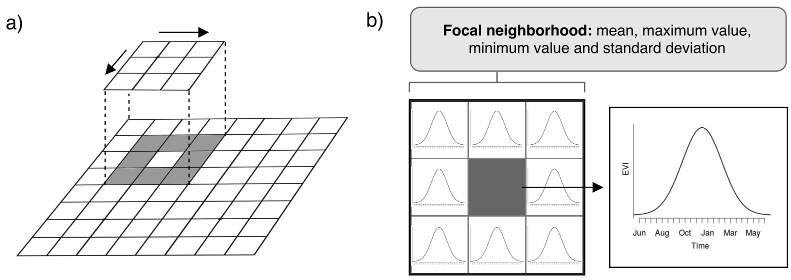
2. Features Extraction from Time Series

Within the context of time series, temporal features are measures extracted from the original a temporal series of data during a timespan. Examples of temporal features are mean and standard deviation for time series with EVI values during a period of 1 year.

Three approaches to extract temporal features to represent land cover classes were explored in this study: 1) statistics measures in annual timespan; 2) statistics measures from 4 sub-intervals in an annual timespan; and 3) statistics measures in annual timespan aggregated for a set of pixels in a given neighborhood configuration. Figure 1 provides a schematic view of the three approaches.



(a) Flowchart with two approaches of features extraction.



(b) Focal feature extraction

Figure 1: Three approaches to features extraction.

In the first approach, for a set of pixels 12 measures (or features) were computed for an annual interval: maximum value, minimum value, standard deviation, mean, amplitude, sum, median, variance, skewness, kurtosis, area under curve and perimeter.

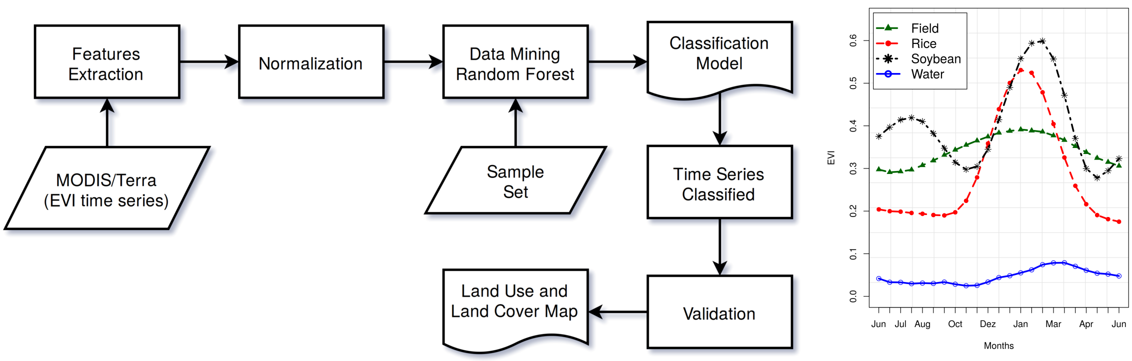
In the second approach, instead of considering the the annual timespan, 4 sub-intervals were considered, and 8 measures were considered potentials for characterisation of a temporal pattern: area under curve, perimeter, maximum value, minimum value, amplitude, mean, standard deviation and sum, for each sub-interval.

In the third approach instead of considering only the one location, we used the Moore neighborhood, 8-connected pixels, to extract an aggregation in space of the features. We extracted 6 features, with four neighborhood aggregation possibilities: mean, maximum value, minimum value and standard deviation. We refer to this as "focal features". For example, the focal feature "maximum_value_mean", computes the maximum value of the metric "mean" calculated in a sample location and in its neighborhood.

3. Case Study

To evaluate the impact of using different approaches to select features from time series we conducted an experiment a classification experiment.

Figure 2(a) summarises the methodology used in the experiment.



(a) Methodology for temporal patterns classification.

(b) Temporal patterns

Figure 2: Classification experiment.

For the acquisition of the EVI time series from MODIS sensor for all years of the study case we used the *Web Time Series Service* (WTSS) (QUEIROZ et al., 2015), which provides the MODIS time series of a given location.

The data was preprocessed with a filtering with Savitzky-Golay filter for cleaning noise, atmospheric error reduction and highlight EVI values of the phonological curve (SAVITZKY; GOLAY, 1964). To remove missing data and gaps in the series we performed an linear interpolation.

In the following step, we extracted the feature according to the two first approaches, considering annual time series, for sub-interval, following by Min-Max Normalisation of the values and finally, focal features extraction. This order was performed to preserve normalised values of the annual time series when focal features extraction was performed over adjacent time series.

Next, a sample set were randomly selected in space and its time series were used as input to the data mining algorithm to build classification model based in a decision tree. In data mining stage we used the Random Forest algorithm (BREIMAN, 2001), implemented in the Weka tool.

After this data mining stage, the features extracted from time series set were classified using the classification model, validated and the use land and cover land maps were created.

4. Experiment: Itaqui municipality - RS

As the study area we have chosen the Itaqui municipality, in the west border of the Rio Grande do Sul state. It has an extension of approximately 3,400 km^2, geographical coordinates latitude 29° 07' 31" South and longitude 56° 33' 11" West, and population of 38,150 inhabitant in 2010, according to IBGE (IBGE, 2016). Itaqui belongs to the Pampa biome, characterised by grassland vegetation composed for graminea, creeping plants and some trees. The municipality have their land used for agricultural production of temporary crops, and is one of main producers of rice n Rio Grande do Sul state.

For this experiment, we use a set of approximately eight hundred thousand time series with EVI data, from MODIS sensor, derived of the MOD13Q1 product for the period from 2000 to 2014 (62620 time series for each year), obtained by WTSS service, filtered and divided in four sub-intervals, in which each entire time series cycle was composed by 23 EVI values.

4.1. Samples training

We random selected 40 sampling points of the study area for each year, in a total of 14 years divided in annual cycles of jun/2000-jun/2001 until jun/2013-jun/2014 generating 560 samples. From these, 520 samples were classified by means visual interpretation of the images from Landsat 5 and 7 satellite, available on Google Earth Engine platform (Google Earth Engine Team, 2015) in 5 classes: field, soybean, rice, water and others. To define the temporal patterns of the rice and soybean crops, we used the temporal patterns adapted from field work performed in Kuplich, Moreira e Fontana (2013), Mengue e Fontana (2015), Figure 2(b). The "water" class was defined by means observation of the samples and the “others" class was defined has having cycles of the EVI values distinct of the others temporal patterns.

4.2. Data mining

In this stage, the set of labelled samples was used as input to the random forest algorithm. The classification accuracy was analysed and evaluated using cross validation method. The model with greater accuracy of classification created was applied for all 14 years of the time series, resulting in the land use and land cover maps showed in Figure 3.

4.3. Evaluation and discussion

To evaluation the resulting classified maps we used coefficient of concordance such as global accuracy, Kappa and Tau.

Table 1 presents the results of the evaluation of the coefficients of concordance for each features set and their combinations. In general, the classification accuracy was considered satisfactory for all set of features, but the classification which combined basic features, for seasons and focal neighborhood has superior performance than others, with 96,78% of global accuracy, 94,18% and 95,98%, for Kappa and Tau indexes, respectively. Kappa value was lower than global accuracy, for considering all cell of the confusion matrix, in contrary of Tau witch showed a value more close of the global accuracy. The classification which combined the three features set was used for creation of the classification model and thematic maps. To determine the thematic classification accuracy was used the confusion matrix, or error, and user's and producer's accuracy, showed in Table 2.

Soybean and rice classes, achieved values of producer's accuracy of 91.84% and 94,87%, respectively. These high value of accuracy indicate a high number of elements which belong the their own classes and low percentage of elements classified to other classes. The user's accuracy obtained percentage of 96,52% e 96,67%, to soybean and rice, and error of commission (inclusion) of 3,48% and 3,33%, that indicates few elements of other classes erroneously

ISBN: 978-85-17-00088-1

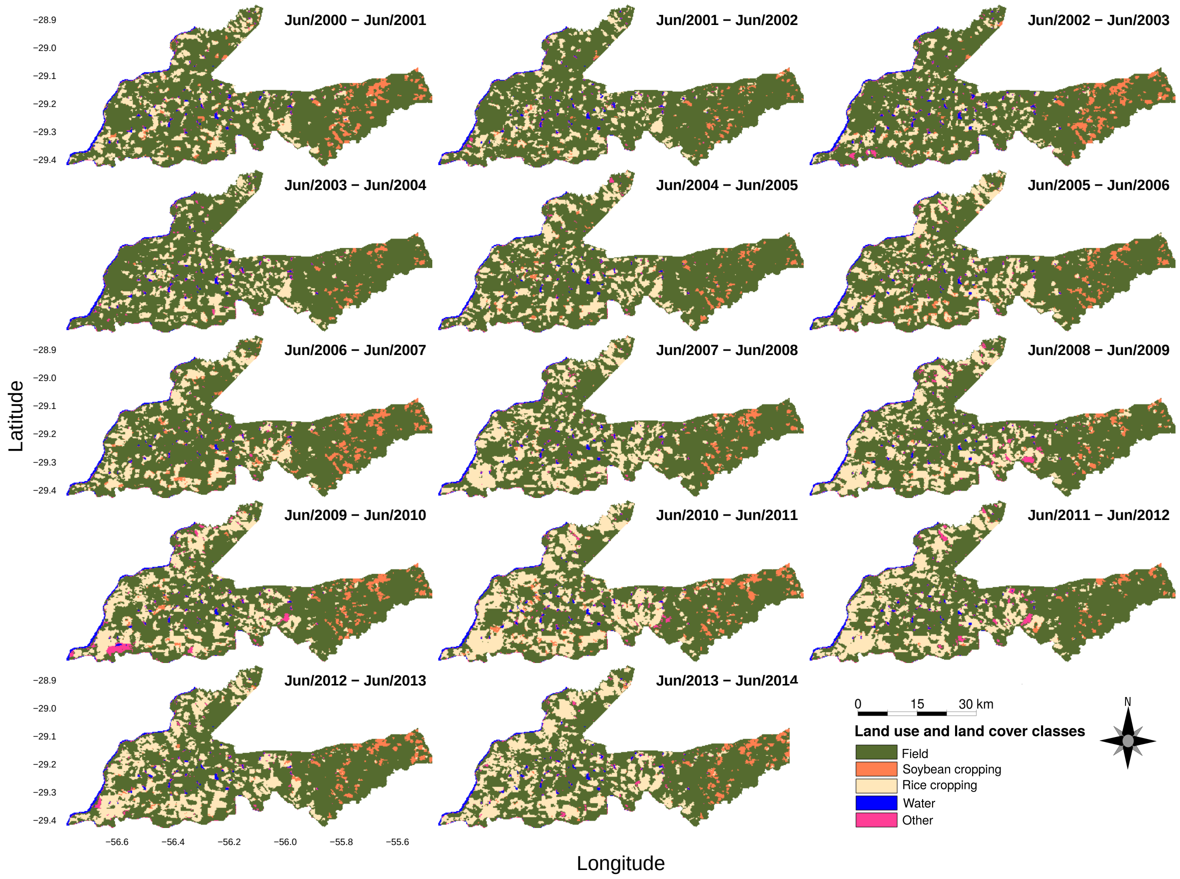


Figure 3: Annual land use and land cover maps to Itaqui municipality, produced by random forest algorithm with EVI/MODIS time series.

mapped as Soybean or Rice.

Field class, has a high value of producer's accuracy, with 98,48% of success, and only 1,52% of error of omission (exclusion), almost 100% of the elements belonging to their own class. The values of user's accuracy also high with 97,58% of the elements correctly mapped as Field and error of commission of 2,42%, in other words, few elements were erroneously mapped as Field.

In general, the classification was good, with global accuracy greater than 95%, which shows the potential of this approach. In this experiment our main goal was distinguish Rice and Soybean classes of the other types of land use and land cover. With this classification we can also observe the progress of rice cultivation areas over time. That shows an increased human activity in the municipality with the expansion of land for this cultivation. Figure 4 shows the percentage of the total area for each class over time.

In order to make a brief comparison of the data obtained of the mapping with data of the IBGE - Agricultural Production by City (PAM), we can verify by graphics in Figure 5, that exists a difference between rice and soybean harvested area from PAM and mapped areas by methodological process, in particular the soybean areas, where is visible a big discrepancy. According to Gass et al. (2015) the differences between PAM data and mapped result, can be by the fact the farmers using rotation in the areas in successive years, which facilitates the soil preparation and management against pests. In this scenario, some rice farmers usually to occupy the double or triple of the area in their properties with one cropping in relation the area which is cultivated for year. Where the areas that are not used for agricultural purposes remain in fallow, may be used as pasture.

Table 1: Comparison of overall classification accuracies, kappa and tau for all classifications performed with random forest algorithm

Features dataset	Overall accuracy(%)	Kappa(%)	Tau(%)
Basics	95.46	91.78	94.33
Sub-interval	96.40	93.47	95.50
Focal neighborhood	92.43	86.30	90.54
Basics + Sub-interval	96.40	93.51	95.50
Basics + Focal neighborhood	96.59	93.85	95.74
Sub-interval + Focal neighborhood	96.59	93.84	95.74
Basics + Sub-interval + Focal neighborhood	96.78	94.18	95.98

Table 2: Confusion matrix of the time series with random forest classifier

		Reference						
	Land use and land cover class	field	soybean	rice	water	others	Total X_{+i}	User(%)
Predicted	field	323	3	4	0	1	331	97.58
Predicted	soybean cropping	2	45	2	0	0	49	91.84
Predicted	rice cropping	3	1	111	0	0	115	96.52
Predicted	water	0	0	0	29	1	30	96.67
Predicted	others	0	0	0	0	4	4	100.00
	Total X_{i+}	328	49	117	29	6	529	–
	Producer(%)	98.48	91.87	94.87	100.00	66.67	–	–

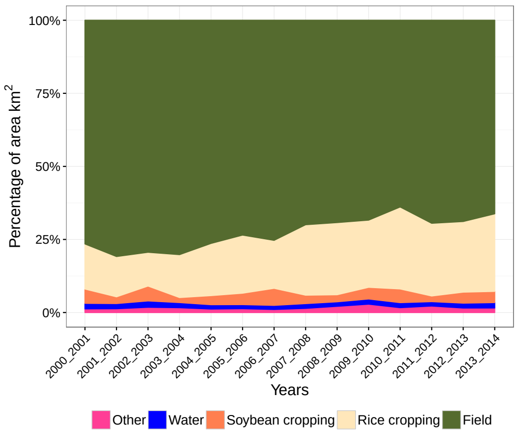
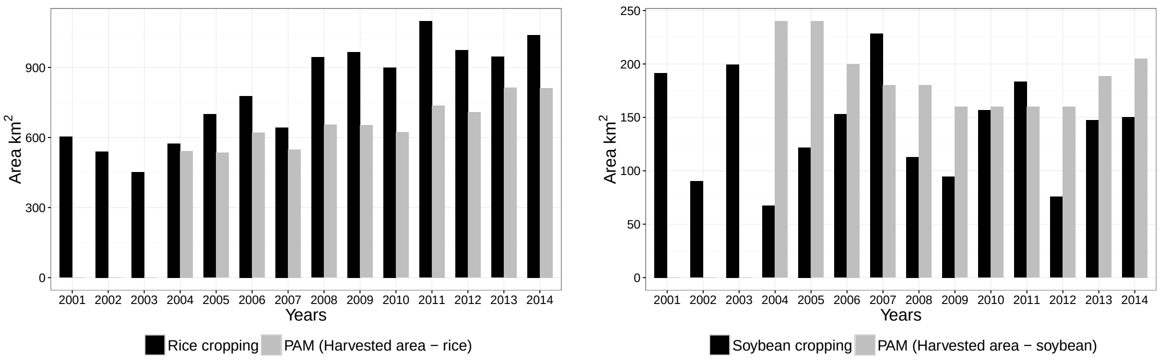


Figure 4: Percentage of area (km^2) for each classe of the land use and land cover from 2000 until 2014 according to classified classes.

5. Final Considerations

This study applied a methodology to feature extraction using sub-interval of entire time series and focal features. The proposal presents in this paper showed that combination of different measures, derived from adjacent time series with value EVI for classification of land use and land cover is important to identification of change in the study area. With time series analyses is possible distinguish gradual and abrupt changes caused by external factors, such as deforestation, new land to agricultural.



(a) Rice temporary cropping.

(b) Soybean temporary cropping.

Figure 5: IBGE - Agricultural Production by City (PAM) (2004 – 2014)

Acknowledgements

The authors thank CAPES and FAPESP e-science program (grants 2014-08398-6 and 2016-03397-7) by financial support.

References

ARVOR, D. et al. Classification of modis evi time series for crop mapping in the state of mato grosso, brazil. *International Journal of Remote Sensing*, Taylor & Francis, v. 32, n. 22, p. 7847–7871, 2011.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

GASS, S. L. B. et al. Estruturação do banco de dados e caracterização básica do município de Itaqui, RS, Brasil, para fins de seu Zoneamento Ecológico-Econômico. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17.(SBSR). *Anais*... São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2015. p. 4073–4081.

Google Earth Engine Team. *Google Earth Engine: A planetary-scale geo-spatial analysis platform*. 12 2015. `https://earthengine.google.com`.

HüTTICH, C. et al. On the suitability of modis time series metrics to map vegetation types in dry savanna ecosystems: A case study in the kalahari of ne namibia. *Remote Sensing*, v. 1, n. 4, p. 620, 2009.

IBGE. *The Brazilian Institute of Geography and Statistics*. 2016. Disponível em: <http://www.ibge.gov.br>.

KUPLICH, T. M.; MOREIRA, A.; FONTANA, D. C. Série temporal de índice de vegetação sobre diferentes tipologias vegetais no Rio Grande do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 17, p. 1116–1123, 2013.

MENGUE, V. P.; FONTANA, D. C. Assessment of spectral-temporal dynamics for mapping the main summer crops in the rio grande do sul state. *Bragantia*, v. 74, n. 3, p. 331–340, 2015.

NITZE, I.; BARRETT, B.; CAWKWELL, F. Temporal optimisation of image acquisition for land cover classification with random forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation*, v. 34, p. 136–146, 2015.

QUEIROZ, G. R. et al. Wtss: um serviço web para extração de séries temporais de imagens de sensoriamento remoto. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17. (SBSR), João Pessoa. *Anais*... São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2015. p. 7553–7560.

RALANAMAHATANA, C. A. et al. Mining time series data. In: ____. *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2005. p. 1069–1103.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, ACS Publications, v. 36, n. 8, p. 1627–1639, 1964.

TOMLIN, C. D. *Geographic Information Systems and Cartographic Modeling*. New Jersey: Prentice-Hall Inc., 1990. 572 p.

ZHANG, X. et al. Land cover classification of the north china plain using modis_evi time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 63, n. 4, p. 476–484, 2008.