THE USE OF LAND COVER CHANGE LIKELIHOOD FOR IMPROVING LAND COVER CLASSIFICATION

*Mariane S. Reis, Sidnei J. S. Sant'Anna, Luciano V. Dutra and Maria Isabel S. Escada**

Brazilian National Institute for Space Research
São José dos Campos, SP – Brazil

Eliana Pantaleão

Federal University of Uberlândia
Patos de Minas, MG – Brazil

ABSTRACT

The likelihood of transitions between pairs of land cover and land use classes in a given time interval and environmental context can be used to impose classification restrictions on an image or to evaluate results. This study presents a methodology for using the likelihood of transitions between classes to improve land cover classification, given a base map (a supposedly accurate map for the same area in another date) and a set of previously classified images. These improved land cover classified images were named conditioned classified images. We aimed to classify one Synthetic Aperture Radar image and an optical one, both from June 2010, using two land cover legends in different level of detail for a region in the Brazilian Amazon. We used both a classified image from 2008 (also in two legends levels) and the data from the Programme for the Estimation of Deforestation in Brazilian Amazon (PRODES) from 2008 as base maps, and presented the likelihood of transitions between the considered classes. The proposed methodology resulted in conditioned classified images with higher Overall Accuracy than the one that does not consider the base maps and the likelihood of transitions. The conditioned classified images presented unlabeled areas due to classification errors in the input data. It is important to highlight that these areas are probably misclassified in maps obtained without using likelihood transition and base maps, since they are impossible to occur in the field.

***Index Terms*—** Image classification, likelihood of transitions between pair of classes, conditioned classifications.

*Funded by CNPq grants #312753/2015-2, #401528/2012-0 and #309135/2015-0. Special thanks to ICMBio(MMA) for SISBIO authorization #38157-2, LBA program and the Monitoramento Ambiental por Satélite no Bioma Amazônia project, process #1022114003005-MSA-BNDES.

1. INTRODUCTION

Land cover (biophysical state of the earth surface) and land use (the purpose for which the land is used) classes have intrinsic relationships that can be helpful for classification and analysis of multi temporal remote sensing data [1]. For instance, land cover classes resulting from ecological succession should occur in a logical order in time and in the space. Additionally, it would be unfeasible finding a well developed forested class in a region that was clear cut in the previous year.

The knowledge about the likelihood of the transitions between classes can be used to impose classification restrictions on an image, to either evaluate results or improve image classification. The Programme for the Estimation of Deforestation in Brazilian Amazon (PRODES), for example, only registers a deforested area in a given year if this area has been classified as forest in all the previous years, because of the adopted definition of deforestation, which supports the whole methodology used in the cited project. In the study carried out by [2], two individually classified images, from the same region in Brazilian Amazon, in two different dates, were compared in order to generate change maps. Using the definition of unlikely and impossible transitions between pairs of classes in that region and date, it was possible to identify errors in up to 30 % of the classifications in a spatially explicit way, with no need of reference data (ground truth). Similarly, [3] created a set of rules regarding the probability of changes from one class to another. The authors used it to separate 'real changes' from possible classification errors. A similar approach was used by [4]. The authors used the characterization of inconsistent transitions in sets of multi temporal classifications to identify areas with classification errors.

In works such as [1, 2, 5, 6], the authors presented the likelihood of transitions between two legends, in a given area and time interval, in a matrix form, herein denominated 'likelihood matrix'. In the present study, we propose a methodology for using likelihood matrices for improving land cover classification. In our case, we propose to classify an image from 2010 with the support of a land cover map from 2008 for the same area and a likelihood matrix of transitions, built to keep the relationships between the land cover legends adopted. We tested land cover classification for two land cover legends, one optical and one Synthetic Aperture Radar (SAR) image and using two different 2008 land cover maps.

2. MATERIALS AND METHODS

This study is focused on obtaining a land cover classification for the year of 2010, in an area of approximately 412 km^2 located in Belterra, Pará state. This area was previously studied in [2] and is illustrated in Figure 1, along geographical references. This is a region of humid tropical climate that presents dense forest vegetation, in which woody lianas, palms and epiphytes can be found. Due to the human occupation process, the study area presents patches of secondary vegetation, pasture and agriculture within the forest matrix.

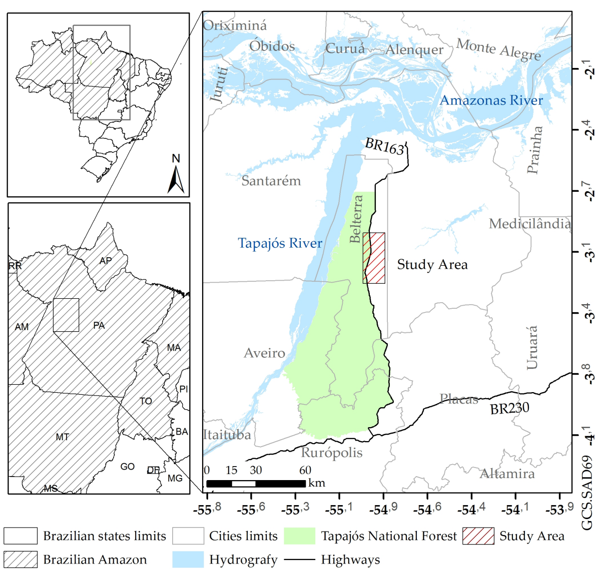

Fig. 1. Study area, with geographical references [2].

Two legends were previously defined by [2] are considered in this work. The first legend (L1) encompasses the following ten land cover classes:Bare Agricultural Soil (BS), Idle Agricultural Area (IA), Cultivated Area (CA), Clean Pasture (CP), Overgrown Pasture (PA), Initial Secondary Vegetation (SV1), Intermediate Secondary Vegetation (SV2), Advanced Secondary Vegetation (SV3), Modified Forest (MF) and Mature Forest (MA). The second legend (L2) is obtained by grouping similar classes from L1, as illustrated in Figure 2. The description of classes and details of generalization can be found in [2].

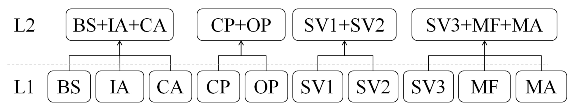

Fig. 2. Legend levels. Modified from [2].

Generally, a supervised classification process is done considering one image, one set of labeled training samples and one classifier, which results in a single classified image. However, it is possible to vary any of the input data in order to generate a set of classified images. One way to do this is to set the image and classifier and to vary the set of labeled samples used to train the classifier. In this case, a final land cover classification can be obtained from this set, in which each pixel in the image is labeled as class more frequently classified in the performed classifications, i.e the mode of the classified images. The proposal of this work is to add another land cover map (namely the base map), from the same area but from another date, to limit the number of land cover classes that are possible to be obtained in the final classification, following the rules described in likelihood matrices. In this case, the final land cover classification does not receive the label mos frequently assigned to the classification set, but the most frequent label that also results in possible transitions when compared to the base map, resulting in a conditioned classified image. This process is illustrated in Figure 3. The former approach is illustrated in black, and the inclusions proposed in this work are illustrated in red.

In this work, two sets of land cover classification of each legend were used to obtain the 2010 final land cover classification. The first one is the same generated by [2], in which the authors classified a LANDSAT5/Thematic Mapper (TM) image from June 29 2010, varying the training samples from two legends and using the pixel based classifier Maximum Likelihood (ML), in order to obtain a set of 100 classified images for each legend. The second one was obtained by the classification of an image from the Phase Array L Band Synthetic Aperture Radar sensor (PALSAR) on board of Advanced Land Observing System (ALOS), from June 21 2010, acquired in Fine Beam Dual (FBD) mode (HH and HV polarizations in L-band), at 1.1 level of processing. This image was orthorectified using Shuttle Radar Topography Mission 4 (SRTM 4) data and the Rational Function Model (RFM) present in PCI 13.0 software, re-sampled to square pixels of 15 meters, filtered using Stochastic Distances Nonlocal Means (SDNLM) filter [7] and used in amplitude format. The ALOS/PALSAR image was classified 100 times for each adopted legend, considering the variation of train-

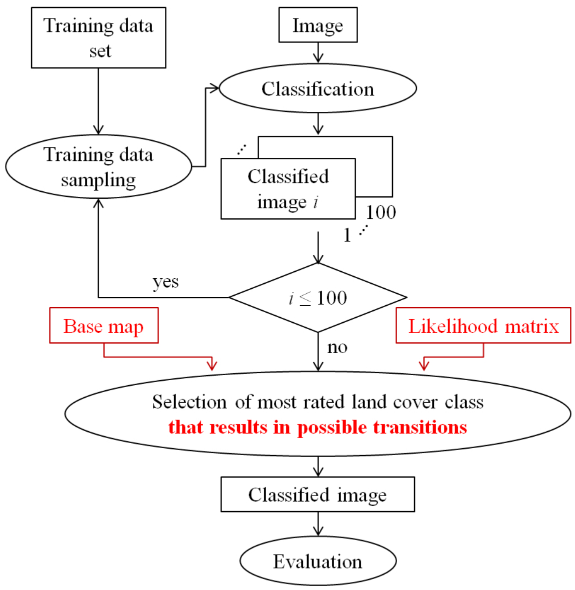

Fig. 3. Supervised land cover classification system, considering variation in the training dataset. Steps illustrated in red refers to the proposed addition of base maps and likelihood matrices to improve land cover classification.

ing samples. Each time, 1200 labeled samples of each class were used to train a pixel based Maximum Likelihood classifier.

Two types of base maps are used. One is the PRODES data from 2008, grouped into three classes: 2008 Deforestation (areas mapped as deforested in 2008), Forest (areas of primary forest) and Other Classes (areas previously deforested, under clouds or otherwise under the PRODES mask). The other type of base map is the mode of the set of 100 classifications of a LANDSAT5/TM image from June 23 2008, as obtained by [2] for legends L1 and L2 and using ML classifier. These base maps, as well as the sets of TM classified images from 2010, were re-sampled to square pixels of 15 meters, to allow comparison.

The likelihood matrices of these data are illustrated in Table 1. Notice that the part describing the likelihood of transitions between L1/L1 legends or L2/L2 legends are from [2], while the ones regarding PRODES legend and either L1 or L2 legends were derived for this work. It is important to highlight that there is no sense in performing a L1/L2 comparison, since its result would be the same as comparing two L2 based classifications.

Four classified images were obtained by calculating the mode of each classifications set (optical and SAR based for L1 and optical and SAR based for L2). These classified images will be denominated as *data_L_mode*, in which *data* refers to the image that was used as input in the classification process (TM or PALSAR, for simplification) and *L* refers to the legend used (either L1 or L2). Using the proposed methodology, eight conditioned classified images were obtained (two legends × two input images × two base maps). These are named *data_L_base* in this work, in which *base* refers to the base map used (2008 PRODES or 2008 TM). These twelve classified images were evaluated using a Monte Carlo approach, in which 100 pixels for each class from a independent test set of samples were randomly selected and used to calculate a confusion matrix and the Global Accuracy index. This process was repeated 10000 times varying the test samples, resulting in 10000 values of the index. The mean and standard deviation of Global Accuracy values were used for comparison.

Table 1. Likelihood of transitions between .

L1 legend transitions

			2010									
2008			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Land Cover Classification (LCC)	**1.Bare Ag. Soil (BS)**	E	E	E	E	E	U	I	I	I	I
		2.Idle Ag. Area (IA)	E	E	E	E	E	E	I	I	I	I
		3.Cultivated Area (CA)	E	E	E	E	E	U	I	I	I	I
		4.Clean Pasture (CP)	E	E	E	E	E	U	I	I	I	I
		5.Overgrown Pasture (OP)	E	E	E	E	E	E	I	I	I	I
		6.Initial S.V. (SV1)	E	E	E	E	E	E	E	I	I	I
		7.Intermediate S.V. (SV2)	E	E	E	E	E	U	E	U	I	I
		8.Advanced S.V. (SV3)	E	E	E	E	E	U	I	E	I	I
		9.Modified Forest (MF)	E	E	E	E	E	U	I	I	E	I
		10.Mature Forest (MA)	E	E	E	E	E	U	I	I	U	E
	PRODES	**2008 Deforestation**	E	E	E	E	E	U	I	I	I	I
		Forest	E	E	E	E	E	U	I	I	U	E
		Other classes	E	E	E	E	E	E	E	E	I	I

Note: Ag.=Agricultural and S.V.= Secondary Vegetation.

L2 legend transitions

			2010			
2008			**11**	**12**	**13**	**14**
	LCC	**11.BS+IA+CA**	E	E	E	I
		12.CP+OP	E	E	E	I
		13.SV1+SV2	E	E	E	E
		14.SV3+MF+MA	E	E	U	E
	PRODES	**2008 Deforestation**	E	E	E	I
		Forest	E	E	U	E
		Other Classes	E	E	E	I

Expected Change (E)
Unexpected Change (U)
Impossible Change (I)

3. RESULTS

The mean and standard deviation values of Overall Accuracy of the classifications using L1 and L2 legend are presented in Tables 2 and 3, respectively. As can be observed in both tables, the mean Overall Accuracy for the classified areas in the conditioned classified images is higher than the values presented by the mode of the set of classified images. However, the proposed methodology presented two expected characteristics: firstly, it showed that the quality of the conditioned classified image depends on the quality of the base map. Secondly, if a given pixel was labeled as one class that results in impossible transitions in all classified images in the original set, this pixel is not labeled in the conditioned classified image. The percentage of the classified area is also presented in Tables 2 and 3. Notice that the classified area is higher using the L2 legend than L1 legend, due to fewer impossible transitions and classification errors. It is important to highlight, however, that when the base map is accurate, as in the case of PRODES data, non classified areas are probably misclassified in images classified by mode without the support of the proposed methodology, i.e. classification errors are highlighted in the conditioned classified images.

Table 2. Accuracy of classified images using L1 legend.

Classified image	Overall Accuracy	Classified area (%)
PALSAR_L1_2008 TM	0.456 ± 0.013	77.51
PALSAR_L1_2008 PRODES	0.431 ± 0.013	82.91
TM_L1_2008 TM	0.778 ± 0.012	79.87
TM_L1_2008 PRODES	0.781 ± 0.012	79.02
PALSAR_L1_mode	0.362 ± 0.013	100.00
TM_L1_mode	0.731 ± 0.012	100.00

Table 3. Accuracy of classified images using L2 legend.

Classified image	Overall Accuracy	Classified area (%)
PALSAR_L2_2008 TM	0.646 ± 0.021	98.78
PALSAR_L2_2008 PRODES	0.690 ± 0.022	81.51
TM_L2_2008 TM	0.848 ± 0.017	99.64
TM_L2_2008 PRODES	0.870 ± 0.016	80.39
PALSAR_L2_mode	0.629 ± 0.021	100.00
TM_L2_mode	0.848 ± 0.017	100.00

4. FINAL CONSIDERATIONS

This study presented a methodology for using likelihood matrices for improving land cover classification, given a base map and a set of classified images. This methodology presented results with higher Overall Accuracy and highlighted areas of classifications errors, which led to unlabeled pixels in the final classified images. These results show the necessity of introducing the use of base maps and the likelihood matrix to limit the classes during the first classification step, by the classifier algorithm itself. Future works should also include the information about unexpected transitions to weight the rules of classes assignment.

5. REFERENCES

[1] C. Gómez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *{ISPRS} Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 55 – 72, 2016.

[2] Mariane S. Reis, Luciano V. Dutra, Sidnei J. S. Sant'Anna, and Maria Isabel S. Escada, "Examining multi-legend change detection in Amazon with pixel and region based methods," *Remote Sensing*, vol. 9, no. 1, 2017.

[3] H. Liu and Q. Zhou, "Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison," *International Journal of Remote Sensing*, vol. 25, no. 5, pp. 1037–1050, 2004.

[4] M. Azeredo, A. M. V. Monteiro, M. I. S. E., K. R. Ferreira, L. V., and T. F. Pinheiro, "Land-cover change trajectory mining in forest degradation studies," *Revista Brasileira de Cartografia*, vol. 4, no. 68, pp. 717 – 731, 2016.

[5] M. S. Reis, L. Torres, S. J. S. Sant'Anna, C. C. Freitas, and L. V. Dutra, "Evaluation of SAR-SDNLM filter for change detection classification," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 2042–2045.

[6] D. Anjos, D. Lu, L. Dutra, and S. Sant'Anna, "Change detection techniques using multisensor data," in *Remotely Sensed Data Characterization, Classification, and Accuracies*, vol. 1, pp. 375–395. crc press, London, 2015.

[7] L. Torres, S.J.S. Sant'Anna, C. C. Freitas, and A. C. Frery, "Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means," *Pattern Recognition*, vol. 47, no. 1, SI, pp. 141–157, Jan. 2014.