SUPERVISED CLASSIFICATION OF SATELLITE IMAGES TO ANALYZE MULTI-TEMPORAL LAND USE AND COVERAGE: A CASE STUDY FOR THE TOWN OF MARABÁ, STATE OF PARÁ, BRAZIL

Priscila Siqueira Aranha[1], Flavia Pessoa Monteiro[1] , Paulo André Ignácio Pontes[5], Jorge Antonio Moraes de Souza[2], Nandamudi Lankalapalli Vijaykumar[4] , Maurílio de Abreu Monteiro[3] and Carlos Renato Lisboa Francês[1]

[1]Federal University of Pará (UFPA), Pará, Brazil
{priscilasa, flaviamonteiro, rfrances}@ufpa.br
[2]Federal Rural University of Amazonia (UFRA), Pará, Brazil
jorge.souza@ufra.edu.br
[3]Federal University of South and Southeast of Pará (UNIFESSPA), Pará, Brazil
maurilio.monteiro@unifesspa.edu.br
[4]National Institute for Space Research (INPE), São Paulo, Brazil
vijay.nl@inpe.br
[5]Federal Institute of Education, Science and Technology of Pará (IFPA), Pará, Brazil
paulo.pontes@ifpa.edu.br

ABSTRACT

Amazon has one of the most diversified biome of the planet. Its environmental preservation has an impact in the global scenario. However, besides the environmental features, the complexity of the region involves other different aspects such as social, economic and cultural. In fact, these aspects are intrinsically interrelated, for example, cultural aspects may affect land use/land cover characteristics.

This paper proposes an innovative methodology to investigate changes of critical factors in the environment, based on a case study in the 26 de Março Settlement, in the city of Marabá, in the Brazilian Amazon. The proposed methodology demonstrated, from the obtained results, an improvement of the efficiency of the classification technique to determine different thematic classes as well as a substantial enhancement in the precision of classified images. Another important aspect is the automation in the process with a minimum computational effort in performing the required analyses

KEYWORDS

Image Classification, LULC, LULCC, LANDSAT 5 satellite, 26th March Settlement, TerraClass, Decision Tree, CART, MPL.

Natarajan Meghanathan et al. (Eds) : COSIT, SEC, SIGL, AIAPP - 2015
pp. 09–19, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.50602

1. INTRODUCTION

In Amazon, actions of political nature have major social, economical and ecological significance in the rural settlement. Families that have been settled tell stories of success and failure[1]. Nowadays, in Legal Amazon there are 3554 settlement projects of the Brazilian Land Reform where 752,000 families live in 767 million hectares. 392 of these settlements have damaged forest coverage reaching around 50% to 75% of the lots, thus, affecting permanent preservation areas (APP) as well as areas of legal reservation (ARL). This has a tremendous impact. The Brazilian Government estimates that 75% of emitted CO_2 is due to change in land use of and forests. Fearnside points out that a part of the Amazon forest turning into pasture and agriculture is responsible, in the 1990s, for an annual emission of 1.20 Pg CO_2 [1].

Deforestation rates in Amazon are quite high and it is controversial to determine those that are responsible for such actions. According to IMAZON (Institute of Citizen and Environment of Amazon), 37% of the settlement areas have been deforested [2]. INCRA (Institute for Settlements and Land Reform) admits the problem and the situation is much worse in the State of Para.

Most of the research work shows that practices and processes in land use in Amazon, with emphasis on forest degradation, may affect regional and global climate and ecosystems [3].
So, the study of the dynamics of the land use and coverage in the settlements of the Amazon region turns out to be an important tool to determine and understand the reasons that lead to environmental degradation of such areas.

The work described in this paper proposes a methodology, using supervised classification, to determine land use and coverage (LULC) and changes in land use and coverage (LULCC) that took place during a defined period. The work uses multi-temporal analysis of LANDSAT 5 satellite images (TM Sensor). In order to demonstrate the efficiency of the approach, a case study used the area of 26th March Settlement of Marabá city, in the State of Pará (Brazilian Amazon).

2. RELATED WORK

Due to continental dimensions of the Amazon, remote sensing is indispensible to monitor the environment, detect fire and deforestation actions. In this context, methods and tools to classify satellite images have been in the spotlight to study the land use and coverage as well as changes occurred along these years, so that some public policies can be put into place to efficiently deter deforestation by aiding settled families to work on this issue but without losing their sustenance. Classification techniques can be supervised and non-supervised. In supervised classification, one has to know a priori a domain specialist information, while in the second, this is not required and may be obtained from existing maps or field work [4].

A methodology that takes advantage of supervised and non-supervised classifications is proposed in [5]. On the other hand, [6] and [7] compared Support Vector Machine-SVM, Artificial Neural Networks-ANN and Maximum Likelihood Classification-MLC to infer which could be the best technique to classify images. All the three showed to be equivalent with a 94% of precision to classify land coverage in coastal areas [6]. [7] showed that ANN had a better performance than SVM and MLC but requires a deeper understanding to clearly define in which situations ANN is in fact better as the authors expected that SVM would perform better. In [8], Classification and Regression Trees–CART, ANN and SVM were compared and showed that SVM could be more generalized than others when using small training samples. Precision obtained was between 70% and 80% for samples between 20 and 800 pixels. ANN produced around 67% to 76% of precision while CART 62% to 73%. MLC was applied to study land use and coverage of Yamuna River bed in Delhi [9] and obtained a precision of around 77.1% and 88.6% with Kappa coefficient of 0.71 to 0.86.

MLC and Decision Tree were employed [10] to classify an area in Victoria, Australia and results showed a precision of 74% to 78% with Kappa coefficient of 0.70 and 0.75 respectively. Decision Tree algorithm had a better performance. [11] analyzed territorial dynamics of land coverage in the State of Rondônia, Brazil, also recognized as deforested area. The authors employed supervised classification based on Bhattacharya algorithm associated to image segmentation by Spring software. The study mapped seven different classes of land use and coverage besides identifying changes occurred in the area.

A study on Tourist State Park in Alto Ribeira, São Paulo State (also in Brazil) investigated the performance of different classification methods to map land coverage [12]. Two methods – hybrid classification per pixel (ERDAS 9.1 software) and classification based on objects (eCognition version 5 software) - were employed. Object-based classification performed better with a Kappa coefficient of 0.8687 (hybrid classification obtained a Kappa coefficient of 0.2224). The expected quality was obtained due to domain specialist knowledge during the classification process.

ANN was evaluated by [13] to classify images to analyze land use and coverage. It showed to be efficient, especially, when that input data were not normally distributed. For complex mapping applications, the authors recommend to employ supervised MLP (Multi Layer Perceptron) networks that could classify images with better precision. They also recommend non-supervised SOM (Self-Organizing Mapping) to analyze spectral features between and within the classes.

3. METHODOLOGY

A methodology has been defined to qualify the deforestation and then classify the satellite images by distributing each pixel in thematic classes. Maps from TerraClass project from 2008 to 2015 [14] were used.

The main objective of TerrClass [14] is to qualify deforestation in Legal Amazon based on deforestation mapped and published by PRODES project [INPE, 2008] and by satellite images. This is done by mapping land use and coverage and also by evaluation this dynamics. It consists of four main stages, executed manually and automatically. The first refers to data pre-processing based on acquiring satellite images. Then, GIS (Geographic Information System) is used to compose bands and geometric correction. Lastly, region of interest is extracted. They are manually performed in sequence for all the obtained images. Once the first phase is over, images are stored in a repository so that they go through classification as shown in Figure 1.

The next phases are automatically executed by an algorithm written in Python. It follows the steps from Figure 1 starting from executing the transformation phase to data normalization as well as definition of input variables to be used during the classification. Results from the classification go through pre-processing and presented in tabular and graphic forms with the percentage values of each area based on thematic class. Percentage figures of each area organized its thematic class are also shown along with the classified image. Details of the phases described in the Figure follow.

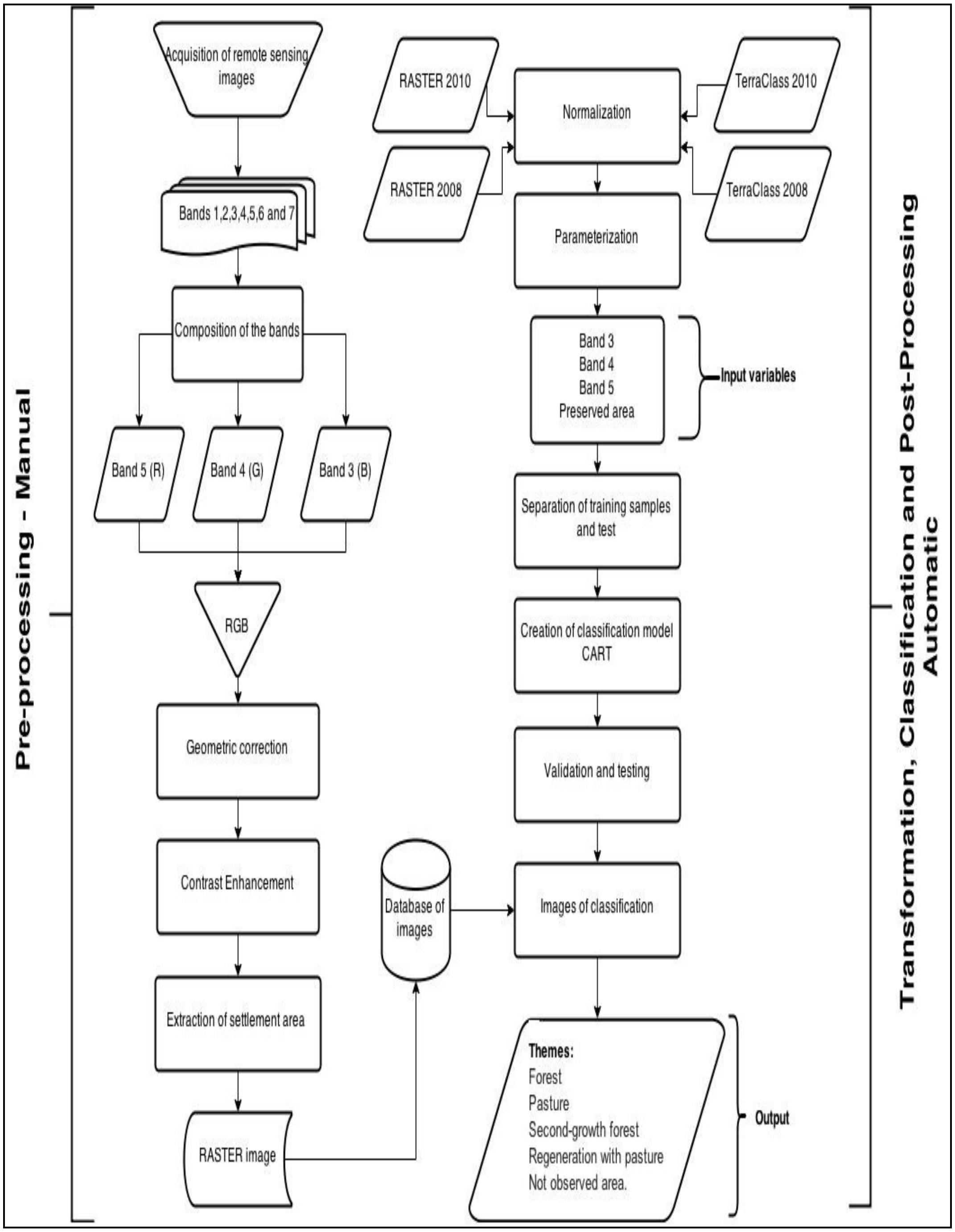


Figure 1. Model of the methodology

3.1. Study Area and Data

The study area corresponds to PA 26th March, created on 19/December/2008. It used to belong to Cabeceiras farm located in Marabá, meso-region of the Southeast of Pará (Figure 2). Its area is 11,919.36 ha with 6 clusters of houses and two headquarters besides 207 family lots [15].

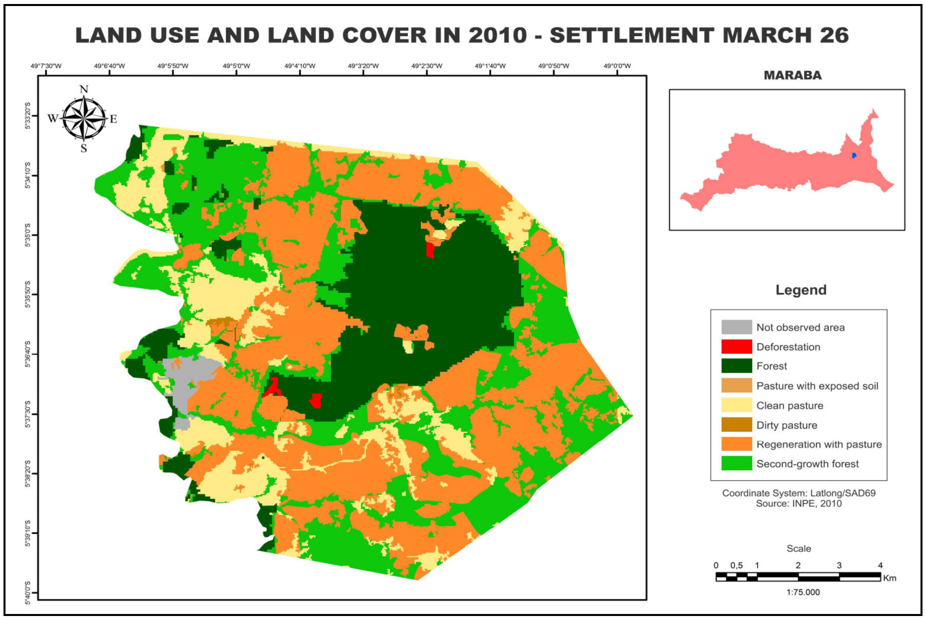


Figure 2. Study Area

In order to perform this study, LANDSAT 5 satellite images, TM sensor, with a geometric resolution of 30 meters for the bands 1, 2, 3, 4, 5 e 7 e 120 meters have been used. After the geo-processing, a database was created adopting a 1:25.000 scale with geographic coordinates SIRGAS 2000. Images were obtained from Brazilian Institute for Space Research (INPE). The collected images were selected according to the percentage of cloud coverage not going beyond 5% threshold. With this criterion, 12 images between 1986 and 2010 were selected also following an interval of 2 to 3 years. In order to minimize seasonal effects of the changes on the land use and coverage, the selected images are from the months between July and September.

3.2. Pre-processing

The original images obtained haven't been pre-processed and present noise due to atmospheric influence and geometric distortions. So, they have to be adjusted in order to avoid problems with quality and validity of extracted information. Once such inconsistencies are sorted out, multi-temporal datasets obtained can now be used in the normalization phase with much more reliability. So, the pre-processing is critical and influences on the quality/validity of the extracted information.

During the pre-processing, ArcGIS, version 10.0 was used. Geometric and radiometric corrections were applied to deal with the positioning and intensity of the pixel values that usually are frequent due to variation of altitude, velocity and atmospheric conditions in the satellite sensors [16]. With an objective of improving the quality of the image, contrast highlighting was applied to distinguish the elements of the image. Empirically, Minimum-Maximum was opted to manipulate the histograms. The values of 45 and 205 were considered as minimum and maximum respectively.

Spatial resolution had to be uniform in 30 m. So, resampling based on closest neighbor was used fixing the error to 0.5 pixel that might generate discontinuity on the resampled image, but without changing the radiometric values of the original image. To minimize the effects due to the combined use of different bands, the same number was used for all the images. In order to compose the bands, 5, 4 and 3 were selected, as they were more representative to distinguish vegetation coverage and to map land use. ArcGIS software was employed to collect training samples and testing samples for the classification phase. TerraClass 2010 was the basis to select the samples. The idea was to qualify deforestation, mapping the land use and coverage in the Legal Amazon from LANDSAT 5 TM satellite images divided into their respective orbit-points.

TerraClass maps all the Legal Amazon and it is expected that the classification may present some inconsistencies when the region of interest is very small as it is the case for Settlement 26th March. Figure 2 shows an example of such an inconsistency. Highlighted area of image A refers to class Non Observed Area but it was classified as Regeneration with Pasture. This could have happened due to high intensity of the spectral return of this class. So, to minimize such distortions, samples were extracted in polygon shapes of those regions whose correspondence between the original and classified images was true.

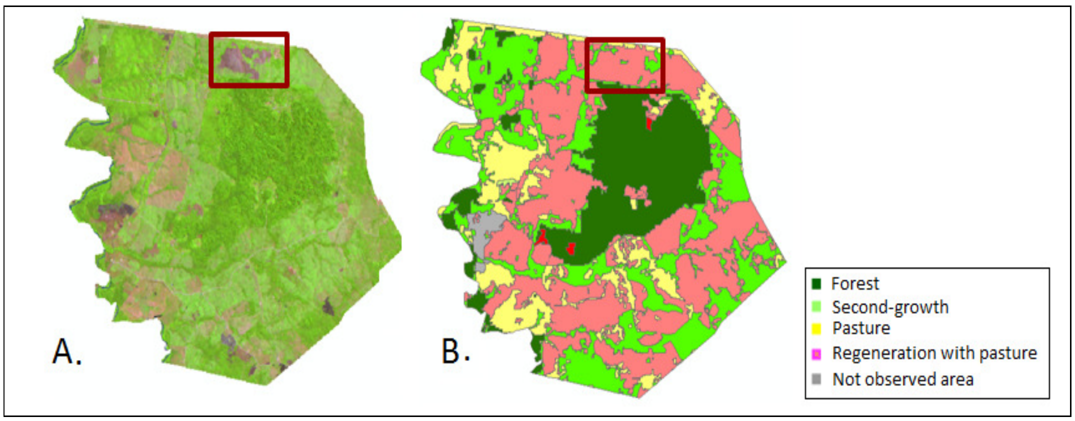


Figura 2. (A) Original Image (B) Classified Image from TerraClass 2010

3.3. Data Transformation

As pointed out in the methodology, data transformation process is divided into two stages: normalization and parameterization. After the pre-processing phase, it is necessary to normalize so that information in the images may be understood during the classification.

For the normalization process, CSV formatted files were created containing information of the positioning of the pixel and gray levels of 3, 4 and 5 bands for those images considered for training and tests. For the parameterization, input values and output thematic classes were defined. Output classes were forest, secondary vegetation, pasture, regeneration with pasture and non-observed area (clouds and forest regional fires) due to their significance for the region of interest. Input values were the gray levels of bands 3, 4 and 5, and the variable “preserved area”, of binary type with 0 or 1. To classify an image of the year x, for example, all the pixels of classified image for the year $x-1$ are verified. If the pixel corresponds to forest, 1 is assigned to the variable, otherwise 0. Its main objective is not to allow that deforested areas are classified as forest.

3.4. Supervised Classification

CART (Classification and Regression Trees) algorithm-based Decision Tree (DT) technique was employed to generate supervised classification model. This is because it associates three important characteristics: speed, precision and simplicity. Besides, it is easy to understand and basically depends on rules based on *if-then*.

Decision Trees represent knowledge and they are an efficient way to build classifiers to predict classes based on attribute values of a set of data. In a Decision Tree, knowledge is tested and represented in each node that conducts a search to its offspring, going down right from the root towards the leaves [17]. CART, proposed in [19] searches for relationships among data besides generating simple decision tree that is readable. To validate the results from employing DT, MLC (Maximum-Likelihood Classification) was compared as it is widely known and applied in several remote sensing problems [18]. Besides, most of the Geo-Processing tools include this algorithm.

3.5. Precision Evaluation

It is important to evaluate the quality of the classification results as it is concerned with extracting intended patterns from the input data. Here, confusion matrix and Kappa coefficient were used to express the precision.. Confusion matrix enables understanding the behavior of the classifier, reacting to effects of wrong prediction. Rows represent expected values and columns real ones. As per Kappa coefficient, it evaluates how good is the model. It is a discrete multi-variate technique that determines the level of the precision varying between 0 and 1. Closer to 1, better is the classification result.

Agreement Level	Values
Poor	0.20 <
Weak	0.21 – 0.40
Moderate	0.41 – 0.60
Good	0.61 – 0.80
Excellent	0.81 - 1

Table 1. Kapp Coefficient Values [19]

The test samples obtained the following results: general precision of the classification technique of DT was 97.7%, with Kappa coefficient 0.97 and the time to classify was 11 seconds. MLC technique yielded 82.75% with Kappa coefficient of 0.78 and within 4 minutes.

Class (%)	Forest		Secondary Vegetation		Pasture		Regeneration with Pasture		Non Observed Area	
	AD	MLC	AD	MLC	AD	MLC	AD	MLC	AD	MLC
Forest	98.75	83.53	0	0	0.76	28.57	0	0	0	0
Secondary Vegetation	0	1.18	97.81	76.67	0	1.50	0	4.35	3.97	7.94
Pasture	1.10	15.29	0	1.67	97.79	66.92	0.84	4.35	0.37	0
Regeneration with Pasture	0.15	0	0	15	1.39	3.01	98.38	91.30	1.08	0
Non Observed Area	0	0	2.19	6.67	0.06	0	0.78	0	94.58	92.06

Table 2. Confusion Matrix from Classification Techniques

Table 2 shows that MLC was inferior to DT with an emphasis on “Pasture” class that was wrongly classified with only 69.2% of correct classification. The difference is of 30.87% when compared to DT. For the “Non-Preserved Area”, MLC was successful in 92.06% of the cases while DT 94.58%, smallest rate of success.

4. RESULTS

For this study, five classes were defined: forest, pasture, regeneration with pasture, secondary vegetation and non-observed area. After employing DT and MLC, it is clear that there is fragility in classifying small areas from TerraClass. It is important to point out that TerraClass is an excellent to tool to deal with areas of continental dimensions, such as Legal Amazon. DT showed to be much more efficient with respect to MLC obtaining a general precision of 97.7% and Kappa coefficient of 0.97. MLC obtained 82.75% and Kappa coefficient of 0.78. Figure 3 shows the results of the classification for the years 2008 and 2010 based on TerraClass, DT and MLC.

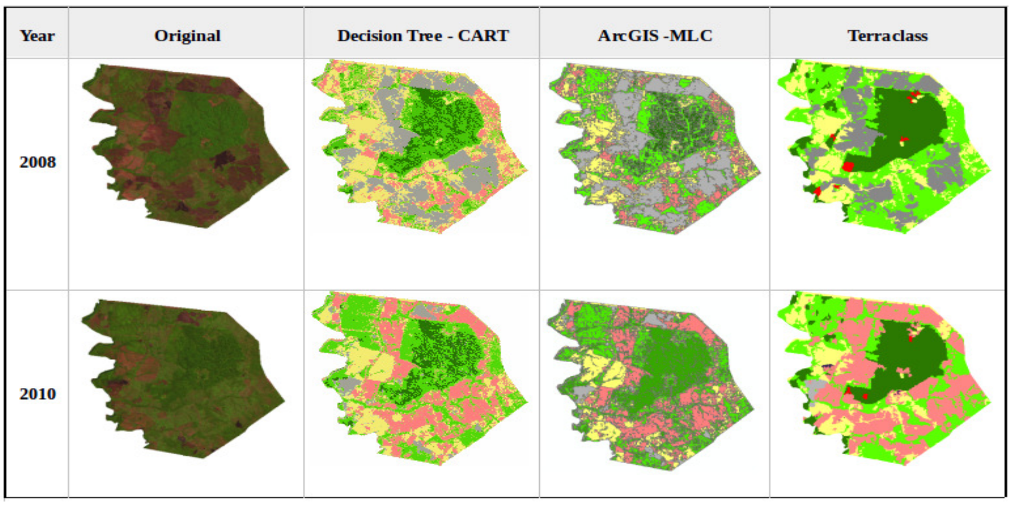


Figure 3. Original Test Image resulted from DT and TerraClass Classification from 2008 and 2009

Based on the classification from DT, one can realize that vegetation coverage of 26th March Settlement suffered a significant degradation, giving place to large areas of pasture. Such dynamics may be explained from the struggle for land ownership between the MST-Movement of Workers without Land and owners of Cabeceiras farm that lead to cutting down a great part of the native forest thus trying to show that the land was productive to avoid expropriation by government in the land reform program. In 2008, when the settlement was created, those that were settled received areas that were highly anthropic. But, deforestation remains, at a slower pace, of course. This occurs due to lack of resources of the settled families that are unable to maintain the pastures clean, leading to, time to time, deforesting new areas. The abandoned pastures give place to regeneration with pasture and at a later stage to secondary vegetation as can be observed in Figure 4.

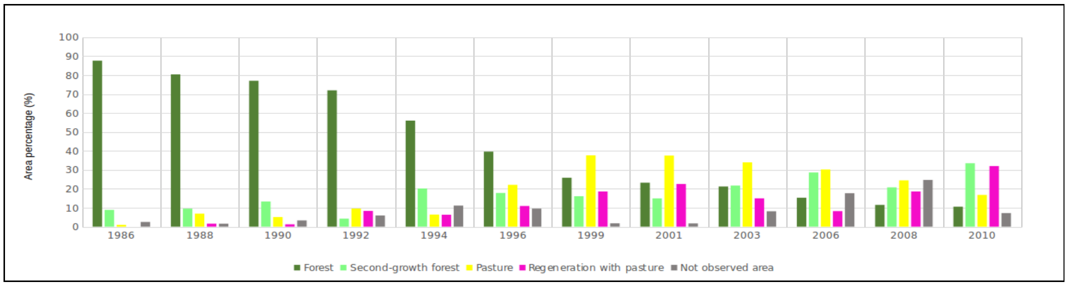


Figure 4. Dynamics of Use and Occupation by Settlement 26th March.

5. CONCLUSION

The study showed that Decision Tree classification technique is more precise than Maximum Likelihood Classification used by ArcGIS 10 tool. The entire process after pre-processing is automatic without the need to train the model every time an image goes through a classification. This approach optimizes time and computational effort. By employing TerraClass, a domain specialist was not required by using thematic classes (resulted from TerraClass) as outputs to construct supervised classification model. Based on the obtained results from DT, it was possible to observe that the landscape of 26th March Settlement is dominated by cultivated pastures and degradation of forest areas increased and directly related to this type of practice. The work also demonstrated with respect to [10] that got a precision of 78% with DT. The work is expected to continue by including more information, such as, other sensors to measure land quality, use of UAV (Unmanned Aerial Vehicle) and socio-economic information of the local population.

ACKNOWLEDGEMENTS

We thank financial support from *National Council for the Improvement of Higher Education - CAPES/Pro-Amazônia and National Council for Scientific and Technological Development - CNPq*.

REFERENCES

[1] Fearside, P. M, (2000)"Greenhouse gas emissions from land use change in Brazil's Amazon region", In: Lal, R.; Kimble, J.M.; Stewart, B.A. Ed. Global climate change and tropical ecosystems. Boca Raton: CRC Press, 2000. p. 231-249. Advances in Soil Science.

[2] Amintas Brandão Jr., Carlos Souza Jr.,Andréia Pinto & Paulo Amaral, (2013) "Situação dos Desmatamentos nos Assentamentos de Reforma Agrária no Estado do Pará.".

[3] Gabriel de Oliveira e Elisabete Caria Moraes, (2003) "Classificação Digital do Uso e Cobertura da Terra na Amazônia com Base em Dados Orbitais ASTER.".

[4] Sílvia Maria Dias Pedro Rebouças, (2011) "Metodologias de Classificação Supervisionada para Análise de Dados de Microarrays." Doutoramento em Estatística e Investigação Operacional (Especialidade de Probabilidades e Estatística)

[5] Hossein Saadat, Jan Adamowski, Robert Bonnell, Forood Sharifi, Mohammad Namdar, Sasan Ale-Ebrahim, (2011) "Land use and land cover classification over a large area in Iran based on single date analyses of satellite imagery." ISPRS Journal of Photogrammetry and Remote Sensing.

[6] Brian W. Szuster, Qi Chen, Michael Borger, (2011) "A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones." Applied Geography.

[7] Prashant K. Srivastava, Dawei Han, Miguel A. Rico-Ramirez, Michaela Bray, Tanvir Islam, (2012) "Selection of classification techniques for land use/land cover change investigation." Advances in Space Research (a COSPAR publication).

[8] Yang Shao, Ross S. Lunetta, (2012) "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points." ISPRS Journal of Photogrammetry and Remote Sensing.

[9] Rucha R., Mangesh Warthe, Sharad Dwivedi, Ritesh Vijay and Tapan Chakrabarti, (2011) "Monitoring changes in lad use land cover of Yamuna riverbed in Delhi: a multi-temporal analysis" International Journal of Remote Sensing.

[10] Liang ZHAI, Jinping SUN, Huiyong SANG, Gang YANG, Yi JIA, (2012) "Large area Land Cover Classification with LANDSAT ETM+ Images based on Decision Tree"

[11] Gabriel de Oliveira e Elisabete Caria Moraes, (2013) "Classificação Digital do Uso e Cobertura da Terra na Amazônia com Base em Dados Orbitais ASTER".

[12] T. Rittl, M. Cooper, R. J. Heck and M. V. R. Ballester, (2013) "Object-Based Method Outperforms Per-Pixel Method for Land Cover Classification in a Protected Area of the Brazilian Atlantic Rainforest Region."

[13] Hui Yuan, Cynthia F. Van Der Wiele, Siamak Khorram, (2009) "An Automated Artificial Neural Network System for Land Use/Land Cover Classification from Landsat TM Imagery." Remote Sens., 1, 243-265; DOI:10.3390/RS1030243.

[14] Lillesand, T.M. and Kiefer, R.W. (2000) "Remote Sensing and Digital Image Interpretation", Wiley, New York, 724 p.

[15] Allison Reynaldo da Costa Castro* Orlando dos Santos Watrin. (2013) "Análise espacial de áreas com restrição legal de uso do solo em projeto de assentamento no sudeste paraense." Geoinformação e Sensoriamento Remoto em Geografia. Geografia Ensino & Pesquisa, vol. 17, n. 2, maio./ago. DOI: 10.5902/22364994/10779

[16] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). "Classification and Regression Trees. Wadsworth".

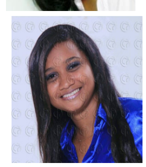
[17] Garcia, S.C, (2000)"O uso de árvores de decisão na descoberta de conhecimento na área da saúde." In: SEMANA ACADÊMICA, 2000. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul.

[18] Landis, J. R. and Koch, G. G. (1977) "The measurement of ob- server agreement for categorical data." Biometrics. 33: 159– 174.

AUTHORS

Studying currently Masters sandwich at the National Institute for Space Research - INPE in Sao Jose Dos Campos.Master's Degree in Electrical Engineering-Applied Computing at the Federal University of Para - UFPA(2013).Has experience in wireless sensor networks and embedded systems.

Master students in Electrical Engineering in the field of Applied Computing at Federal University of Para.Graduated in Computer Engineering from the Community College of the Amazon(IESAM).Has experience in application development for Digital TV, Web systems development using Java and PHP and the study of network simulations using the Network Simulator 3



Graduated in History (1991),with a Masters in Development Planning(1996),both from the Federal University of Para(UFPA).She holds a doctorate in Sustainable Development in the Humid Tropics(2001),also by UFPA, institution where he is a professor since 2002.He has conducted research in which addresses economic, social and spatial implications of the conversion of natural resources into commodities,with special attention to the Amazon Brazil's east

Graduated in Data Processing Technology in the University of the Amazon(1996) and master's degree in Computational Mathematics from the University of Sao Paulo(2000).He is currently a doctoral student at the Federal University of Para and assistant professor in the Amazon Federal rural University.He has experience in computer science,with emphasis on data communication technologies,mainly in the following themes:performance evaluation and wireless sensor networks

PhD in Electronic Engineering(area:Computer) from the Technological Institute of Aeronautics(1999).Participated in the Post - doctoral program at University College Cork(UCC) in Coastal and Marine Resources Centre(CMRC),Cork, Ireland in the area of computer modeling coastal environment. He has experience in computer Science, with emphasis on performance Evaluation, Software Testing with Formal Specification, Time Series Analysis and Computational Modelling of coastal environment

Effective Teacher of the Federal Institute of Education, Science and Technological Para - IFPA. Have Graduation in Technology in Data Processing by the Para State University Center - CESUPA (2003). Specialist in Application Development for internet by the Federal University of Para- UFPA (2004). Master in Electrical Engineering in computing applied by UFPA (2013). PhD in Electrical Engineering in Computing applied by UFPA (2014)

Graduated in Bachelor of Computer Science from the Federal University of paru(1995), master`s degree in Computer Science and Computational Mathematics from the University of Sao Paulo (1998) and a PhD in Computer Science and Computational Mathematics from the University of Sao paulo (2001). Has experience in computer science, with emphasis on performance evaluation, acting on the following topics: Digital TV, Access Technologies, performance Markov models and simulation, correlation techniques (Bayesian networks)and optimization.