DISCRIMINATION OF LOW BIOMASS CLASSES IN A TROPICAL ENVIRONMENT USING PALSAR FDB IMAGERY

PI No 108

Luciano V Dutra[1], Michelle O Parreira, Rogerio G Negri, Wagner B Silva, Luciana O Pereira, Corina C Freitas, Sidnei J S Sant´Anna

Instituto Nacional de Pesquisas Espaciais, Brazil (National Space Research Institute)
[1]dutra@dpi.inpe.br, Tel +5512 3208 6514, Fax +5512 3208 6468
Av.dos Astronautas 1758, São José dos Campos, SP, 12227-010 - Brazil

1. INTRODUCTION

During the last four years, the Microwave Research Program at INPE has been significantly evolved due to the Institution´s participation in the Scientific Research Program established by JAXA, with the important assistance of ASF. Land cover/use classification, [1,2,3,4], deforestation monitoring [5,6], InSAR products evaluation [7,8,9] are the main research areas studied and still in full development by INPE´s researchers and students. One of the main outcomes of INPE´s participation in this research program was the preparation of a new set of professionals, now fully capable of proper use of RS microwave products. For the mapping problems with classes covering the whole range of biomass inside the tropical environment it is possible to discriminate up to 7 classes [1,3,4].

The focus of this work is the assessment of a land use/land cover mapping task composed of a fairly large number of low biomass classes, including different agriculture types and stages, pasture, leftover of plantations, and others, which normally gives a low classification precision when using SAR data. This type of study is important because the mentioned set of classes is the most important one for land use change monitoring.

2. STUDY AREA AND ANCILLARY DATA

The study area is located east of the Tapajós National Forest, a federal conservation unit established by Decree No. 73684 - February 19, 1974, in an area covering approximately 600,000 hectares of land located in four counties: Aveiro, Belterra, Placas, and Rurópolis in the western Brazilian state of Pará [10]. This Mesoregion of the Lower Amazon, located between coordinates 55° 1' 51" W 3° 22' 51" S and 54° 42' 16" W 2° 56' 53" S (Figure 1) is characterized by having a hot and humid climate, with a predominance of dense rain forest. This type of forest is characterized by having a green appearance, with emergent trees up to 40 meters tall and dense scrub.

This region experienced a large mechanized agricultural expansion in the last two decades, characterized by the production of soybeans and other grains such as rice, maize, sorghum and beans. Therefore, the region presents large areas of degraded forest, mainly caused by burning [11]. Thus, there are forest patches in different stages of regeneration as a result of human disturbances near the highway BR-163 (Santarém–Cuiabá). The agricultural calendar begins in December at the start of the rainy season, with one or two planting activities during the year, with harvests, respectively, in March/April and July/August.

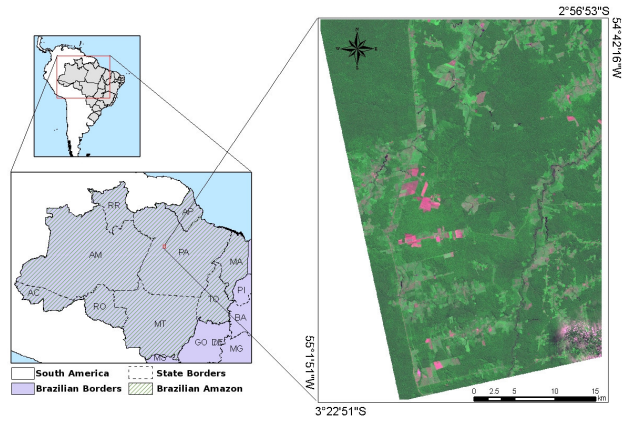

Fig. 1 Location of study area. Color composite R3G4B2

The remote sensing data used in this work were set by a LANDSAT-5 TM scene acquired on 29/06/2010, and a ALOS PALSAR FDB scene acquired on 21/06/2010 with correction level 1.5. The small time gap between these acquisitions made these data particularly suitable for this study. Figure 2 shows details of both scenes.

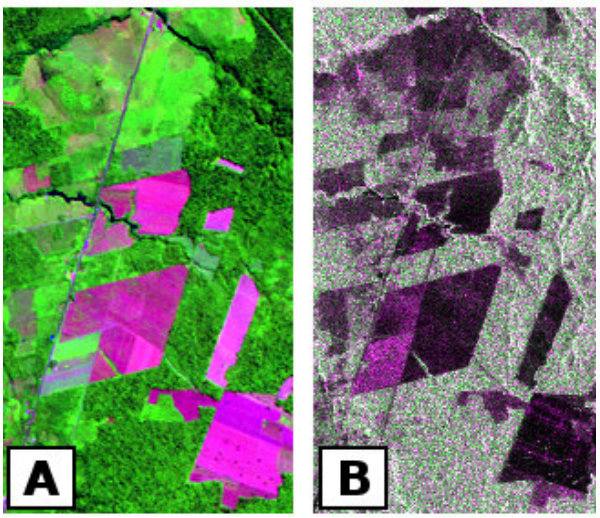

Fig. 2 Clipping of images: (A) TM and (B) SAR.

3. METHODOLOGY

Training and test samples of each class are shown in Figure 3 for a bigger area. Table 1 shows the amount of pixels used for training and validation. These classes were defined from fieldwork carried out on dates consistent with the image acquisitions. It should be noted that the high biomass classes degraded forest (DF), old regeneration (OR) and intermediate regeneration (IR) were grouped into a single class, giving eight classes of use and land cover.

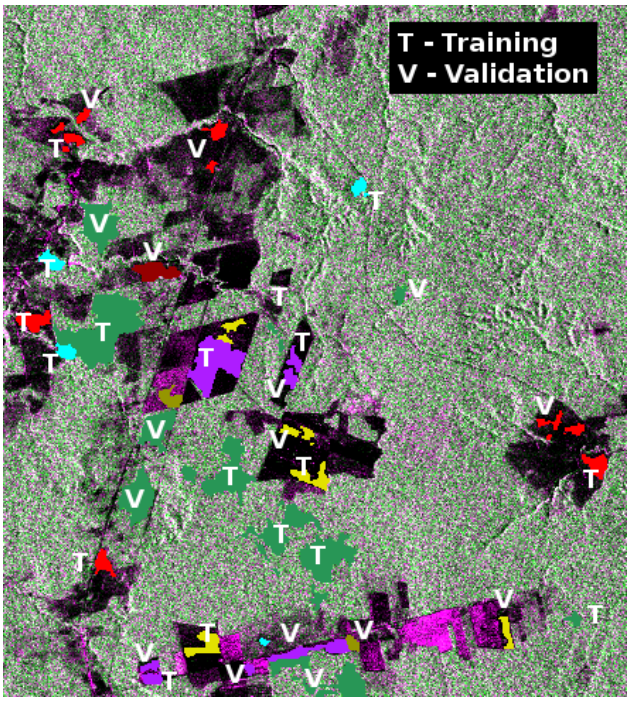

Fig. 3 Class definition with training and test areas

Table 1. Training and test samples collected along the study area

Land use Class	#Pixel Training	#Pixel Test	Color
Primary Forest (PF)	13858	7258	
Degraded Forest (DF), Old and Intermediate Regeneration (OR, IR)	5480	2846	
New Regeneration (NR)	1623	546	
Clean Pasture (CP)	510	292	
Dirty Pasture (DP)	1252	495	
Fallow Agriculture (FA)	1049	450	
Soybean1 (S1)	562	306	
Soybean2 (S2)	142	68	

Per point classification, either contextual or not, and region based classification have been applied to the FDB data. Classifications of contemporaneous TM data set are also presented as comparison. In the sequence a brief explanation of these methods are presented.

a) Per point classification

A contextual Bayesian classification procedure, known as ICM, (section 3.1), the Support Vector Machines (SVM) machine learning method (section 3.2) are employed and compared with the standard Maximum Likelihood (ML) method. The resulted precision is then compared with the ML and ICM classification of TM data. Injection of radar information onto optical data, by means of a fusion approach (section 3.3), is also tested to observe if there would be a precision improvement.

b) Region based classification

Region based classification paradigm has been applied to this RS imagery. Firstly, objects (uniform regions according some criteria) are identified inside the image and then classified. The procedure follows 3 steps:

1. Segmentation
2. Feature extraction
3. Classification of segments.(or regions)

Initially, the FDB scene is processed by an in-house multilevel segmentor (section 3.4) specially developed for radar data, which divides the image into uniform backscatter regions. Secondly, features are extracted from these segments to be used as input data to the following classifiers: the Bhattacharyya distance, the decision tree (DT) and the SVM.

3.1. INTERATED CONDITIONAL MODES (ICM) CONTEXTUAL CLASSIFIER

Any feature vector x representing a pixel is classified to certain class ω_i according to Equation 1.

$$x \in \omega_i \text{ se } g_i(x) > g_j(x) \text{ for each } i \neq j \qquad (1)$$

where $g_i(x)$is the discriminating function. This function, including what is called the context around the position of pixel x is given by the Equation 2, considering normal distribution [12].

$$g_i(x) = -\frac{1}{2}\ln\left|\sum\nolimits_i\right| - \frac{1}{2}(x-m_i)^t \sum\nolimits_i^{-1}(x-m_j)$$
$$+ \ln[p(\omega_i)] + \exp(\beta \#\{t \in \partial_x : class(x_t) = \omega_i\}) \qquad (2)$$

m_i and $\sum_i$ are mean vector and covariance matrix for class ω_i, ∂_x is a certain neighborhood around the x position and t is a coordinate inside ∂_x, # represents 'cardinality of', ':' represents 'given that', β is called an 'attraction factor' and can be estimated by an *apriori* class mapping or user adjusted.

ICM is an iterative estimator of the Pott-Strauss model which describes the distribution of spatial arrangements of classes [13].

3.2. SUPPORT VECTOR MACHINE (SVM) CLASSIFIER

Support Vector Machine (SVM) is a classification method that has received a great deal of attention in recent years due to its excellent ability to generalize and be independent of data distribution. This method consists of finding a separation hyperplane between training samples which have the biggest margins. The separating hyperplane is a function of Equation 3.

$$f(x) = w^T \cdot \phi(x) + b \tag{3}$$

where w^T represents the transpose of the vector orthogonal to the hyperplane of separation, $b/\|w\|$ is the distance from the hyperplane to the origin and $\phi(x)$ is a function adopted, if necessary, to remap the vectors into a new space. The parameters of Equation 3 are obtained from the quadratic optimization problem [14], as shown in Equation 4 and 5.

$$\max_{\lambda}\left(\sum_{i=1}^{l}\lambda_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \left\langle \phi(x_i), \phi(x_j)\right\rangle\right)$$
$$\text{restricted to: } \begin{cases} 0 \le \lambda_i \le C, i = 1,\ldots,l \\ \sum_{i=1}^{l}\lambda_i y_i = 0 \end{cases} \tag{4}$$

where λ_i are Lagrange multipliers, $y_i = \{-1,+1\}$ define the class of x_i , once SVM is a binary classifier, and C acts like a upper bound of λ values. Let's $SV = \{(x_i, y_i) \mid \lambda_i \neq 0; i = 1,\ldots,l\}$, known as support vector set, the hyperplane parameters w and b are computed by Equation 5.

$$w = \sum_{i \in SV}\lambda_i y_i x_i$$
$$b = \frac{1}{\#SV}\left(\sum_{i \in SV} y_i + \sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \left\langle \phi(x_i), \phi(x_j)\right\rangle\right) \tag{5}$$

To apply this method to a multiclass problem, one can use the majority rule or make use of functions which transform the linear SVM classification rule to a non linear case, which can handle the one against all strategy.

3.3. FUSION

One possible concept of image fusion [15] is "the combination of two or more different images or scenes obtained from different sources to form a new image using a particular algorithm".

The fusion technique used in this study is based on the Principal Component Analysis (PCA). One of the channels rotated to the principal axes is substituted by another one from a different source and rotated back to the original axis. The new scene is used for classification purpose [16].

3.4. REGION-BASED CLASSIFICATION

An alternative to pixel based classifiers, usually noisy, is to 1) detected uniform regions or segments inside an image and then 2) using all the information inside the segment or from the collective of these pixels, to provide a classification for the whole segment or region. The results are, in general, better than those obtained by per pixel classifiers.

For this study, for phase 1) it was used a region-based segmentation algorithm, called SegSAR [17], designed specifically for radar data. The segmentation parameters chosen, based in a previous and careful image analysis was 0.5 dB of similarity, 50 pixels of minimum area, equivalent number of looks equal to 3.0 and six compression levels. See [17] for details. After the segmentation the resulted segmented scene is submitted to the classification phase. Three classifiers were tested. The first alternative extracts from segments a vector with nine features and use these extracted features to inform two classifiers – SVM and Decision Tree classifiers. The second alternative is based on what is called statistical distances, which uses second order statics extracted considering all points inside the segment.

3.4.1. Feature Extraction

From all scene segments have been extracted a nine positions vector with segment means (2), standard deviations (2), the variation coefficients (2), the modulus of the complex correlation coefficient between the two bands, and the ratio and difference between the HH and HV means of the segments. After extracting the segment features, it was generated an image band for each one, obtaining a file with nine bands, each one corresponding to a specific feature of the segments.

The feature vectors from regions of interest, representing the classes previously defined, have been then selected to serve as input examples to train a decision tree (DT) and SVM classifiers.

3.4.2. Decision tree (DT)

The parameters list generated for each labeled sample was classified by a supervised non-parametric method called Decision Tree. This method consists of the representation of a decision table or a set of rules regarding the form *IF-THEN*, where each *leaf-node* is the result of a rule [18]. The Decision Trees of this study were built through the J48 algorithm [19]. This algorithm uses a divide-and-conquer strategy where the division is done by the calculation of the Information

Gain of each attribute [19]. Basically the decision trees calculate the entropy of each attribute from training data, selects the attribute that has the highest entropy, and finally, generates a node with the selected attribute in way to maximize the information gain. The algorithm runs recursively expanding the children's node generated [20].

The entropy is calculated using Equation 6 [19], where $H(X)$ is the entropy of a specified attribute X to be valued and P is the probability that X assumes some defined value. The sum is calculated to each value that X can assume.

$$H(X) = -\sum P(X)\log_2 P(X) \tag{6}$$

The information gain (GI) is calculated on the value of entropy of each attribute, conditioned to all other attributes. The GI is measured between the attribute X and any other attribute of the set X', according to Equation 7.

$$GI(X \mid X') = H(X) - H(X' \mid X) \tag{7}$$

3.4.3. Region-based classification stochastic distances

Region-based classifiers can make use of stochastic distances, which measure the separability between probability density functions of the internal data from two different regions on the images. The joint distribution of multi-look intensity pair [21] of FDB data, whose density is shown in the Equation 8, can be used to statistically model this kind of data.

$$f(z_1, z_2) = \frac{N^{(N+1)}(z_1 z_2)^{\frac{N-1}{2}} e^{\left(-N\frac{\frac{z_1}{h_{11}}+\frac{z_2}{h_{22}}}{1-|\rho_c|^2}\right)}}{(h_{11}h_{22})^{\frac{N+1}{2}}\Gamma(N)(1-|\rho_c|^2)|\rho_c|^{(N-1)}} * I_{N-1}\left(\frac{2N|\rho_c|}{1-|\rho_c|^2}\sqrt{\frac{z_1 z_2}{h_{11}h_{22}}}\right) \tag{8}$$

Where z_1 and z_2 are the variables representing the two intensity images; h_{11}, h_{22}, ρ_c and N are distribution parameters which can be estimated by the data, as explained in [21]. The symbol $\Gamma(.)$ denotes the gamma function and I_η denotes the modified Bessel Function of the first kind and order η.

For this work, it was chosen the Bhattacharyya distance, due to its large usage in the Remote Sensing tasks, mainly when using the Gaussian model. Considering $f(x)$ and $g(x)$ two densities with the same support S, the general Bhattacharyya Distance formula can be computed by Equation 9 [22], [23].

$$d_B = -\log\int_S \sqrt{f(x)g(x)}dx \tag{9}$$

In the context of the present work, $f(x)$ and $g(x)$ are two bivariate functions similar to the one shown in Equation 8. Due to the high analytical complexity of the multi-look intensity pair density, the calculus of the Bhattacharyya Distance was made numerically. For this purpose, it was used a Interactive Data Language (IDL) routine which computes the double integral of a bivariate function using iterated Gaussian quadrature. As the density support varies from zero to infinity, it was necessary to implement a routine to verify a value from each variable that corresponds to the infinity. It was done by modeling each variable with a Gamma Cumulative Distribution Function. From that point, an iterative search was used to find values that correspond to the 100% of cumulative distribution in the two dimensions.
Using the procedure detailed above, the Bhattacharyya Distance from each segment data distribution to the reference class data distribution was numerically computed. The segments were associated to class whose distance was minimum.

4. RESULTS

Figure 4A presents the result of ICM classification of SAR data in the selected portion of study area (Figure 3B) and Figure 4B the result of SVM (non contextual) of the same portion and data. SVM was run with linear kernel, penalty of 100 and one to one classification strategy. Although ICM results appear to be much less noisy, the overall Kappa coefficient (Table 2) is about the same. SVM result is noisier but ICM result completely misclassified the fallow agriculture (FA) class, which should have been classified as Soybean1.

Classification of TM is presented at Figure 4C for the same area, the same set of classes and almost the same acquisition date. The overall accuracy supremacy of optical data is patent.

Figure 4D presents the result of ICM classification of the fusioned product where the information of HV channel was injected onto the TM PC2 channel and rotated back to the original axes. In this particular case, data fusion did not bring any overall improvement although it was possible to observe that the producer´s accuracy of Fallow Agriculture increased from 61% to 73.7% in when using the fusioned product. Slight improvements for the producers accuracy was also observed for Soybean1 and Soybean2 classes in the fusioned product. Table 1 also shows the important increase on accuracy when using contextual classification.

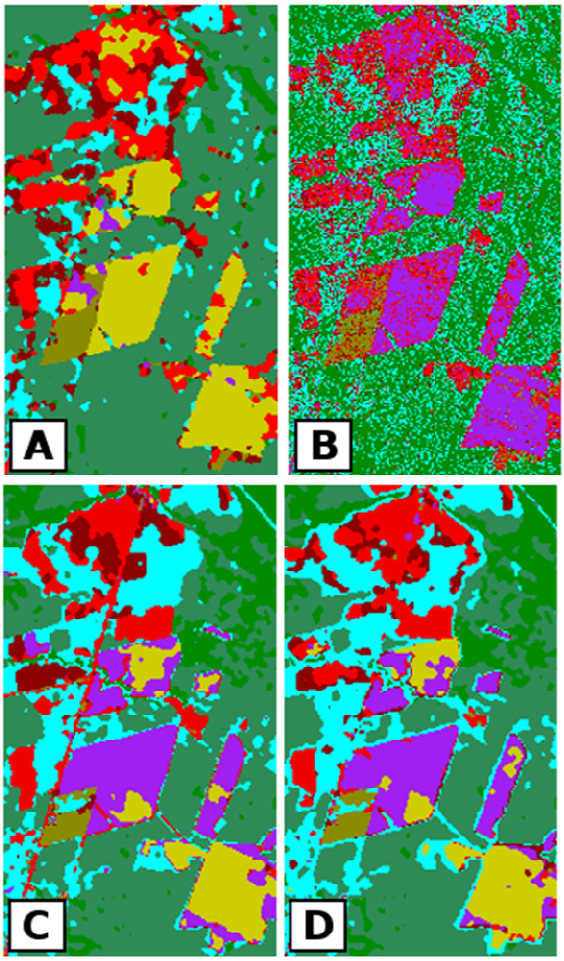

Fig. 4 Clipping of pixel/contextual classified image. (A) ICM (SAR). (B) SVM-Per-point (SAR). (C) ICM (TM). (D) ICM (FUSION/PCA2-HV)

Table 2. Results pixel/contextual classified (%)

		SAR	TM	Fusion	SAR (SVM)
Per-point	Overall Accuracy	31.9	80.0	80.0	48.6
	Kappa Coefficient	13.1	68.1	64.0	16.8
Contextual (ICM)	Overall Accuracy	32.6	92.0	92.3	_
	Kappa Coefficient	15.2	86.4	87.0	_

Trying to improve these results, region based classification was tested. Figure 5A represents the cartoon model (each segment pixel has as value the mean of that segment in that channel) of the radar data segmented by the SegSAR algorithm (HH and HV channels). Figure 5B presents the result of using DT classification of the segmented scene (using the nine elements tall feature vector) and Figure 5C the result of region based classification using BD. Both results are significantly better than per point classification, specially for PF, 'DF+OR+IR' e CP classes. NR class was also satisfactorily classified using BD classifier. Class FA and DP was, both cases, classified as CP, which means that CP has good producer´s accuracy but poor user´s accuracy.

Figure 5D shows the classification from SVM using RBF kernel, 200 of penalty one to one strategy. The result is worse than previous ones maybe because the use of just one example pattern for class S2 is too few for SVM classification when RBF kernel is used. RBF is non linear and needs more examples for adequate learning. Nevertheless the usage of region based classification showed promising prospects for using HH-HV products alone, when no optical data is available. Table 3 shows the accuracy figures.

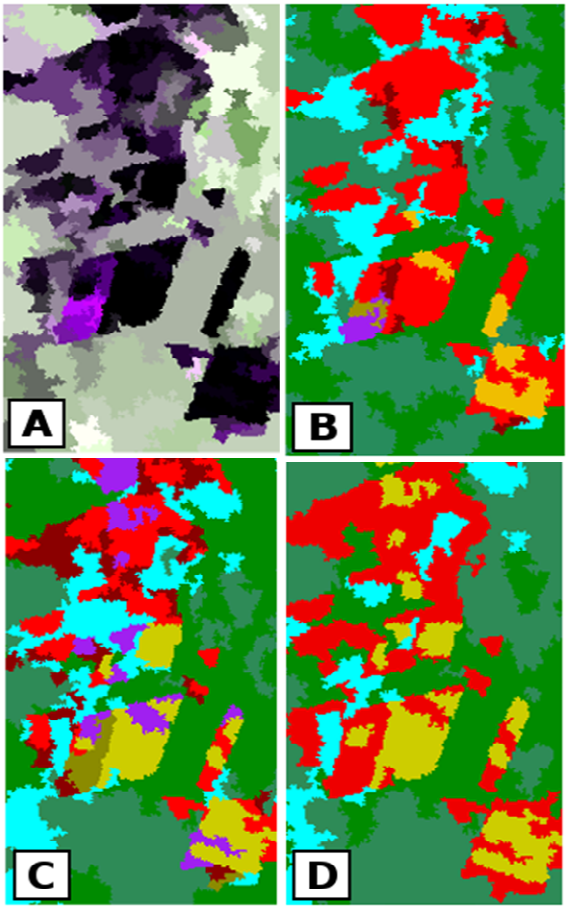

Fig. 5 Clipping of images. (A) Segmented. Classified by: (B) DT (SAR). (C) BD (SAR). (D) SVM-Region (SAR).

Table 3. Results region-based classified (%)

	DT	BD	SVM
Overall Accuracy	65.7	68.4	36.9
Kappa Coefficient	46.0	50.3	13.6

5. CONCLUSIONS.

This paper reinforces the general observation that the use of dual SAR for land use/cover classification did not reach the same general precision than optical data. In the best scenario, using region based classification and dual polarization SAR, one can discriminate up to 4 classes with adequate confidence. But still poor discrimination is common among these low biomass classes. Region based methods showed superior performance. As availability of optical imagery is an issue for tropical area monitoring, the research on the improvement of information extraction methodologies from SAR data is clearly necessary, maybe including texture and shape features and/or higher level context information in the analysis procedures.

Acknowledgments

The authors would like to acknowledge the support of JAXA (108) and ASF for providing the imagery and organizing all the scientific meetings. This research is also partially supported by CNPq/INCT/PQ, FAPESP (2008/58112-0), CAPES (Procad NF2009/486) and National Science Foundation (Grant #BCS0850615)

6. REFERÊNCIAS

[1] R. G. Negri, "Avaliação de dados polarimétricos do sensor ALOS/PALSAR para classificação da cobertura da terra da Amazônia". 2009. 170 p. Dissertation (Masters in Applied Computing) - National Institute for Space Research, São José dos Campos.

[2] L. R. Sartori, N. N. Imai, J. C. Mura, E. M. L. Moraes Novo, T. S. F. Silva, "Investigations on the full polarimetric PALSAR data to discriminate macrophytes species in the Amazon floodplain wetland", *IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2010*, pp.413-416, 25-30 July 2010 doi: <10.1109/IGARSS.2010.5652492>

[3] L. V. Dutra, G. B. Scofield, S. R. Aboud Neta, R. G. Negri, C. C. Freitas, D. L. A. Silva, "Land Cover Classification in Amazon using Alos Palsar Full Polarimetric Data", *IV Simpósio Brasileiro de Sensoriamento Remoto*, p. 7259-7264, April 2009.

[4] R. G. Negri, L. V. Dutra, C. C. Freitas, G. B. Scofield, D. L. A. Silva, S. R. Aboud Neta, "Classificação da cobertura da terra na Amazônia utilizando Imagens Polarimétricas em Banda L e Maquina de Vetores Suporte", *IV Simpósio Brasileiro de Sensoriamento Remoto*, 2009, p. 7863-7869, April 2009.

[5] S. R. Aboud Neta, "Uso das Imagens do ALOS/Palsar para detecçao incremento de desflorestamento na Amazonia". 2009. 264 p. Dissertation (Masters in Remote Sensing) - National Institute for Space Research, São José dos Campos.

[6] S. R. Aboud Neta, C. C. Freitas, L. V. Dutra, "Uso de imagens alos/palsar multipolarizadas para detecção de incremento de desflorestamento na Amazônia. (Use of multipolarized ALOS/PALSAR image for detection of deforestation increment in the Amazon)", *Revista Brasileira de Cartografia* - RBC. V. 62 pp. 417-431, September. 2010.

[7] D. L. A. Silva, "Geração e avaliação de produtos interferométricos dos dados ALOS/PALSAR FBD e PLR para classificação de cobertura da terra na Amazônia". 2009. 159 p. Dissertation (Masters in Remote Sensing) - National Institute for Space Research, São José dos Campos.

[8] D. L. A. Silva, G. B. Scofield, S. R. Aboud Neta, R. G. Negri, L. V. Dutra, C. C. Freitas, "Utilização de imagens de coerência interferométrica em banda L para classificação de cobertura da terra na região de Tapajós-PA", *IV Simpósio Brasileiro de Sensoriamento Remoto*, 2009, p. 7489-7496, April 2009.

[9] G. B. Scofield, L. V. Dutra, D. L. A. Silva, R. G. Negri, S. R. Aboud Neta, C. C. Freitas, "Mapeamento da cobertura da terra na Floresta Nacional de Tapajós - PA utilizando imagem de coerência polarimétrica", *IV Simpósio Brasileiro de Sensoriamento Remoto*, 2009, p. 6227-6232, April 2009.

[10] Secretaria de Estado de Meio Ambiente (SEMA). "Decreto nº 73.684 – Decreto de criação da FLONA do Tapajós - 19/02/1974". Available in: <http://www.sema.pa.gov.br/interna.php?idconteudocoluna=2018&idcoluna=9&titulo_conteudocoluna=73684>. Acesso em: Jan/2011.

[11] A. Venturieri, "Zoneamento ecológico-econômico da área de influência da rodovia BR-163 (Cuiabá-Santarém): diagnóstico do meio sócio-econômico, jurídico e arqueologia". Belém: Embrapa Amazônia Oriental, 229-252 p. 21, 2007.

[12] J. A. Richards and X. Jia, "Remote sensing digital image analysis: an introduction". 4 ed. Berlin: Springer-Verlag, 2006. 439 p.

[13] A. C. Frery, A. H. Correia and C. D. C., Freitas, "Classifying Multifrequency Fully Polarimetric Imagery With Multiple Sources of Statistical Evidence and Contextual Information". *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, n. 10, p. 3098-3109. doi: 10.1109/TGRS.2007.903828, 2007.

[14] S. Theodoridis and K. Koutroumbas, "Pattern Recognition", Fourth Edition. 4th. ed. [S.l.]: Academic Press, 2008.

[15] C. P. Pohl and J. L. Van Genderen, "Multisensor image fusion in remote sensing: concepts, methods and applications". *International Journal of Remote Sensing*. v. 19, n.5, p. 823-854, 1998.

[16] J. Dong, D. Zhuang, Y. Huang and J. Fu, "Advances in Multi-Sensor Data Fusion: Algorithms and Applications". *Sensors*, v. 9, n. 10, p. 7771-7784. doi: 10.3390/s91007771, 2009

[17] M. A. Sousa Júnior, "Segmentação multi-níveis e multi-modelos para imagens de radar e ópticas". 2005. 131 p. PhD Thesis – National Institute for Space Research, São José dos Campos. 2005.

[18] J. R. Quinlan, "C4.5: Programs for Machine Learning". São Francisco: Morgan Kaufmann, 1993.

[19] C. E. Shannon and W. Weaver, "The mathematical theory of communication". Urbana: University of Illinois Press, 1949.

[20] T. M. Mitchell, "Machine Learning". New York: McGraw-Hill, 1997.

[21] J. S. Lee, K. W. Hoppel, S. A. Mango and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery". *IEEE Transactions on Geoscience and Remote Sensing*, v. 32, n. 5, p. 1017-1028. doi: 10.1109/36.312890, 1994.

[22] A. R. Webb, "Statistical Pattern Recognition". 2nd ed. Chicester, West Sussex: John Wiley & Sons Ltd, 2002.

[23] A. G. Wacker and D. A. Landgrebe, "Minimum distance classification in remote sensing", LARS PRINT 030772, Purdue University, Lafayette, Indiana, 1972.