SUPPORTED MAPPING WITH MULTI SENSOR IMAGES THROUGH STRATEGY FOCUSED ON CUSTOMIZATION AND INTEGRATION OF GENERALIZED CLASSES BY GEOBIA

C.B.M. Cruz [a], P.M.M. Almeida [a,b], R.S. Barros [a], R.S. Vicens [c], E.M.F.R. Souza [a], E.P.A. Caris [a], M.C. Fernandes [a], P.M.L. Menezes [a 1*]

[a] Dept. of Geography, Federal University of Rio de Janeiro - (cmad, almeida.pmm, barros.rafael, manoelcoutofernandes, elizabethmfr, elizabethmfr, pmlmenezes)@gmail.com
[b] UCB, Castelo Branco University, Dept of Teachers Formation (Geography)
[c] UFF, Dept. of Geography - rsvicens@gmail.com

KEY WORDS: Land use, Land cover, Data Integration

ABSTRACT:

Rio de Janeiro presents itself as a land of contrasts, with the second largest metropolitan area in Brazil and a total area of 43,778 km^2. Its landscape consists of remnants of the Atlantic Forest and environments with different levels of anthropic impacts. The complexity of its territory combines natural and anthropic covers related to each other at different intensity levels, being a challenge in defining mapping techniques of land use and land cover with greater spatial and temporal detail. Supported by the Secretary of State for Environment, this initiative accepted the challenge of developing a methodology for mapping about 45% of the state in 1:25,000. This mapping aims to support decision making regarding the land use planning and the monitoring of deforestation actions. Using GEOBIA techniques and images of different resolutions, we structured a methodology for the classification of four macro-classes very different spectrally (Natural Forested Areas, Natural Non-Forested Areas, Anthropic Agropastoral Areas, Anthropic Non-Agropastoral Areas), identifying objects on the ground from 0.5ha, to meet demands from the Rural Environmental Registry (CAR, in portuguese) and the forest monitoring. The mapping was divided into Working Groups (WG) that customized solutions through Process Trees in eCognition environment. This choice is based on the need to minimize inconsistencies in the interpretive process, improving the level of specialization in short time. Thus, each WG is responsible for the classification of a set of functionally related classes. The integration process of the classes in a single mapping was also supported by GEOBIA. It is believed that the presentation of this strategic view may contribute to challenges of similar mapping, allowing the achievement of cartographic goals in short term.

1. INTRODUCTION

Institutional mapping aiming at applications related to land use planning is hardly anchored in complex methodological solutions that aim at improving results. When the definition of unusual methods which can be implemented in short time, ensuring quality results, is necessary, it is common for the State and research groups, usually linked to universities, to form partnerships.

This paper presents the experiment for mapping, in the 1:25,000 scale, part of the Atlantic Forest corridor, located in the state of Rio de Janeiro, area of fundamental importance for the maintenance of the water resources, the forest remnants and the biodiversity. The state, with approximately 43,000km^2, presents the main remnants of the Atlantic Forest in Brazil, which occupy about 30% of its territory. The current mapping covers an area around 20,000km^2, composed by four hydrographic regions of the state, with a minimum mapping area of 0.5ha.

The map intends to offer support to decision-making statewide, related both to issues from the Rural Environmental Registry (CAR, in portuguese) as to actions of forest conservation and recuperation. CAR's goal is to promote identification and integration of environment information of rural properties and possessions, aiming at environmental planning, monitoring, regularization and combating deforestation. As rural properties in the state of Rio de Janeiro are mostly small, the aid of the government through technical support is provided by law. Thus, a system with the necessary bases for the survey of spatial information to define Permanent Preservation Areas (PPA) and Legal Reserves (LR) is indicated. In the specific case of the state of Rio de Janeiro, greater detail scales are desired to meet the demands. Thus, this proposal is justified by the need for compliance with legal requirements related to CAR, in addition to related interests for the forest monitoring with great spatial and temporal detail, in support of the fight against deforestation, among other malpractices.

It is noteworthy that there is no consolidated solution to generate a mapping of this nature, especially if we include the need for short deadlines for implementation. This framework points to the integration of research from different academic groups that, together and supported by experienced professionals from other institutions, participated in this technical solution.

The work is anchored on a Partnership Agreement celebrated together by the company *Porto do Açu Operações S.A.* and the institution Coordination of Projects, Research and Technology Studies (COPPETEC/UFRJ), having as an intermediate the Secretary of State for Environment (SEA, in portuguese) of Rio de Janeiro. The Laboratory of Remote Sensing and Environment Studies ESPAÇO of UFRJ, with the support from

1 * Corresponding author

the laboratories of Cartography GEOCART (UFRJ) and of Physical Geography LAGEF (UFF), built the methodological solution of this work.

The mapping area is composed by the Hydrographic Regions (HR) of the Médio Paraíba do Sul, Piabanha, Dois Rios and part of the Baixo Paraíba do Sul and Itabapoana, respectively identified by HR III, IV, VII and IX (figure 1).

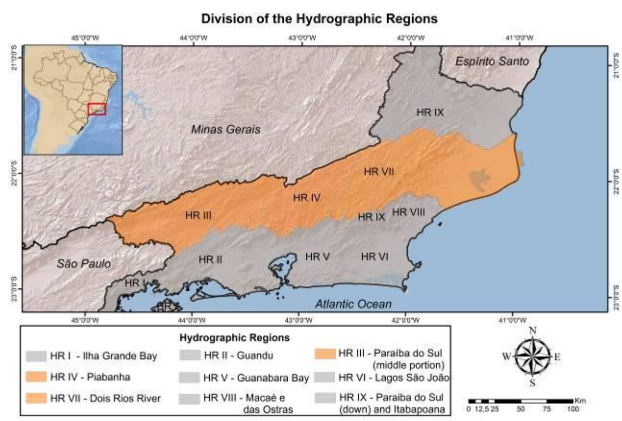

Figure 1: Division of the Hydrographic Regions of the state of Rio de Janeiro (the mapping area is identified in orange)

The mapping was based on orbital imaging through different sensors and spatial resolutions. The integration of different sensors aimed at meeting the peculiarities of each class, as the temporal dynamics and the best spectral and spatial characterization.

The legend proposed for the final mapping was defined by two hierarchical levels, which were developed in two steps (Table 1).

MACRO-CLASSES	DETAILING
NATURAL FORESTED AREAS (ANF, in portuguese)	Forested Coastal Sandbar Vegetation and Mangrove
NATURAL NON-FORESTED AREAS (NNF)	Non-Forested Coastal Sandbar Vegetation
ANTHROPIC AGROPASTORAL AREAS (AAG)	Silviculture
	Areas of Consolidated Use (previous to June of 2008)
ANTHROPIC NON-AGROPASTORAL AREAS (ANA)	
WATER BODIES	

Table 1: Mapping Legend

Each macro-class corresponds to a grouping of different, but functionally related, covers. This is another complexity for the knowledge modeling in support of the classification by GEOBIA. The classification was performed in hierarchical levels, supported by spectral, temporal, thematic and numeric data, in addition to the specialized knowledge about the spatial context in which each class is inserted.

Working Groups (WG) were defined in order to focus, in a specialized way, on each macro-class. The working teams were leveled so that the used interpretive criteria were homogeneous. The classification method to be adopted considered the most advanced technological processes available:

- Integration of multisensor and multidate data.
- Knowledge modeling, organizing bases and thresholds defined in the bibliography which helps us to build rules for classification.
- Data mining assisting the definition of parameters which characterize the classes, when the previous knowledge is considered not sufficient.
- Classification based on geographic objects, which allows the use of complementary data to the spectrum, increasing chances of achieving more complex legends.

The final product sought consistency in terms of connectivity for the entire state of Rio de Janeiro (link between hydrographic regions) and its structure is compatible to the manipulation through Geographic Information Systems (GIS).

2. DATA PRE-PROCESSING

In order to meet the proposed objectives, we used images acquired by four sensor systems with different resolutions: TM/Landsat 5 of 2008, OLI/Landsat 8 of 2014, RapidEye of 2014 and Worldview-2 of 2014, in addition to the Digital Elevation Model (DEM) from the SRTM mission. The year of 2014 was taken as reference for the mapping, being accepted images acquired within one year before or after this reference. In some few areas it was not possible to obtain RapidEye and Worldview-2 images without cloud cover in the proposed period, being necessary to move further away from the year of reference.

The Landsat 5 and 8 images were acquired for free on the USGS website with the spectral bands of visible light, near infrared and short wave infrared, all of them with 30 meters of spatial resolution and orthorectified. These images were used mainly for the detection of changes in order to identify the areas considered consolidated (agropastoral use previous to June of 2008). The consolidated areas need to be identified throughout Brazil on account of CAR, which deals with such areas in a different way.

The RapidEye images are acquired with 6.5 meters of spatial resolution and provided orthorectified with pixels resampled to 5 meters. In Brazil these images were acquired by the Ministry of the Environment (http://geocatalogo.mma.gov.br/) in a license option that allows them to be available to all government agencies with no extra cost, which encourages their use. For containing the near infrared band, in addition to the bands of visible light, these images were useful for generating indexes that allowed the identification of certain classes.

The Worldview-2 images were the main product in this mapping, being acquired from the DGBM (DigitalGlobe Basemap) system, which had five acquired licenses for one year for the development of this project, as well as for the generation of other products for the Secretary of State for Environment. The detailing presented by these images, as well as the relatively lower cost of acquisition by the DGBM was really important in this project. The fact that the license is valid for one year was very interesting, because it allowed new images to be acquired whenever necessary, for example for scenes with excessive presence of clouds or with problems in the geometric correction. Some major disadvantages were also observed, such as the absence of the near infrared band, which is very important in studies related to vegetation, and the problem of the great radiometric diversity observed between different scenes, which required a great effort of radiometric equalization. The images available on DGBM correspond only to the bands of visible light, merged with the panchromatic and resampled to 8 bits.

The SRTM DEM was acquired with pixel of 1 arcsecond (approximately 30 meters in Ecuador) and it was very useful for the separation of classes that can be differentiated, among other factors, depending on the altimetry, as was the case of high fields and mangroves.

The methodology predicted that the preliminary effort of pre-processing would be performed by a specific WG, including the preparation of the seed KIT for classification. Thus, all the WG received a standard eCognition project, or seed, to start their classification. This caution was taken, among other things, to facilitate the integration of the classes for the final mapping. This way the images would be distributed in KITs corresponding to pre-defined parts of the official Hydrographic Regions of the state of Rio de Janeiro. The mapping area was divided into 15 parts or KITs, containing, in addition to the images and mentioned DEM, the same segmentation and the mapping of the existing water bodies (water class). The segmentation was generated, after several tests, with a Scale Factor of 30, and weight of 0.5 for shape and 0.1 for compactness. The figure 2 presents the division in subprojects.

All the images were evaluated in terms of geometry and radiometry before being used in the mapping. The Landsat 5 and 8 images were obtained with orthorectification and atmospheric correction directly from the USGS website. Because they have already been the subject of several geometric reviews by our research group, it was decided not to assess the Landsat images in this project (CRUZ and BARROS, 2012). Considering their role of supporting the detection of changes on the cover, it was performed the radiometric normalization of the images of 2008 (Landsat 5) and the current images (Landsat 8). The equalizations were made on the Orthoengine software from PCI Geomatica 2015, while the normalization was made on the R software (CANTY et al, 2004; SCHROEDER et al., 2006).

The RapidEye and Worldview-2 images required more caution in the pre-processing step. The RapidEye images used in this mapping had the support of the official Cartographic Base of Brazil in the 1:25,000 scale for the orthorectification, which contributed for a better geometric correction. The Worldview-2 images are available orthorectified on DGBM. Thus, a geometry review was made comparing the relative displacement between the Worldview-2 and RapidEye images, as well as their absolute position using 20 control points obtained from the 1:25,000 cartographic base, identifiable in the images available for the teams. The displacements between the images and the cartographic base were below 15 meters in the tested areas and were acceptable to the 1:25,000 scale.

After several tests involving the detailing desired for the mapping in the 1:25,000 scale and the processing capacity and available storage, yet combined to the mapping deadline, it was decided to resample the Worldview-2 images to 2 meters. This way we optimize the processing without compromising the interpretation to the final scale.

Both RapidEye and Worldview-2 images passed by the radiometric equalization on Orthoengine from PCI Geomatica 2015, in order to minimize the problems caused by the large radiometric variation observed. In figure 3 it is possible to see the variation between the Worldview-2 images that compose the KIT, named HR3C, before and after the equalization.

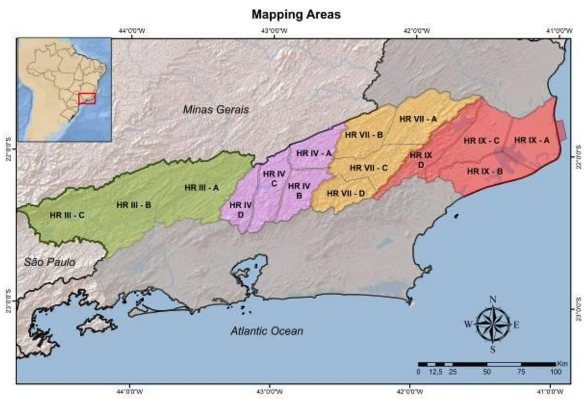

Figure 2: Subprojects for mapping

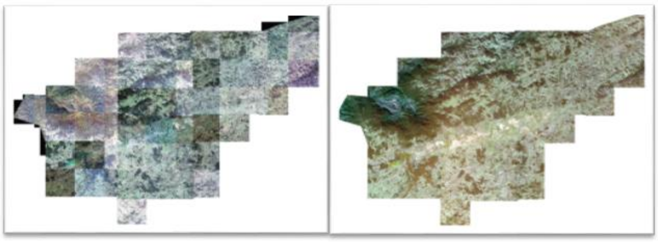

Figure 3: Effect from the Radiometric Equalization between Wordview-2 scenes

After the pre-processing steps, the elaborated KIT for each subproject was delivered with all the mentioned data, as well as the mapping of the Water Bodies class, thus ensuring the further integration of the parts. From these KITs, the teams started the mapping of the macro-class under their responsibility.

3. KNOWLEDGE MODELING AND CLASSIFICATION

3.1 Mapping of the Macro-Classes

The choice of specialized modeling by WG was fundamental for the conclusion of the mapping in short term, because it allowed the fast leveling of the photo interpreters and the enhancement of the method along the sequential classifications.

The specialized modeling by macro-classes presents benefits, such as: i) lower dependence on previous know-how by the team – once each group is able to focus on the modeling of a single class; ii) mapping speed; iii) continuous verification of commission errors by class/WG; iv) identification of the most complex classes for modeling, candidates for further increase of efforts.

The four macro-classes mapped in this work are detailed in table 2, which presents the set of functional classes that were grouped in each one of them. The subclasses diversity demanded a lot of effort on the modeling and the classification.

CLASS	DEFINITION	DETAILING
Natural Forested Areas (ANF)	It comprises a set of forest structures on different successional stages of development,	Dense Forest, Open Forest; Seasonal Forest; Mixed Rain Forest, Coastal

	distributed by different environments and geographic situations.	Sandbar vegetation and Mangroves. Different conservation states were included.
Natural Non-Forested Areas (NNF)	Natural areas with, eventually, non-forested vegetation (shrubby or herbaceous), in addition to other natural non-forested indentation.	Relic communities; outcroppings; active gullies; coastal sandbar vegetation, marsh areas.
Anthropic Agropastoral Areas (AAG)	Areas used for production of food, fibers and agribusiness commodities. Both heterogeneous agricultural areas and extensive pasturelands.	Temporary and permanent crops; planted pastures; silviculture; in addition to proven agropastoral areas. Then the consolidated areas were separated (previous to June of 2008).
Anthropic Non-Agropastoral Areas (ANA)	All types of land use of non-agropastoral, non-forested and non-water nature.	Urban areas (villages, cities, industrial complexes); mining areas; exposed soil areas, parks.

Table 2: Detailing of the Macro-classes

Each WG worked independently on the definition of the best way to identify their mapping class. For this, they have considered: (i) the team's previous knowledge on the software and image interpretation; (ii) the experience with GEOBIA; (iii) the spectral characteristics of the targets inserted in the class and; (iv) the scope of the class. These two last aspects, in addition to the complexity of the used data, reflected in the difficulty to achieve a single classification model, replicable for all the 15 projects.

The ANF class presents great extent on the mapping area, mainly on the HR III and IV, which comprise part of the Mountain Region of the state. The modeling of this class was one of the most complex, because its distribution occurs in different lighting conditions due to sharp relief and the incorporation of different vegetation structures, with different compositions and conservation states. The team chose to use, on each project, the tiles division feature available at eCognition, starting with a general classification model, for further tile-to-tile adjustments. In a general way, the scale factor of the selected segmentation varied between 20 and 50, considering only the WorldView images. The main adopted descriptors were: the first PCA component; HUE and Intensity; in addition to an extra descriptor, resulting of the subtraction of HUE by PCA.

The modeling of the NNF class was performed with three different approaches, proper for the identification of: (i) altitude fields and rocky outcroppings; (ii) other types of outcrops and; (iii) exposed soils. For the altitude fields, the adopted technique was the GEOBIA, through the NDVI average (of the RapidEye image) and the altitude (SRTM). As the outcroppings and exposed soils have a very punctual occurrence, their identification was performed through visual classification.

The AAG class has the largest spatial expression in the area. It presents significant spectral diversity, as it incorporates both “green” and “non-green” regions, and compared to this characteristic the modeling of this class followed the top-down logic that first separates “green” and “non-green” targets, further applying data mining to eliminate areas which do not belong to this macro-class. Within this class, the silviculture areas were treated in a particular way. Their identification used a mixed of techniques, including the use of a previous mapping from 2009 and some textural patterns (AMORIM et al., 2012).

The greatest challenge for mapping the ANA class was its subclasses' heterogeneity. One of the greatest difficulties was the separation between urban and rural areas. In the case of urban areas, different construction intensities were found, besides the presence of exposed soils and small green areas. The relief was another difficulty factor, because many cities were shaded, being hard to delimit. The variation of the angle of incidence of high-resolution sensors was revealed to be problematic in these cases. Thus, a top-down and multiresolution approach was chosen, starting from macro segmentation with scale parameter between 40 and 50, based on the Landsat and RapidEye images. In the modeling, the main adopted descriptors were the Red band average of the RapidEye and Landsat images.

Each WG was responsible for the reintegration of the parts of each Hydrographic Region. This reintegration was performed on the original project with the seed segmentation through the use of mode function. Thus, a single segmentation for the mapping and the association of the predominant class for each object was reestablished.

3.2 Mapping Integration

The mapping integration of all WGs left off the individual delivery of each macro-class and the silviculture class, in raster format, with two identified classes: the one in the WG's interest and the area of no interest (or areas not classified by the WG). The spatial resolution of these five raster files was equal to the 2m original. Choosing the raster structure is justified by the fact that the classes integration is made through the *Combine Raster* function on ArcGIS, which was shown to be the most operational way, given the data volume involved.

The *Combine Raster* results in a combination of all classes against all classes, presenting, at the end, a raster file with the total number of combinations found. In order that the operation occurs satisfactorily, the original files must be organized. The table which accompanies the *Combine raster*, was done on Excel, where the integration statistics were more easily calculated, allowing the establishment of rules for the next step of the integration, on eCognition. The calculated statistics by HR, were: (i) percentage of classification without confusion (pixels identified by a single WG); (ii) percentage of unclassified areas (pixels not identified by any WG); (iii) percentage of the confusions diversity between WG, being separated by types of confusion. In average, all the HR presented the following percentages: (i) 80 to 90% of classification without confusion; (ii) 2 to 8% of unclassified areas and; (iii) 10 to 15% of confusion areas. The *Combine* file was then regrouped in a set of 12 to 15 classes, representing the observed arrangements. For each situation, the best actions to be taken in support of the final integration modeling on eCognition were discussed in group.

The final project created on eCognition was structured with the images used by the WG, the most detailed segmentation (20 or 40) and the Combine file regrouped in raster format. Figure 4 presents the main steps performed on eCognition, which aimed at solving the identified confusions and omissions.

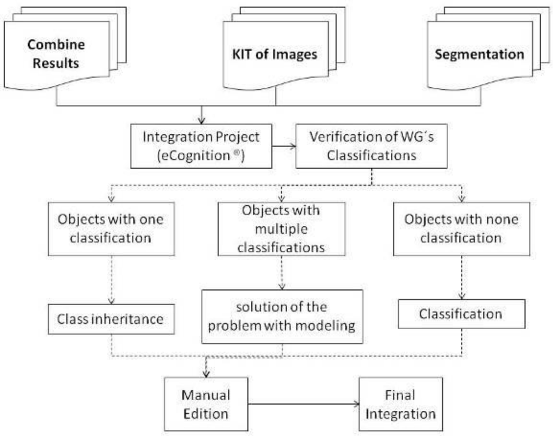

Figure 4: Flowchart with the integration steps on eCognition

First, the mapping areas without confusion were identified, associating the objects of segmentation by mode to the macro-classes. Next, the confusion areas, mapped by more than one WG, were treated. Areas smaller than 0.5ha were ignored in this step for further treatment with generalization operations. Confusions between the ANF and AAG classes were predominant in all HR, which was expected given that they are the classes with the most spatial expression in the area and for the difficulty to separate the initial secondary vegetation, usually in transition areas. The modeling used, mainly, the NDVI average of the RapidEye image. Smaller confusions, between other classes, were treated in a particular way. The areas not classified by any WG were modeled from spectral characteristics and context.

After the modeling, the obtained results passed by a manual edition, prioritizing more complex cases. The final mapping, exported from eCognition, passed yet by generalization operations on ERDAS (*Clump* and *Eliminate features*) to eliminate noise and small unidentified areas. Redefining the topology was one last operation performed by the ArcGIS software.

Each HR, totally integrated, was returned to the WG for a last verification of the decisions taken regarding the unclassified and confusion areas. The final product was taken to the field for validation and the elimination of possible doubts.

Figure 5 presents the final layout of the Hydrographic Region of the Médio Paraíba do Sul, HR III, where it can be perceived the predominance of the ANF (Natural Forested Areas) and AAG (Anthropic Agropastoral Areas) classes.

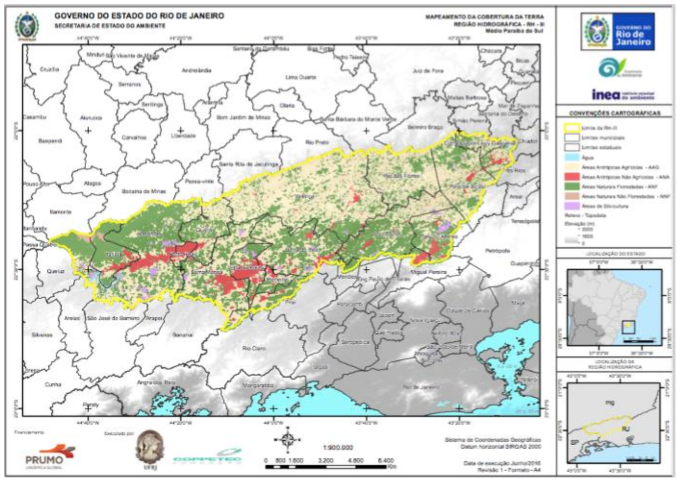

Figure 5: Final layout of the Hydrographic Region of the Médio Paraíba do Sul, HR III. (Legend: ● Water Bodies; ● AAG; ●ANF; ● NNF; ● ANA and ●Silviculture

4. RESULTS AND DISCUSSION

The use of GEOBIA in cartographic production is still rare in Brazil, in spite of the country having an enormous territorial size (approximately 8.5 million km^2) and presenting innumerous difficulties for the generation and update of several mappings, as much as in high detail scale, as in thematic.

With experience in other two great mapping initiatives supported by GEOBIA, also aimed at cartographic production in partnership with governmental institutions, the research group from the ESPAÇO laboratory, coordinated the mapping of the Atlantic Forest's biome (CRUZ et al., 2007) in the 1:250,000 scale and the land use and cover of the state of Rio de Janeiro regarding the Ecological and Economic Zoning (CRUZ et al., 2009) in the 1:100,000 scale. Both experiences presented interesting results, in terms of team learning and knowledge improvement for modeling, as well as the investments and deadlines.

The present experiment contributes to solutions that require detailing (1:25,000 scale), accuracy and agility on execution. The total performance time of mapping an area of approximately 20,000 km^2, with a minimum area of 0.5ha, was divided into: (i) 4 months of pre-processing; (ii) 6 months of mapping and integration (simultaneous with 2 months of pre-processing); (iii) 2 months for final editions and validation.

The greatest difficulty in this method was the absence of infrared bands on the DGBM (DigitalGlobe Basemap) system, specially the near infrared, demanding the incorporation of another sensor, RapidEye's, and more efforts of processing and data integration; in addition to increasing the data volume to be treated on eCognition and difficulties in modeling. Currently the DGBM system is starting to incorporate images with NIR.

The common cartographic challenge of mapping extensive areas with great detail requires methodological solutions which offer viability and a good relation between cost and benefits. In Brazil, a country whose cartography lacks detail and, mainly, update, it is foreseen that investments would be made.

The project's current step is the validation by obtaining field data. In spite of still not having the accuracy obtained on the final mapping, the integration process of the parts through

Combine Raster allowed the comprise of greater weakness and certainties of the mapping, directing the following edition step.

Thus, the analysis of the mapping final accuracy is in progress. The field works for collecting points are being structured with mixed teams of members from all the WG, increasing the decision's synergy.

REFERENCES

AMORIM, H.B.; FRANCELINO, M.R.; SALAMENE, S.; PEDREIRA, L.O.L.; ASSUMPÇÃO FILHO, L.I.; CAPITANO, R.C.; MOURA, T.A. Estimate of the Area Occupied by Reforestation Programs in Rio de Janeiro State. Cerne, Lavras, v. 18, n. 1, p. 27-32, jan./mar. 2012.

CANTY, M.J.; NIELSEN, A.A.; SCHIMIDT, M. Automatic radiometric normalization of multitemporal satellite imagery. Remote Sensing of Environment, 91, 441-451, 2004.

CRUZ, C.B.M.; BARROS, R.S. Contribution of New Sensors to Cartography. In: Cartography, Chapter XX, Editora InTech, 2012, v.01, 20 pg. ISBN 979-953-307-910-2.

CRUZ, C.B.M.; ROSÁRIO, L.S.; ABREU, M.B.; ALMEIDA, P.P.M.; VICENS, R.S.; CRONEMBERGER, F.M. Classificação Orientada a Objetos na Geração do Mapa de Uso e Cobertura da Terra do estado do Rio de Janeiro. XIV Simpósio Brasileiro de Sensoriamento Remoto, INPE. 7789-7796. Natal, RN. ISBN: 978-85-17-00044-7. INPE, 2009.

CRUZ, C.B.M.; VICENS, R.S.; SEABRA, V.S.; REIS, R.B.; FABER, O.A.; RICHTER, M.; ARNAULT, P.K..; ARAUJO, M. Classificação Orientada a Objetos no Mapeamento dos Remanescentes da Cobertura Vegetal do Bioma Mata Atlântica na escala 1:250.000. XIII Simpósio Brasileiro de Sensoriamento Remoto, INPE. 5691-5699. Florianópolis, SC. ISBN: 978-85-17-00031-7. INPE, 2007.

SCHROEDER, T.A.; COHEN, W.B.; SONG, C.; CANTY, M.J.; YANG, Z. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. Remote Sensing of Environment, 103, 16-26, 2006.

SOKOLONSKI, H.H. "Manual Técnico de Uso da Terra." Rio de Janeiro: IBGE (1999).