REGIONAL SCALE LAND USE/LAND COVER CLASSIFICATION USING TEMPORAL SERIES OF MODIS DATA

M. Jonathan [a, *], M. S. P. Meirelles [a, b], J-P. Berroir [c], I. Herlin [c]

[a] Dept. of Computer Engineering – Geomatics, State University of Rio de Janeiro (UERJ), 20550-900, Rio de Janeiro, RJ, Brazil - mjonathan@mail.com

[b] National Center for Soil Research – Brazilian Agricultural Research Corporation (Embrapa-CNPS), 22460-000 , Rio de Janeiro, RJ, Brazil - margaret@cnps.embrapa.br

[c] The French National Institute for Research in Computer Science and Control (INRIA), BP105 - 78153 - Le Chesnay, France – {jean-paul.berroir, isabelle.herlin}@inria.fr

KEY WORDS: remote sensing, multitemporal image processing, land use classification, land use change detection, MODIS

ABSTRACT:

This paper describes a methodology for systematic land use/land cover classification on a regional scale, with emphasis on a low cost and highly automatized approach. This methodology is based on multitemporal analyses of surface reflectance data from the Moderate Resolution Imaging Spectroradiometer (MODIS), which is located on board NASA's Terra and Aqua satellites and features high temporal frequency, extensive coverage, and extremely low costs for data acquisition. A sequence of automatized procedures were developed for MODIS data pre-processing, as well as for the training and execution of a supervised classification algorithm, where temporal profiles are fitted to smooth polynomial curves and intelligent curve features are then computed in order to reduce data dimensionality and improve profile interpretability, thus providing a more robust classification approach. A case study was performed in the High Taquari Basin, in the states of Mato Grosso do Sul and Mato Grosso, Brazil, which showed that the method was indeed capable of generalizing well over the entire region of study (over 25,000km^2), effectively discriminating between areas of agriculture, pasture, and savannah. The methodology was also seen to be quite successful in identifying areas of deforestation, which is of particular interest for the monitoring of land use and land use change in the region.

1. INTRODUCTION

It has already been widely recognized that land use and land cover (LULC) changes play a very important role on regional to global scales, with impacts over ecosystem functioning, ecosystem services, and biophysical and human variables such as climate and government policies (Meyer and Turner, 1994). However, even though LULC assessments using high-resolution remotely sensed images have been quite successful over the last years, it has become clear that using this approach to analyze large areas on a regular basis ends up yielding prohibitive computational and financial costs.

As an alternative, several authors have proposed the exploitation of the rich temporal information contained in sequences of freely available coarse resolution satellite data (Holben and Shimabukuro, 1993; Bouzidi *et al.*, 2000; Meirelles *et al.*, 2004). Traditionally, many of these approaches have employed data from NOAA's AVHRR sensor (1.1km resolution), but nowadays data from NASA's MODIS sensor is also available, featuring better spatial resolution (up to 250m) and superior standards of calibration, georeferencing and atmospheric correction, as well as detailed per-pixel data quality information.

As such, a number of researchers have started to apply this kind of data for land cover assessments (Strahler *et al.*, 1999; Lobell and Asner, 2004; Wessels *et al.*, 2004). In this work, a methodology for supervised LULC classification on a regional scale is proposed, with specific application to the Cerrado tropical savannah biome in mid-western Brazil. Additionally, the particular use of the methodology for detecting areas of deforestation is also evaluated, in view of the fact that this corresponds to the most critical land cover change process in the region.

2. STUDY AREA

The study area used in this work corresponded to the High Taquari Basin, which is almost entirely contained within the state of Mato Grosso do Sul, in mid-western Brazil (Figure 1). A small portion in the northern part of the basin is located in the neighboring state of Mato Grosso. According to the definition given in (Silva, 2003), the High Taquari Basin is limited by the coordinates (17^{o}10'S, 53^{o}10'W) and (19^{o}45'S, 55^{o}10'W), and comprehends a total area of 28,046 km^2.

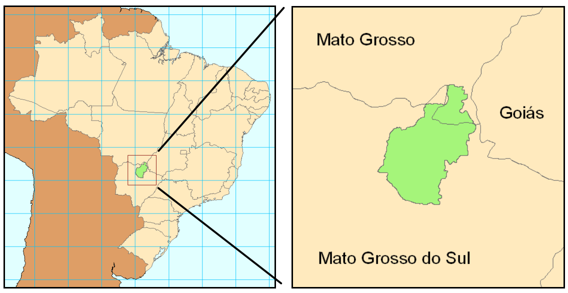

Figure 1. Location of the High Taquari Basin, the study area for this work, in mid-western Brazil

The Taquari river itself is a very important tributary of the Paraguai river and plays an essential role within the Pantanal ecosystem, an area that was designated a UNESCO World Biosphere Reserve in 2000 due to its extremely rich and unique biodiversity. Unfortunately, however, the river has been suffering a severe silting-up process due to uncontrolled land use expansion in the High Taquari Basin, leading to an increase in soil erosion processes that were seen to have grown exponentially over the last 25 years (Godoy *et al.*, 1999). As such, this phenomenon is already being considered the cause of the most important environmental and socio-economical problems within the Pantanal region, leading to such consequences as inundations, reduction of natural pasture areas, and impacts over animal and vegetal life cycles.

For these reasons, the monitoring of land use and land cover in the region can be seen to be an extremely important and urgent issue. Nevertheless, the surveillance of the entire basin corresponds to an expensive and complex effort, and it is thus not currently viable on an operational basis. Additionally, the basin has already been the subject of numerous studies (*e.g.*: Godoy *et al.*, 1999; Bueno *et al.*, 2003; Silva *et al.*, 2003), and therefore it can be considered particularly suitable as a pilot area for research.

3. MATERIALS

In order to perform this work, daily Terra/MODIS atmospherically corrected surface reflectance data were acquired (product MOD09GQK, containing red and near-infrared bands and corresponding metadata at 250m resolution). Images were obtained so as to cover the High Taquari Basin during a one-year period from August 2000 to July 2001, a period that corresponds to the local annual vegetative cycle. In addition to that, 1km resolution quality metadata (product MOD09GST) were also acquired for the same period, which uses information from several other MODIS bands in order to infer more detailed information regarding cloud cover, aerosols, and other special conditions.

Other than that, a detailed LULC classification for the entire basin was obtained for the year 2000, which was based on LANDSAT TM imagery from July 2000 and extensive field trips (Silva, 2003). This classification was to be used as "ground truth", so as to provide means to train and then test the methodology described in this paper. It contained 14 classes: agriculture, water bodies, eucalyptus, open savannah, close savannah, dense savannah, lower montane forest, alluvial forest, 4 mixed natural vegetation classes, pasture, and urban areas.

4. METHODOLOGY

4.1 Data Pre-processing

Once all data were available, the daily MODIS images and metadata were processed in order to select the best quality data and filter out areas contaminated by clouds, cirrus, shadow and high levels of aerosol. For this purpose, the detailed 1km metadata were used to screen out images that did not meet minimum clear-sky and average observation coverage conditions (a measure related to viewing angle), leading to a selection of 71 better quality images. Then, for these remaining images, each pixel was evaluated by combining the detailed 1km metadata with the 250m quality and coverage metadata. As such, a correction algorithm was implemented so that pixels rejected by the evaluation process were replaced using linear temporal interpolation. Finally, it was also necessary to register the resulting MODIS image sequences to the available classification image for the year 2000.

Indeed, it was observed that the MODIS images did present good georeferencing properties, but, on the other hand, some limited problems related to cloud and shadow detection were also noticed. These made it necessary to develop and apply a number of filtering processes in order to achieve more trustworthy temporal data sequences. After that, temporal profiles for the red and near-infrared bands were computed, as well as temporal profiles for the Normalized Differential Vegetative Index (NDVI). This index, which is based on a ratio of the red and near-infrared bands, is given by:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

where:

- *NIR* corresponds to the reflectance value in the near-infrared band
- *R* corresponds to the reflectance value in the red band

In fact, it has been shown (Sellers *et al.*, 1992) that NDVI is directly related to the amount of photosynthetically-active radiation intercepted by the vegetation canopy, and thus it is widely used for differentiating areas that contain healthy vegetation. In addition to that, it also displays interesting properties of being relatively robust to atmospheric interference, slope, illumination variations, and other effects that influence both bands (Lillesand and Kiefer, 2000). Therefore, being simple to compute, easy to interpret, and relatively robust, it is one of the most common measures used for LULC assessment. For these reasons, it will be used extensively throughout this work.

At this point, the study area was arbitrarily divided in two halves (north and south) so that the classes were well represented in both of them. As such, the southern part was to serve as a general area of training and testing, while the northern part was kept aside to be used exclusively for validation purposes. Such a sharp cut was more interesting than a random selection of points, since distinct areas may present variations in spectral response due to different geographical and meteorological status. Thus, this approach would help to test the ability of the methodology in dealing with larger regions, where differences in climate and local conditions may become more important.

Finally, it was also observed that, given the 250m spatial resolution of the MODIS data, approximately 70% of the pixels within the area of study were seen to cover only one single class (i.e., to be "pure" pixels). This way, it was decided that using only pure pixels would be an interesting approach for many of the analyses to be subsequently performed.

4.2 Adaptation of the Original Classification Information for MODIS Data

After that, a process of class merging and splitting needed to be performed on the original classification data, so as to define a set of meaningful classes effectively discernible with MODIS temporal reflectance information. This procedure was performed in two steps: first, NDVI temporal profile means and variances were analyzed in order to assess class separability, and classes considered to be non-separable were merged together; after that, a *clustering* process was performed on the observed temporal profiles for the training data using a *k-means* clustering algorithm, in order to identify whether the original classes included more than one sub-class with respect to the MODIS temporal profiles. Indeed, this phenomenon was observed for the *agriculture* class, which could be successfully subdivided into distinct agriculture practices, and for the *savannah* classes, where deforestation behavior could be identified (Figure 2). As such, after this merge/splitting process, the original 14 classes were finally grouped into 7 classes considered to be discernible with MODIS red and near-infrared temporal data: 3 distinct agriculture practices, savannah, pasture, urban areas, and deforestation. The original *lower montane forest* class did indeed present a distinct reflectance profile, but it was considered to be too insignificant at a spatial resolution of 250m in order to be considered as a separate class for assessments with MODIS in the region. Finally, it should be noted that, in this area of study, the temporal reflectance profiles for class *urban areas* simply reflect severe lack of vegetation and are thus equivalent to profiles for areas of bare soil.

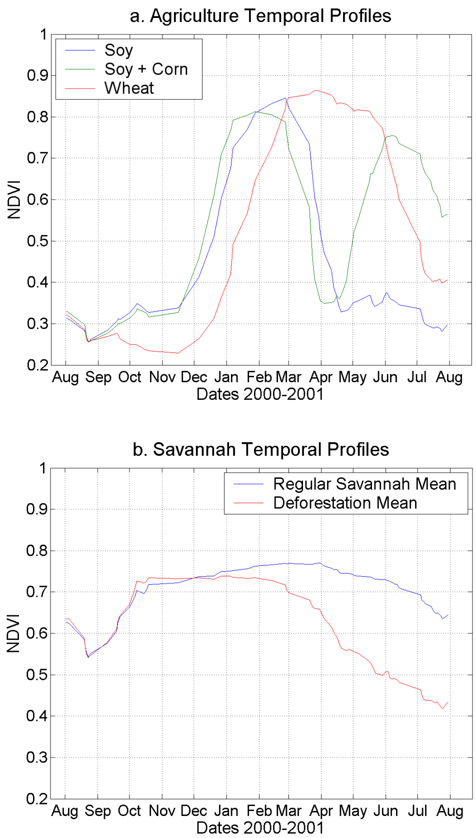

Figure 2. Sub-classes identified by clustering the original classification training data. Splitting the original classes for agriculture and savannah, it was possible to detect a) distinct agriculture practices and b) areas of savannah that suffered deforestation during the period of analysis

4.3 Feature Computation and Selection

The first step in this procedure consisted of modeling the NDVI temporal profiles by fitting a smooth curve to the previously computed data. As such, this curve fitting process would provide a reduction of dimensionality that should prove to be very important for the classification algorithm (it should be noted that the original profiles have dimension 71, which corresponds to the number of better quality images selected for the sequence). Moreover, the fitting would also help to minimize the residual noise and to increase profile interpretability. More specifically, the idea was that the resulting curves would be used as a basis for computing several curve features that should provide more intelligent and compact information about the profiles, thus helping the classification process. Thus, the profile features computed were to include attributes such as curve mean, slope measures, number of modes, and date of maximum value, among others (Figure 3).

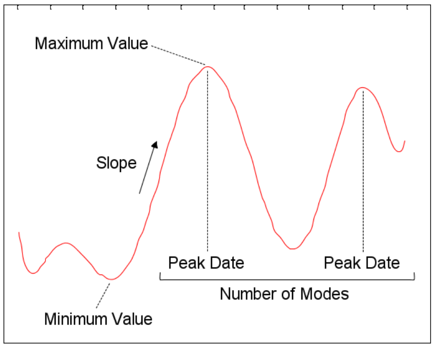

Figure 3. Examples of profile features computed for a temporal profile

Once these several profile features were computed, a *feature selection* process was performed so as to determine the subset of features most appropriate for discriminating between the 7 classes defined before (section 4.2). For this purpose, the Correlation-based Feature Selection strategy (CFS) was employed, which corresponds to an information theory approach based on two steps: a search strategy (*e.g.*, *best first search* or *genetic algorithms*), and a measure of merit/fitness of a subset of features (Hall, 1999). This measure is actually based on the mean correlation between each feature *i* (independent variables) and the classification (dependent variable), as well as on the inter-correlations between each pair of features *i* and *j*, correlations of which are all computed using entropy and information gain approaches.

This way, in the end of all this process, a subset of 12 features was selected to be used for subsequent temporal profile classifications, namely: *NDVI mean, NDVI minimum, date of maximum NDVI, beginning NDVI, global NDVI amplitude (max – min), global NDVI gain (end – begin), 150-day NDVI gain (mean of the first 215 days – mean of the last 150 days), number of modes, mean of the absolute slope values, standard deviation of the unfitted data values, date of the peak of the main mode, width (in days) of the main mode*.

Therefore, it was possible to drastically reduce the dimensionality of the input data (from 71 to 12) and also make it more robust to slight profile variations due to intra-class variability and noise, thus facilitating the classification process as a whole.

4.4 Classification

At this point, the process of *regional scale classification* could begin. Recalling that approximately 70% of the 250m MODIS pixels turned out to be pure with respect to the original classification, it was assumed that assigning a single class to each pixel (*i.e.*, performing a "hard" classification) would be appropriate for most part of the basin. Moreover, considering the elaborate process of profile feature computation and selection, it was expected that at this point the data would be well treated and able to provide good separability between the classes. As such, any well-established classification approach should then be capable of achieving satisfactory results. Therefore, it was decided that the maximum likelihood classification algorithm (ML) would be an appropriate choice, since it is an extremely simple and easily implemented algorithm, but that at the same time is very well known and has already been successfully applied to a broad range of remote sensing problems (Lillesand and Kiefer, 2000).

The basic ML algorithm is a purely statistical approach based on Bayes' formula, which assumes that the variables (*i.e.*, the features) are continuous and follow a multi-dimensional gaussian or normal distribution. However, in the particular case of this work, one of the features selected for classification turned out to be actually a *discrete* variable (the number of modes in the temporal profile). Therefore, it was important to modify the basic algorithm in order to treat this variable appropriately, since it would make the procedure more consistent and also allow for the existence of classes with variance zero for this feature (*e.g.*, classes whose profiles never have a mode). The model was thus extended so as to admit one discrete variable, so that the posterior probability $P(w_j|x)$ for a class w_j given an input set of features x is given by:

$$P(w_j \mid x) = \frac{p(x_{cont} \mid w_j, x_{disc}).P(x_{disc} \mid w_j).P(w_j)}{\sum_i p(x_{cont} \mid w_i, x_{disc}).P(x_{disc} \mid w_i).P(w_i)} \qquad (2)$$

where:

- x_{cont} corresponds to all continuous variables
- x_{disc} corresponds to the discrete variable
- $P(w_j)$ corresponds to the prior probability for class w_j
- $P(x_{disc}|w_j)$ corresponds to the conditional probability mass function of the discrete variable, given class w_j
- $p(x_{cont}|w_j,x_{disc})$ corresponds to the probability density function of the continuous variables, given class w_j and a value of the discrete variable x_{disc}

As stated earlier, a pure-pixel or "hard" classification approach such as this one was considered to be appropriate for about 70% of the area of study. However, this algorithm can be expected to face serious difficulties when confronting pixels covering more than a single LULC class (*i.e.*, "mixed" pixels). In this context, it was reasoned that, since the posterior probability given by the ML algorithm is a good measure of the confidence of the classification, this measure could be used in order to separate easily classifiable, well-behaved (probably pure) pixels from the more problematic ones. As such, it was decided that pixels with low posterior probabilities (*e.g.*, under 95%) should not be assigned the class selected by the ML algorithm. Instead, a *proportion estimation* was carried out for each pixel left unclassified, considering only those classes found within a local window of arbitrary size (this way, the procedure would also be able to take advantage of the well-known spatial correlation properties of remotely sensed data). In order to do this, characteristic red and near-infrared reflectance profiles for each class were estimated using pure pixels from the training data, and then a *linear mixture model* (Holben and Shimabukuro, 1993; Bouzidi *et al.*, 2000) was used in order to estimate the proportions of each class within each MODIS pixel to be re-evaluated. This model considers that the final observed reflectance R_i of a given pixel i is the result of a linear combination of the reflectances R_j of each class j (learned before) and their proportion ρ_{ij} within the pixel. As such, class proportions within each pixel were estimated by computing the set of values that minimized the distance between the observed and modeled reflectances for each date and for each band. At last, each re-evaluated pixel was assigned the class with the highest estimated proportion, thus completing the classification procedure.

5. RESULTS AND DISCUSSION

5.1 Classification Results

For validation purposes, the methodology described above was applied to the northern part of the area of study, based on the training classification data with 7 classes computed in section 4.2. Then, the results obtained were compared to the original classification information (section 2) for the northern part, which would thus serve as a "ground truth" for evaluating the performance of the algorithm. In this context, it should be noted that none of the sub-classes identified during the clustering process (section 4.2) could be evaluated quantitatively, since there was no corresponding ground truth in the reference data. Therefore, a final set of 4 classes was finally used for performance assessment of the methodology, namely: *agriculture, savannah, pasture* and *urban areas*. The results obtained are displayed in the confusion matrix below (Table 1):

	Reference Data					
	Agriculture	**Savannah**	**Pasture**	**Urban**	**Total**	**Usr. Acc.**
Results **Agriculture**	10,239	674	980	2	11,895	86.08%
Savannah	305	87,811	18,372	10	106,498	82.45%
Pasture	2,756	13,149	105,523	163	121,591	86.79%
Urban	0	3	44	12	59	20.34%
Total	13,300	101,637	124,919	187	240,043	
Prod. Acc.	76.98%	86.40%	84.47%	6.42%		

<u>Overall Accuracy</u>: 84.81% <u>Kappa Coefficient</u>: 0.7217

Table 1. Confusion matrix and accuracies for classification in the validation area (northern part of the basin)

These results show that the performance of the classification methodology was in general quite good, reaching an overall accuracy of 85%. Indeed, it could be seen that the algorithm was capable of generalizing very well for the *agriculture, savannah* and *pasture* classes. In addition to that, it is also important to note that some of the discrepancies observed between the computed results and the reference data can in fact be associated with differences between the *land use* information in the reference data and the actual *land cover* verified during the period of analysis. As such, it was particularly relevant to observe from the temporal profiles that significant areas labelled as *agriculture* actually showed no sign of crop activity during this particular year, suggesting that these areas were indeed assigned for pasture use during this period (Figure 4).

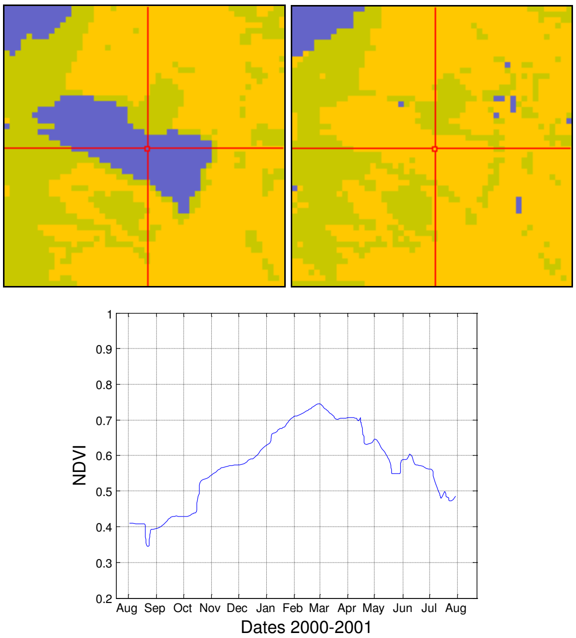

Figure 4. Differences between the reference data (top left) and the classification result (top right) in certain agricultural areas, due to lack of crop activity during the period of analysis (as verified by the observed temporal profile, typical of pasture use). Orange stands for pasture, green for savannah, and purple for agriculture.

5.2 Deforestation Assessment

Since there was no ground truth information available for evaluating deforestation, it was necessary to develop a methodology in order to assess the accuracy of this detection, which was given as one of the sub-classes identified in section 4.2. As such, it must first be recalled that the available reference data referred to July 2000 (section 2), which corresponded to the beginning of the temporal profiles evaluated with MODIS data. Thus, it became reasonable to admit that areas identified as deforestation were supposed to be labelled as *savannah* in the reference data, otherwise one could safely considerer such an identification to be incorrect. In addition to that, the areas that did correspond to *savannah* were further analyzed by inspecting LANDSAT images from before and after the period of analysis, in order to visually verify the disappearance of a forested area (see Figure 5 below).

In this manner, the results for the classification procedure showed that the identification of deforested areas was indeed quite good. First of all, it was observed that 75% of the areas identified did correspond to pixels labelled as *savannah* by the reference data. In fact, the remaining 25% corresponded almost entirely to areas of pasture use, which reflects the fact that pasture fields left "resting" (*i.e.*, unused in order to recover) often resemble areas of open savannah. Thus, when the landowner finally cleans the field, the apparent effect may then be confused with that of a deforestation phenomenon. In any case, for the 75% correctly identified, it could be readily verified from the LANDSAT images that all the regions detected did correspond to real deforested areas. Additionally, it was also possible to infer the actual date of the deforestation from the temporal profiles, as seen in Figure 5.

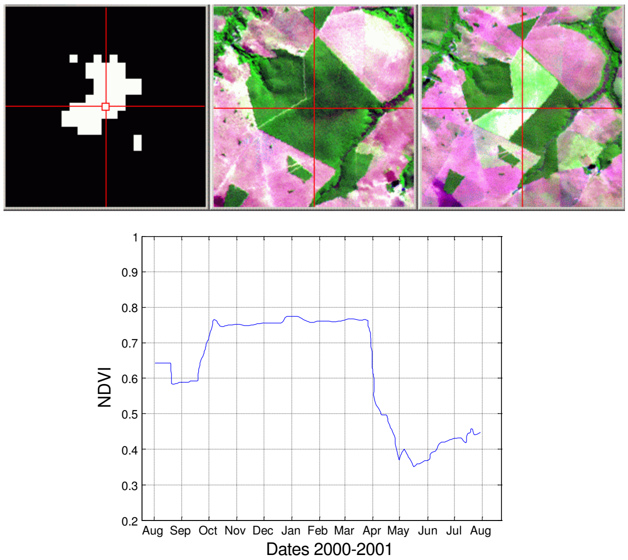

Figure 5. Assessment of deforestation detection. Looking at LANDSAT images, it could be seen that pixels classified as deforestation (top left) corresponded to areas that were covered by forest before (top center) but that were observed to be altered afterwards (top right). The very date of the deforestation event can be inferred from the corresponding temporal profile (beginning of April in this case).

As such, it can be considered that, with prior knowledge of the regions originally covered by forests, it is perfectly viable to use the methodology proposed here in order to reliably detect the occurrence of deforestation. Finally, it was also possible to compute an estimate of the total area of deforestation for the entire basin during the period of analysis, which was seen to correspond to 102.37 km^2. It should be stressed, though, that omission errors have not been considered in this assessment, and also that deforestation area estimates based on MODIS data have already been shown to underestimate the actual values, given that small deforestation activities cannot be reliably detected (Morton, 2005). Therefore, this estimate should be considered as a conservative value for the actual deforestation verified in the region during the period of analysis.

6. CONCLUSIONS

First of all, it can be stated that the results of this work confirm that MODIS 250m red and near-infrared surface reflectance data are indeed appropriate for performing regional scale LULC assessments, due to its moderate spatial resolution and excellent overall quality. In addition to that, it could also be observed that the methodology presented in this paper was capable of generalizing quite well over the entire High Taquari Basin, an area of 28,046 km^2, accurately identifying areas of agriculture, pasture and savannah, and providing a very high level of automation. Moreover, it was seen that the assessment of different crop types was also possible, which was used to facilitate the classification of agriculture areas. Finally, it was also shown that the method was capable of successfully identifying large areas of deforestation, and it may thus serve as the basis of an alert system for environmental monitoring applications.

Regarding future work, subsequent research will focus in the incorporation of pluri-annual data into the methodology, so as to extend the one-year period of analysis investigated in this paper. As such, the consideration of 2 to 4 years of data should provide a much more comprehensive view of land use in the region, thus helping the algorithm to differentiate between natural vegetation and unused pasture, as well as significantly improving its capability to understand the land use dynamics associated with agricultural practices. In addition to that, the possibilities of long-term monitoring based almost exclusively on MODIS data should also be investigated in the future. As such, it is expected that it will become possible to monitor the High Taquari Basin and its adjacencies in an affordable and effective way, thus enabling researchers to more easily assess land use and land use change in the area over time, particularly in regard of the conversion of natural vegetation to agriculture and pasture use.

REFERENCES

Bouzidi, S.; Lahoche, F.; Herlin, I., 2000. Land use classification at meso-scale using remotely sensed data. In: *Proceedings of 19th ISPRS Congress and Exhibition, geoinformation for all*. Amsterdam, Netherlands.

Bueno, M. do C. D.; Meirelles, M. S. P.; Coutinho, H. L. da C., 2003. Utilização de redes de dependência para a avaliação do impacto erosivo da Bacia do Alto Taquari. In: *IV Congresso Brasileiro da Sociedade Brasileira de Informática Aplicada à Agropecuária e à Agroindústria*, vol. 1, pp. 233-238. Porto Seguro, Brazil.

Godoy, J. M.; Padovani, C. R.; Pereira, J. C. A; Vieira, L. M.; Carvalho, Z. L.; Galdino, S., 1999. Evaluation of the river Taquari, Pantanal, Brazil, silting up based on the 210Pb geochronology of floodplain lakes sediments. *3rd International Symposium Environmental Geochemistry In Tropical Countries*. Nova Friburgo, Brazil.

Hall, M. A., 1999. *Correlation-based feature subset selection for machine learning*. PhD thesis, Department of Computer Science, University of Waikato, New Zealand.

Holben, B. N.; Shimabukuro, Y. E., 1993. Linear mixing applied to coarse spatial resolution data from multispectral satellite sensors. *International Journal of Remote Sensing*, 14(11), pp. 2231-2240.

Lillesand, T. M.; Kiefer, R. W., 2000. *Remote sensing and image interpretation*, 4th ed. Wiley and Sons, New York, NY, USA.

Lobell, D. B.; Asner, G. P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93(3), pp. 412-422.

Meirelles, M. S. P.; Costa, G. A. O. P. da; Singh, D.; Berroir, J.-P.; Herlin, I.; Silva, E. F. da; Coutinho, H. L. da C, 2004. A methodology to support the analysis of environmental degradation using NOAA/AVHRR data. In: *Proceedings of 20th ISPRS Congress*, pp. 534-540. Istanbul, Turkey.

Meyer, W. B.; Turner, B. L. II, 1994. *Changes in land use and land cover: a global perspective*. Cambridge: Cambridge University Press. 549 p.

Morton, D. C., Espírito-Santo, F. del B., Shimabukuro, Y. E., DeFries, R. S., Anderson, L. O., 2005. Validation of MODIS annual deforestation monitoring with CBERS, Landsat, and field data. In: *Proceedings of XIIth Brazilian Remote Sensing Symposium (SBSR)*, April 16-21, pp. 3159-3166, Goiânia, GO, Brazil.

Sellers, P. J.; Berry, J. A.; Collatz, G. J.; Field, C. B., Hall, F. G., 1992. Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sensing of Environment*, 42(3), pp. 187-216.

Silva, J. dos S. V. da, 2003. *Análise multivariada em zoneamento para planejamento ambiental - Estudo de caso: bacia hidrográfica do alto rio Taquari, MS/MT*. PhD Thesis, Faculdade de Engenharia Agricola, Universidade Estadual de Campinas, Campinas, Brazil.

Strahler, A. H.; Muchoney, D.; Borak, J. S.; Friedl, M.; Gopal, S.; Lambin, E.; Moody, A, 1999. *MODIS land cover and land cover change*. Algorithm Theoretical Basis Document (ATBD), v. 5.0, GSFC, NASA, USA.

Wessels, K. J.; DeFries, R. S.; Dempewolf, J.; Anderson, L.O.; Hansen, A.J.; Powell, S.L.; Moran, E.F, 2004. Mapping regional land cover with MODIS data for biological conservation: examples from the Greater Yellowstone Ecosystem, USA and Pará State, Brazil. *Remote Sensing of Environment*, 92(1), pp. 67-83.

ACKNOWLEDGEMENTS

This work was supported by a MCT/CAPES M.Sc. scholarship, and by the Programme Alban (European Union Programme of High Level Scholarships for Latin America, scholarship no. EO4M045790BR). It is also a part of the ECOAIR project, sponsored by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and INRIA (Institut National de Recherche en Informatique et en Automatique, France).