

**Considerations for regional scale long-term monitoring of land use with MODIS data
and its application for the high Taquari basin, MS/MT, Brazil.**

Milton Jonathan
Margareth Simões Penello Meirelles
Jean-Paul Berroir
Isabelle Herlin

Embrapa Solos - CNPS
R. Jardim Botânico, 1024
22460-000 - Rio de Janeiro - RJ, Brasil
mjonathan@mail.com, margaret@cnps.embrapa.br

Institut National de Recherche en Informatique et en Automatique - INRIA
BP105 - 78153 - Le Chesnay - France
{jean-paul.berroir, isabelle.herlin}@inria.fr

Resumo. Este artigo discute uma metodologia para a classificação sistemática do uso e cobertura do solo em escala regional e para períodos extensos de tempo. O cerne desta metodologia está baseado no trabalho descrito em Jonathan et al. (2005, 2006) e se fundamenta na análise de dados multitemporais provenientes do sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*), o qual se encontra a bordo dos satélites Terra e Aqua da NASA e apresenta propriedades de alta frequência temporal, extensa cobertura, e baixíssimo custo para aquisição de dados. Neste trabalho, realizou-se uma avaliação dos maiores potenciais e possíveis impedimentos da aplicação desta metodologia para o monitoramento plurianual do uso e cobertura do solo, concluindo-se ser de fato viável monitorar grandes regiões por longos períodos de tempo com baixo custo e altos níveis de automação. Neste sentido, tomou-se como área de estudos a Bacia do Alto Taquari, localizada nos estados de Mato Grosso do Sul e Mato Grosso, de forma a se enfatizar a relevância da aplicação desta abordagem para a detecção e quantificação de fenômenos como o desflorestamento e a dinâmica agrícola, que são por sua vez essenciais para o melhor entendimento dos principais processos de degradação relacionados ao ecossistema do Pantanal.

Palavras-chave: sensoriamento remoto, processamento de imagens multitemporais, classificação do uso do solo, detecção de mudanças do uso do solo, MODIS.

Abstract. This paper discusses a methodology for the systematic classification of land use/land cover on a regional scale and over a period of several years. The basis of this methodology has been presented earlier by Jonathan et al. (2005, 2006) and is centered on multitemporal analyses of 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. In this article, special emphasis is given to the evaluation of the possibilities and potential hindrances regarding the use of such a methodology for the pluriannual monitoring of land use and land cover. It is concluded that the monitoring of large regions for an extended period of time with low cost and high levels of automation is indeed a viable proposition. In this sense, the High Taquari Basin, located in the states of Mato Grosso do Sul and Mato Grosso, Brazil, is used as a study area for the analyses, so as to point out the relevance of the approach for detecting and quantifying phenomena such as deforestation and agriculture dynamics, which are themselves essential for the better understanding of the degradation processes within the Pantanal ecosystem.

Key-words: remote sensing, multitemporal image processing, land use classification, land use change detection, MODIS.

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 to exploit 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, with great success (Strahler et al. (1999); Lobell and Asner (2004); Wessels et al. (2004)).

Regarding the Pantanal biome in particular, it can be stated that the Taquari river, as a very important tributary of the Paraguai river, plays an essential role within this ecosystem. Unfortunately, however, the river has been suffering a severe silting-up process due to uncontrolled land use expansion in the High Taquari Basin, in the Brazilian states of Mato Grosso do Sul and Mato Grosso (**Figure 1**), 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.

2. Objectives

This work aims to provide an overview of the methodology proposed by Jonathan et al. (2005, 2006) for supervised LULC classification on a regional scale, and then to discuss its applicability for long-term monitoring of large areas with low cost. As such, the main objective sought here is to contribute to the development of tools that can allow a rapid and accessible assessment of land use change processes over time. In this way, it is expected that this will facilitate the efficient monitoring of processes such as deforestation and vegetation dynamics, with significant benefits to the improvement of scientific knowledge and surveillance activities that are essential to the preservation of the Pantanal biome.

3. Materials and Methods

3.1 Study area

The study area used in this work corresponds 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°10'S, 53°10'W) and (19°45'S, 55°10'W), and comprehends a total area of 28,046 km².

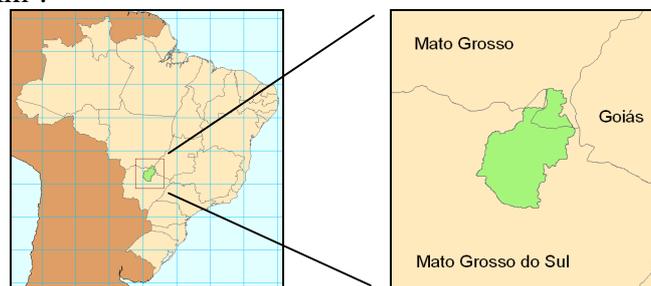


Figure 1. Location of the High Taquari Basin, the study area for this work, in mid-western Brazil

Aside from the reasons depicted in section 1, it should also be noted that this basin has already been the subject of numerous studies (e.g.: Godoy et al. (1999); Silva (2003)), and therefore it can be considered particularly suitable as a pilot area for research.

3.2 Previous work: overview of the basic methodology for regional scale classification

The basic methodology for regional scale classification consists of a supervised strategy, and is described in detail in Jonathan et al. (2005, 2006). In essence, this methodology is based on the following two sets of input data:

- MODIS data covering the area under study for an entire year, in such a way as to capture the entire vegetative cycle in an appropriate manner. This data should include NDVI values and red and near-infrared surface reflectance values for each point within the area.
- A reliable classification for at least part of the region under study, which should be valid for the same year from which MODIS data was gathered. This classification

should include areas that belong to every LULC class of interest, in such a way that each of these classes is well represented in the data. This way, the classification can serve as a reference to provide the supervised classification algorithm with a set of training data.

Once all data is gathered, an important step of the methodology consists of the processing of the original MODIS temporal signal observed at each point within the area of the study. This processing step is aimed at simplifying the raw and redundant temporal data gathered sequentially along the year, in such way that a more compact and meaningful description of the temporal behavior can be obtained for each point. For that matter, the MODIS NDVI temporal profiles are first modeled by fitting a smooth curve to the previously computed data, thus providing a reduction of dimensionality for the original data. Moreover, this fitting also helps to minimize the residual noise and to increase profile interpretability. This way, it becomes possible to use the resulting curves as a basis for computing several *curve features* that provide more intelligent and compact information about the profiles, thus helping the classification process (**Figure 2**).

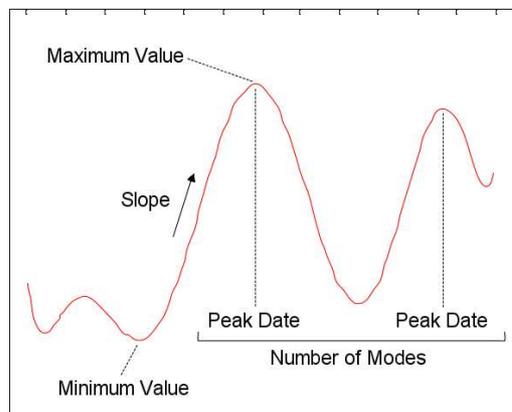


Figure 2. Examples of profile features computed for a MODIS NDVI temporal profile, which provide more meaningful and compact information about the temporal behavior observed at each point.

After that, the supervised classification process itself can be executed. For this purpose, MODIS data are analyzed at the areas belonging to the training set, and then the patterns learnt for each class are used by the classification algorithm for the remainder of the region under study. In this sense, the following collection of features are computed and used to describe the temporal behavior characteristic of each class: *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 original unfitted data values*, *date of the peak of the main mode*, *width (in days) of the main mode*.

3.3 Materials

In order to perform this work, 16-day Terra/MODIS atmospherically corrected surface reflectance data were acquired (product MOD13Q1, containing NDVI, red and near-infrared bands and corresponding metadata at 250m resolution). Images were obtained so as to cover the entire High Taquari Basin during a three-year period starting in August 2000 up to July 2003, where each August-to-July period corresponds to the local annual vegetative cycle.

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)). The classification legend was adapted to MODIS resolution as described in Jonathan et al. (2005, 2006), and thus could serve as a reference for the training of the classification algorithm during the 2000-2001 year. The adapted MODIS legend consisted of 3 distinct agriculture classes, savannah, pasture and deforestation.

3.6 Long-term Monitoring

Once a classification has been successfully achieved for the starting year (Aug 2000 – July 2001), it was then possible to start to analyze the possibilities for monitoring the dynamics of land use and land cover over subsequent years by use of MODIS data.

In that respect, it must be noted that it is particularly important to be able to monitor the region as cheaply and automatically as possible, so as to make this monitoring viable on an operational basis. As such, it should be pointed out that the most expensive and difficult process within the proposed methodology (and supervised classification approaches in general) corresponds to the acquisition and adaptation of appropriate training data. Indeed, training data for this methodology must be acquired by use of higher-resolution (and therefore more expensive) data such as that provided by LANDSAT, coupled with *in situ* observations provided from (even more expensive) field trips.

Therefore, in light of these considerations, it is proposed here that, once a classification for an initial year has been performed, the monitoring of subsequent years should not require *any* data other than those provided with little or no cost. Thus, it is suggested that detailed training data from LANDSAT or field trips should ideally only be necessary in order to provide a starting point for long-term monitoring, so that subsequent years could be evaluated *exclusively* by use of MODIS data alone. As such, it is expected that, after an initial effort and investment for the first classification of a base year, such an approach would make it possible to monitor large regions such as the High Taquari Basin for an extended period of time, with extremely low costs and in a highly automatized manner.

In this case, a simple methodology for long-term monitoring would consist of the following steps:

1. Use detailed training data from LANDSAT and/or field measurements in order to identify an appropriate number of samples for each LULC class, and then use these samples together with MODIS temporal data from the same year, in order to train a supervised classification algorithm as described in Jonathan et al. (2005, 2006). In this manner, the algorithm will learn the MODIS temporal patterns that are characteristic of each class, and these can then be readily used to classify the entire region under study for the referred year.
2. Once the MODIS temporal patterns typical of each class are known, these could then be used repeatedly for subsequent years, applying the same classification algorithm based on MODIS data from each following year.

Needless to say, in order for such a methodology to be successful, it is strictly necessary that the information provided during the initial training of the methodology be sufficient for the successful classification of subsequent years. This means that the temporal patterns that were found to be typical of each class during the first year *must* occur similarly during the following years, so that the classification algorithm can successfully identify the correct LULC class.

Unfortunately, however, inevitable changes in climate conditions, soil conditions, or general land use policies from one year to the next will surely result in pattern variations over time. As such, these variations may in turn present some level of difficulty for the correct identification of the LULC classes, depending on the particular patterns used for characterizing each class. Therefore, in order to assess the possibilities of applying this methodology for reliable long-term monitoring of land use and land cover, two issues must be firstly investigated:

1. Do MODIS temporal profiles in general present similar behaviors from one year to the next? More specifically, do the same sort of patterns arise from the data, and do overall values observed change substantially?
2. Which specific temporal patterns or features are not overly affected by inter-annual variability, so as to be considered “stable” from one year to the next? In particular, which features can be considered to be reliable “temporal signatures” of each class, so that they can be consistently used across the years for discriminating between the classes?

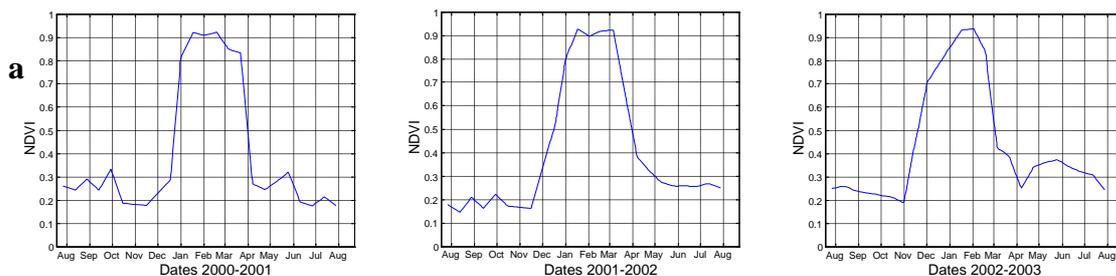
In order to study these questions, MODIS data from three consecutive years were examined, the results of which are presented in the following section.

4. Results and Discussion

Given the MODIS data for the period between August 2000 and July 2003, NDVI temporal profiles were generated for each point within the area of study, and their behavior over the three years was observed. In that respect, focus was given to areas whose LULC class was known to be unchanged throughout the period of observation.

As such, temporal profiles were observed for areas of three agricultural subtypes, as well as for areas of savannah and pasture. It should be noted that examples of deforestation were not investigated in this manner, given that deforested areas cannot be deforested again over subsequent years.

First of all, areas of agriculture were investigated, as shown in **Figure 3** below. Observing the NDVI temporal profiles generated, it was seen clearly that the general patterns characteristic of each agricultural practice remained evident throughout the period of observation. More specifically, for agricultural type **a** (soy), the single central mode observed between the months of November and April revealed itself clearly for all three years. Likewise, agricultural type **b** (soy followed by winter corn) displayed both the typical soy mode and the posterior mode for corn, located between the months of April and August. Finally, agricultural type **c** (sorghum) maintained its characteristic larger mode, with greening up around January, and senescence occurring during the month of June.



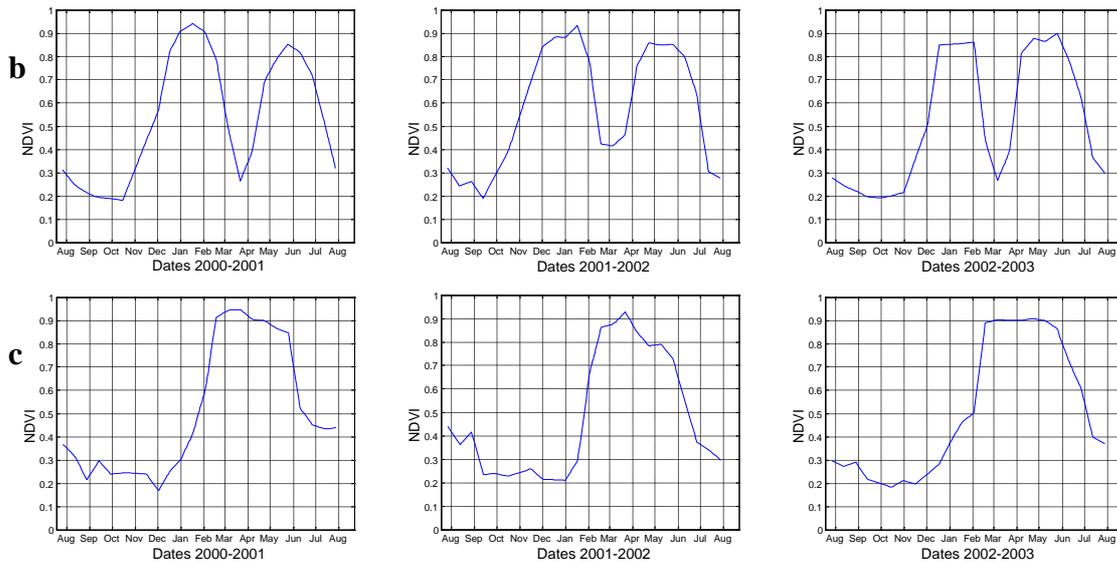


Figure 3. NDVI temporal profiles built from MODIS data for areas of three distinct agricultural practices, namely: type **a**, soy (first line); type **b**, soy + winter corn (second line); type **c**, sorghum (third line). The temporal profiles generated cover the annual periods 2000-2001 (first column), 2001-2002 (second column) and 2002-2003 (third column). It can be readily seen that the major temporal patterns remained clearly visible for all three agricultural types throughout the period of analysis.

In a similar way, NDVI temporal profiles were then generated for areas that remained continuously covered by the LULC classes *savannah* and *pasture* during the three years analyzed. Indeed, these profiles, shown in **Figure 4** below, show once again that the varying inter-annual conditions did not significantly affect the spectro-temporal response provided by each class. Nevertheless, it is important to note that the temporal profiles for these classes do not differ a lot in terms of shape, meaning that in this case class discrimination must rely almost exclusively on features related to overall absolute NDVI values. Furthermore, it should also be observed that, for these classes, the differences in overall NDVI are sometimes not very large, so that one can anticipate a certain difficulty for the proper classification of areas of pasture and savannah.

In this sense, it has been shown before that, with proper training, this information is indeed sufficient for adequate discrimination between the two classes (Jonathan et al. (2005, 2006)). However, when one considers the applicability of this training for subsequent years, it is necessary to point out that, in this case, even small variations of the absolute NDVI responses over the years will have significant effects on the correct identification of these classes. Thus, it can be reasoned that additional information would be of great importance if reliable long-term discrimination between *pasture* and *savannah* is to be achieved, a subject that is further discussed in section 5.

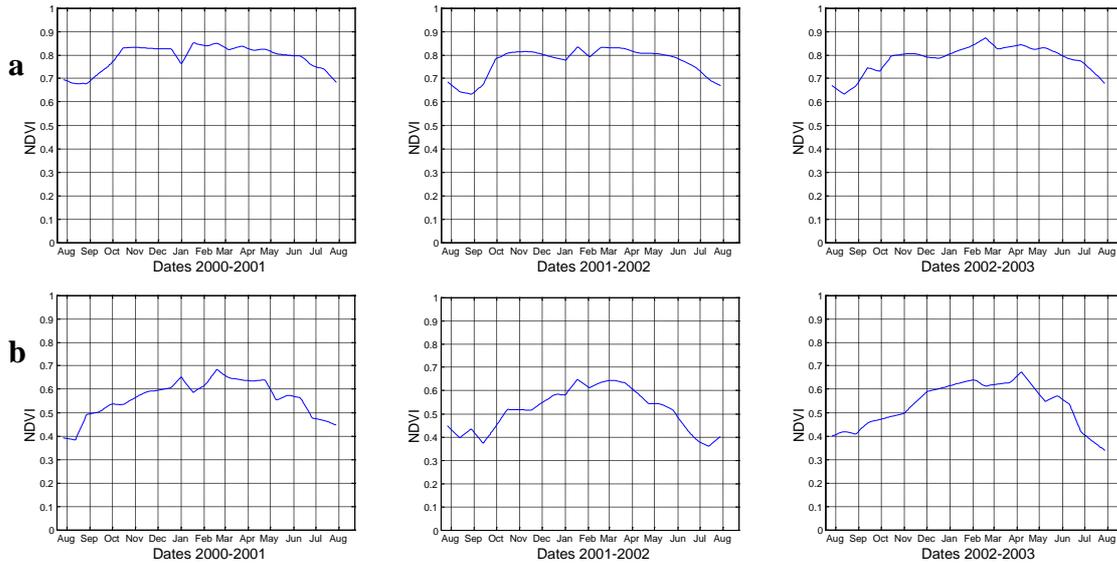


Figure 4 NDVI temporal profiles for areas of LULC classes *savannah* (a) and *pasture* (b), for years 2000-2001 (first column), 2001-2002 (second column) and 2002-2003 (third column). Overall response patterns can be seen to remain stable over the years. However, since the two class profiles differ very little in shape, class discrimination must rely almost exclusively on occasionally subtle differences in absolute NDVI values.

Following these observations, another analysis was performed in order to investigate the second issue pointed out at the end of section 3.6: which temporal features computed by the methodology can be considered to be more stable or reliable over subsequent years, to the point that they can be adopted as “temporal signatures” of the LULC classes?

In order to approach this question in a very general way, the responses observed for the *entire basin* for each year were taken into consideration. As such, the overall mean and standard deviation were computed for each feature and each year, so that it could be quantified how much the responses as a whole deviated from the base year to the subsequent ones.

In practice, this quantification of deviation was performed in the following manner: assuming a gaussian or normal distribution for each feature observed over the entire basin, the total *overlap* was computed between the distributions of distinct years, so that a 100% overlap would correspond to no deviation at all. This way, it was expected that certain features would display good robustness to inter-annual variability, with very little observed deviation (and hence a high degree of distribution overlap over the years), whereas other features might exhibit more sensitivity to these varying conditions and consequently less overlapping distributions. It should be noted that the feature “*number of modes*” was not considered in this analysis, since it is discrete in nature and therefore does not yield a continuous distribution.

Table 1 Overlap computed between each temporal feature’s distributions over the three years analyzed. Features with higher overlap values correspond to the ones less sensitive to inter-annual variability, and hence more appropriate for long-term monitoring applications.

Feature	Distribution Overlap	
	2000/2001 – 2001/2002	2000/2001 – 2002/2003
mean	91.19%	95.65%
maxDate	93.82%	85.75%
begin	75.77%	80.44%

deltaBeginEnd	52.30%	55.72%
delta150days	83.13%	86.11%
slopeAbsMean	81.91%	79.24%
min	93.67%	90.99%
deltaMaxMin	96.00%	88.22%
valueStddev	87.40%	85.77%
mode1DatePeak	91.84%	86.71%
mode1DateWidthPeak	95.90%	86.07%

Indeed, by observing the results presented in **Table 1**, it can be noted that the feature *deltaBeginEnd* displayed the greatest variability over the years, followed by the *begin* feature. This phenomenon can actually be noticed even by visual inspection of the temporal profiles displayed in **Figure 3** and **Figure 4**, where it becomes evident that the NDVI values at the beginning and end of the temporal profile are very dependent on the actual date of the beginning and end of the vegetative cycle. As such, given that these dates may vary significantly due to changing inter-annual conditions such as climate, it becomes clear that the arbitrary cut used to divide the overall temporal profile into yearly profiles is not entirely adequate. Therefore, this suggests that these two features should only be used if this cut is to be adjusted in an appropriate manner (i.e., by moving it to the effective beginning of the vegetative cycle of each year).

5. Conclusions and Suggestions

In this paper, a general investigation of the behavior of pluriannual NDVI temporal profiles from MODIS data was performed, with the purpose of analyzing the applicability of the methodology described in Jonathan et al. (2005, 2006) for long-term monitoring. Several classes of land use and land cover were taken into consideration, and it was observed that in general terms the profiles displayed a very satisfying degree of stability, meaning that they are not significantly affected by changing conditions from one year to the next.

Furthermore, an analysis was carried out in order to evaluate, in practice, if the different temporal features employed successfully in Jonathan et al. (2005, 2006) could also be considered appropriate for the identification of these LULC classes throughout the years. In this sense, it was observed that, while the majority of the features did remain relatively stable throughout the years, some of them demonstrated a higher sensitivity to variations from one year to the next, especially those related to the conditions at the beginning and end of the temporal profile. Therefore, it can be concluded that, while some care is needed, it is indeed possible to define a set of features that are in principle appropriate for actually capturing “temporal signatures” of each class, in such a way that they could be effectively applied to data of any given year.

Finally, it must also be noted that, in certain cases, the discrimination between specific classes can be more sensitive to inter-annual variations and thus more complicated in terms of pluriannual classification. More specifically, it was suggested that, since the discrimination between the classes *pasture* and *savannah* is overly dependent on subtle differences in NDVI responses, this differentiation might not be viable on the long term without the incorporation of additional information to the methodology. For that matter, one possibility that will be investigated in the future refers to the consideration of the history of land use for each area over the years. Thus, for each point, the actual evolution of LULC classes over time could be analyzed in conjunction with a state-transition diagram (STD), in order to guide the

methodology by inferring the most probable outcome for future land use. In this sense, it would become possible to state, for instance, that areas previously covered by pasture should not be confused with native savannah, whereas areas covered with savannah would not be able to become pasture directly, without the previous observance of a deforestation process.

As a final note, given all these observations, it can be concluded that long-term monitoring based on MODIS data alone is indeed a viable proposition, and can have a very positive impact on current capabilities for the assessment of land use change of larger regions over the years. The current methodologies will be further investigated and perfected in the future, so that important insights into the land use dynamics of priority areas such as the High Taquari Basin can be gained routinely and with low cost.

9. References Bibliográficas

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

Godoy, J. M.; Padovani, C. R.; Pereira, J. C. A.; Vieira, L. M.; Carvalho, Z. L.; Galdino, S. 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, 1999.

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

Jonathan, M. **Classificação do uso e cobertura do solo em escala regional a partir de seqüências temporais de dados MODIS**. 2005. 147 p. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Computação – Área de Concentração Geomática, Universidade do Estado do Rio de Janeiro, Brasil. 2005. Disponível em <http://www.geomatica.eng.uerj.br/teses/miltonj.tar.gz>. Acesso em: 25 set. 2006

Jonathan, M.; Meirelles, M. S. P.; Berroir, J.-P.; Herlin, I., 2006. Regional scale land use/land cover classification using temporal series of MODIS data. In: **Proceedings of ISPRS Mid-term Symposium 2006** “Remote Sensing: From Pixels to Processes”, Enschede, Netherlands.

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

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. A methodology to support the analysis of environmental degradation using NOAA/AVHRR data. In: **Proceedings of 20th ISPRS Congress**, p. 534-540. Istanbul, Turkey, 2004.

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

Silva, J. dos S. V. da. **Análise multivariada em zoneamento para planejamento ambiental - Estudo de caso: bacia hidrográfica do alto rio Taquari, MS/MT**. 2003. 307 p. Tese de Doutorado, Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, Brasil. 2003.

Strahler, A. H.; Muchoney, D.; Borak, J. S.; Friedl, M.; Gopal, S.; Lambin, E.; Moody, A. **MODIS land cover and land cover change**. 1999. 66 p. Algorithm Theoretical Basis Document (ATBD), v. 5.0, GSFC, NASA, USA. 1999. Disponível em http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf. Acesso em: 25 set. 2006

Wessels, K. J.; DeFries, R. S.; Dempewolf, J.; Anderson, L.O.; Hansen, A.J.; Powell, S.L.; Moran, E.F. 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**, v. 92, n. 1, p. 67-83, 2004.