ABSTRACT

Time series of MODIS vegetation indices are widely used to map vegetation. However, some noise can affect the temporal profiles. Thus, many techniques have been developed to smooth them. Four algorithms are applied on crop pixels in the Brazilian Amazonian State of Mato Grosso. Comparisons led to the selection of the Weighted Least Squares (WLS) algorithm and the Savitzky-Golay (SG) filter. Those techniques were computed on MODIS data in order to detect six crop classes. Tests of separability show that the smoothed data improved the potential of separability at each MODIS sub-period. Moreover, supervised classifications were then realized. The WLS data refined efficiently the classification result when using C4.5 decision tree. When using the Maximum Likelihood and Spectral Angle Mapper classifiers, the smoothed data did not improve the classification results as compared with those obtained through original MODIS data. However, it required fewer input MODIS images to reach good results. The SG filter led to better results than the WLS algorithm when using those classifiers.

Index Terms— Image classification, Smoothing methods, Vegetation mapping, MODIS, Mato Grosso

1. INTRODUCTION

Crop monitoring and more generally land use change detection are of primary importance in order to analyze agricultural spatio-temporal dynamics and its environmental impacts. This aspect is especially true in such a region as the State of Mato Grosso (south of the Brazilian Amazon Basin) which hosts an intensive pioneer front. Deforestation in this region has often been explained by soybean and pasture expansion in the last three decades [1][2]. Remote sensing techniques may now represent an efficient and objective manner to quantify how crops expansion represents a factor of deforestation through crop mapping studies.

The Moderate Resolution Imaging Spectroradiometer (MODIS) data with a near daily temporal resolution and 250 m spatial resolution has proven to be an adequate resource to crop mapping [3][4]. Especially, multitemporal vegetation indices (VI) studies have been currently used to realize this task [5][6][7]. In this study, 16-days compositions of EVI (MOD13Q1 product) data are used. However, although these data are already processed by a Constrained View-Maximum Value Composition (CV-MVC) algorithm [8], multitemporal VI profiles still remain noisy due to cloudiness (which is extremely frequent in a tropical region such as the southern region of the Amazon Basin), sensor problems, errors in atmospheric corrections or BRDF effect [8][9][10][11][12]. Thus, many authors developed algorithms that smooth the multitemporal VI profiles in order to improve further classification.

This work aims at comparing and testing different smoothing algorithms in order to select the one which satisfies better to the demand, which is classifying crop classes.

2. STUDY AREA AND USED DATA

The study area is localized in the Brazilian Amazonian State of Mato Grosso, which area represents 906 000 km². This state has known the highest deforestation rates in Amazonia for more than 20 years. 104 000 km² has been deforested between 1992 and 2005 [13]. Mechanized agriculture, and more especially soybean, is considered to be the main driver that led to this high deforestation rates. Indeed, during the same period, soybean cultivated area grew from 1.5 to 6 millions hectares. Others crops as corn, cotton, rice or sugarcane also expanded intensively during those years.

In order to collect ground truth data for the supervised classification, an intensive field trip has been realized. 93 424 ha and 151 627 ha were mapped for the 2005-2006 and 2006-2007 harvests respectively. The data were obtained at a field scale (field’s mean size = 176 ha) through interviews. Six classes were considered including single-crop (soybean + millet, soybean + sorghum, soybean + corn and soybean + cotton) as well as double-crop production practices (soybean + millet, soybean + sorghum, soybean + corn and soybean + cotton).
MODIS/TERRA images (MOD13Q1 product) were acquired from July 2005 to July 2007. Thus, annual temporal profiles of vegetation indices were extracted. The profiles are composed of 23 images, which spatial resolution is 250 m and temporal resolution is 16 days. The Enhanced Vegetation Index (EVI) has been chosen to study crops phenological cycles due to its ability to better avoid atmosphere and soil disturbances [8]. It is also more sensitive than NDVI in areas of high vegetation activity [8], such as Mato Grosso.

3. METHODOLOGY

MODIS EVI data can be affected by cloudiness, sensor problems or BRDF effect. A CV-MVC method is applied to eliminate part of this noise [8]. It consists in selecting maximum values of EVI on a fixed period (16 days in that case). However, noise still remains especially in tropical regions were cloudiness is very high during the rainy season. Thus, smoothing algorithms have been proposed to ameliorate the profiles’ quality. Most of those techniques are reviewed in [14] and [15].

In this study, four algorithms are compared: BISE (Best Index Slope Extraction) [9], MVI (Mean Value Iteration) [10], WLS (Weighted Least Squares) [11] algorithms and SG (Savitzky-Golay) filter [12]. Each of these techniques relies on parameters as thresholds or moving window size that must be fixed.

The algorithms are compared on five 2-years pixels to evaluate their smoothing quality. This evaluation is based on the analysis of 4 criteria proposed in [14]. These criteria are: “(i) high frequency noise should be diminished to the highest possible degree, (ii) overall intensity given by an area-specific EVI profile (low frequency feature of profile) has to be retained, (iii) high values should be retained to highest possible degree due to the low frequency of false highs and (iv) timing of profile features should not be influenced by the smoothing or to the lowest possible degree”. Notes (from 0 to 3) are assigned for each criterion in order to evaluate which technique(s) is (are) the best smoother. The chosen algorithm(s) is (are) then computed on MODIS EVI images for the two studied years. Tests of separability based on the ground truth data of the year 2005-2006 are realized based on the Jeffries-Matusita distance. It aims at showing if the algorithms managed in improving the potential of differentiation between the classes. Those tests are realized on the overall profile (comprising 23 MODIS images) as well as on each MODIS sub-period of the profile as proposed in [6].

This last test is a double interest process because it allows comparing the smoothing techniques and also enables to select a set of images which carries more information on the separability between the classes.

Supervised classifications are then applied. The training data are based on the ground truth data collected in 2005-2006 whereas the 2006-2007 data are used to control the classification quality. The classifiers tested are the Maximum Likelihood, the Spectral Angle Mapper and the C4.5 decision tree [16]. Various classifiers are tested in order to evaluate if the optimal smoothing method differs depending on the classification process. Moreover, smoothed data and original data are integrated in the classification processes to quantitatively estimate the improvement carried by the smoothing methods.

4. RESULTS

Parameters have been previously tested and assigned for each method. Moving windows for WLS, SG filter and BISE are 5, 5 and 3 MODIS periods respectively. The EVI threshold for MVI application

![Figure 1. Visual comparison of 4 smoothing algorithms on a 2-years pixel.](image-url)
is 0.1. The four smoothing algorithms are then compared on five 2-years pixels. An example is shown in figure 1.

Notes are then assigned to each method according to the criteria proposed by [14]. These notes vary from 0 (not satisfying) to 3 (satisfying) (table 1). The WLS algorithm and the SG filter led to better results. Thus, they were chosen to be computed on MODIS images in order to carry out further analysis.

<table>
<thead>
<tr>
<th></th>
<th>WLS</th>
<th>BISE</th>
<th>MVI</th>
<th>SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>High frequency noise</td>
<td>2.5</td>
<td>3</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>Overall intensity</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>High values retained</td>
<td>2.5</td>
<td>3</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>Timing of profile</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>10.5</td>
<td>9</td>
<td>9</td>
<td>10.5</td>
</tr>
</tbody>
</table>

Table 1. Notes assigned to the smoothing algorithms for each criteria.

Those selected algorithms are programmed and applied to the MODIS/TERRA EVI data (16-days composition periods). Tests of separability based on the Jeffries-Matusita (JM) distance and applied to the entire profile of the 2005-2006 harvest show that better separability index of the 6 classes is obtained by original and SG filter smoothed data (JM distance = 0.658). WLS smoothed data show lower result (JM distance = 0.636). However, the same test applied for each MODIS period showed that the smoothed data (either by WLS or SG filter) improved the potential of separability between classes (fig. 2).

The low JM distances calculated mean that it is difficult to separate the 6 defined classes. Indeed, due to similar agricultural calendars, only 3 classes are actually separable: cotton (single crop), soybean + cotton (double crop), soybean (single or double crop with corn, millet or sorghum). Therefore, the classification tests will be established based on these 6 classes and then regrouped to get 3 classes classifications. Moreover, the JM distances allow ranking the dates or images that carry more information on separability. Thus, various classification tests are made integrating images one by one according to the established rank. The classification tests are realized with 3 classifiers: Spectral Angle Mapper (SAM), Maximum Likelihood (ML) and C4.5 decision tree. The C4.5 decision tree algorithm already induces a variable selection process. Thus, this classifier was only tested by integrating all MODIS images.

For SAM and ML classifiers, results obtained by original data are as good as those obtained with SG filtered data when the 23 MODIS dates are integrated as input in the classification. WLS smoothed data gave lower classification results. However, the two smoothed data reach their maximum Kappa indexes values by integrating less images (MODIS dates) than original data. With the ML classifier, Kappa index is superior than 0.8 when integrating 8 MODIS dates whereas 13 MODIS date are necessary when using original data. The SAM classifier gives the best classification result and the Kappa Index of 0.8 is reached with 8, 9 and 11 MODIS dates by the WLS, SG filtered and original respectively.

The C4.5 decision tree gives the lower results (Kappa index = 0.73 with WLS smoothed data). However, the importance of smoothing appears more clearly. Indeed, the WLS smoothing algorithm managed increasing the Kappa index from 0.65 to 0.73. On the contrary, it is not explained why the SG filtered data gave lower results than the original data (Kappa index = 0.54).
5. CONCLUSION

MODIS Enhanced Vegetation Indices (EVI) data have been used to map crops in the Brazilian State of Mato Grosso through supervised classification. In order to improve the classification results, remaining noises were eliminated by computing smoothing algorithms. Over four tested methods, two were selected due to their ability to take into account 4 defined criteria proposed by [14]. Those methods are the Weighted Least Squares method and the Savitzky-Golay filter. Tests of separability applied on MODIS EVI temporal profiles for 6 classes show that the smoothed data improve the potential of separability for many MODIS sub-periods. The classification results confirmed this fact. Indeed, fewer MODIS images were necessary to efficiently map crops. Even if SG filtered data showed better results, original data also reached good results. It means that the noise contained in each individual original data was compensated when integrating more images in the classification process. Thus, a main advantage in smoothing EVI profiles is that it requires fewer MODIS data to get efficient results. Thus, it could reduce the necessary lag time to get agricultural statistics. The C4.5 decision tree classifier led to lower results and was much more dependent on the input data.

6. ACKNOWLEDGMENTS

This work is helped by the CNPq (through the ENVlAIR project linked to the CnPq-INRIA relations), the IDRC (International Reseurch Centre, Canada) and the IAI (Inter American Institute) for Global Change Research through project: “Land use change, biofuels and rural development in the La Plata Basin”, the ANR (Agence Nationale de la Recherche, France) through the DURAMAZ project : “Sustainable development in the Brazilian Amazon” and the European Union through the SENSOR-TTC project : “Land use change, biofuels and rural development in the La Plata Basin”.

7. REFERENCES


