Monitoring crop phenology in Mato Grosso (Brazil) using remote sensing data

Léna Maatoug [1]
Damien Arvor [2]
Margareth Simões [1,3]
Agnès Bégué [1]

[1] CIRAD - UMR TETIS
Maison de la Télédétection, 500 rue Jean-François Breton, 34093 Montpellier, France
{lena.maatoug,agnes.begue}@teledetection.fr

[2] IRD - UMR 228 ESPACE-DEV
Maison de la Télédétection, 500 rue Jean-François Breton, 34093 Montpellier, France
damien.arvor@ird.fr

[3] EMBRAPA - Programa LabEx Europa and Rio de Janeiro State University
PPGMA/DESC/UERJ
Agropolis International, 1000 Avenue Agropolis, 34394 Montpellier, France
margareth.simoes@embrapa.br

Abstract. In Mato Grosso (Brazil), agricultural practices are evolving rapidly and their monitoring (e.g. sowing dates or double cropping systems) becomes a relevant issue. The objective of this paper is to test the potential of MODIS satellite data to detect vegetation dynamics over agricultural land in Mato Grosso. We especially focus on the estimate of the sowing dates of soybean crops. First, the MODIS MCD12Q2 product, composed of phenology transition dates, was tested at the field scale, but turned out to be unusable due to a large number of missing data and inconsistencies. We then developed an alternative method based on the MOD13Q1 Enhanced Vegetation Indices (EVI) time series. We applied a 3x3 window Savitzky-Golay filter to the EVI time series, and extracted the 2006-2007 growing period. We then calculated the dates at which different EVI values were reached (from 0.1 to 0.9, step 0.1), and correlated these dates to a set of sowing dates observed in the fields over the same period. The best result was obtained with a EVI threshold value of 0.6, with a Pearson coefficient equals to 0.8.

Keywords: MODIS, time series, sowing date, agriculture, soybean, Mato Grosso.

1. Introduction

The state of Mato Grosso is the main producer of soybeans in Brazil. The agricultural model relies on highly industrialized and intensive practices, which have partly been incentivized by the Brazilian federal government. For example, it publishes recommendations aimed at helping farmers to determine the optimal sowing date and thus maximize the yields. As a consequence, agricultural practices are evolving rapidly and their monitoring (e.g. sowing dates or double cropping systems) becomes a relevant issue (Arvor et al., 2011). If we can accurately detect the dates of vegetation start of season, then we can estimate the sowing dates used by farmers, which in turn can be linked to crop yields data. Looking at these relationships over several years and locations could potentially lead to establishing more specific recommendations depending on the localization of the parcels.

Optical remote sensing data presents several advantages to monitor vegetation phenology over large areas (Curnel et al., 2006). In the case of MODIS sensors (on board of NASA Terra and Aqua satellites), images are freely released globally up to a daily basis at medium spatial resolution (250 m). The robustness of vegetation indices such as the Enhanced Vegetation Index (EVI) or the Normalized Vegetation Index (NDVI) to estimate photosynthetically active vegetation has long been tested and acknowledged (Huete at al., 2002). However in

tropical environment the use of optical remote sensing data is challenging and questionable because of the persistence of clouds and the high variability of Aerosol Optical Thickness. The MODIS sensors offer an interesting both spatial and temporal resolution to study vegetation dynamics that can help to overcome these issues (Ganguli et al., 2010).

The objective of this paper is to test the potential of MODIS data to detect vegetation dynamics over agricultural land in Mato Grosso. We especially focus on the estimate of the sowing dates of soybean crops.

2. Study area

Brazil is the world's second largest producer of soybeans after the United States. The soybean crop has faced a rapid expansion. Mato Grosso is now a farm State where dominant crops are soybeans, corn and cotton. For those different crops, there is a constant concern for understanding cultural practices, starting with agricultural calendars, which are partly determined by rainfall regimes. Particularly, in some regions, the rainy season extends over a period of 6 months (from September-October to May) allowing the development of double cropping systems, which combine usually soybean and corn or soybean and cotton. Soybean is always cultivated first, being planted between late September and mid-November. Soybean sowing date is established based on the first precipitation, which is characterized by a high spatial and temporal variability and can, consequently, sometimes condition yield losses. There is, therefore, a particular interest in following the soybean sowing date in order to improve the quality of the advice given to the producers.

3. Data

3.1. Ground data

In order to compare the vegetation dynamics observed by satellite, we used a dataset collected in the field between 2006 and 2007 (Arvor, 2009). This dataset contain a map in ArcGis format (shapefile), where agricultural parcels are identified. In total, 617 fields are mapped and filled with sowing dates, for the crop year 2005-2006, as well as 319 fields for the year 2006-2007. Additional attributes of those agricultural parcels are: field identification, name of the farmer, name of the owner, area, type of crop, planting date (for the first crop only), date of harvest (for the first culture only), soybean yields.

3.2. Satellite data

The satellite data used are those from the MODIS sensor. MODIS is a multispectral sensor that records the spectral response of surfaces in 36 spectral bands between 0.4 to 14.4 µm and a spatial resolution ranging from 250 m to 1000 m. MODIS data are available by NASA in open access, free of charge. For our study we used three different data sets:

- MCD12Q2 products: MCD12Q2 product contains data on the phenological vegetation transition dates (Ganguli et al., 2000). For each MODIS tile, there is a MCD12Q2 file per year with a spatial resolution of 500 m. The phenological transition dates represented in the file correspond to four phenological indices (Figure 1): 1) the beginning of vegetation growth (SOS - Start of Season), 2) the beginning of the maximum of vegetation (SMAX - Start of MAXimum greenness), 3) the end of the maximum of vegetation (EMAX - End of MAXimum greenness) and 4) the end of the cycle of vegetation (EOS – End Of Season).
- MCD43A4 products: this product contains the NBAR data (surface reflectance adjusted by the BRDF) used to determine the phenological indices MCD12Q2. The MCD43A4 data are distributed every 8 days at a spatial resolution of 500 m.

– MOD13Q1 and MYD13Q1 products: these products provide data on indices of vegetation NDVI and EVI with a resolution of 16 days and a spatial resolution of 250 m. For each pixel the vegetation indices were calculated from reflectance data after atmospheric and geometric corrections.

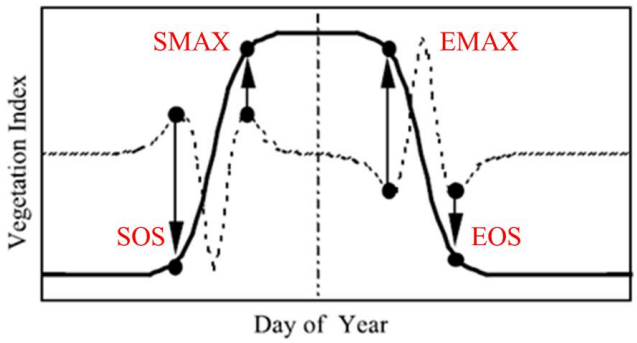

Figure 1. Schematic representation of the phenological transition dates calculated from the second derivative of logistics functions (based on Zhang et al., 2003).

4. Method

The main objective of our work is to assess the quality of MODIS MCD12Q2 phenological indicators, to estimate the soybean sowing dates on agricultural land in Mato Grosso, where we have field data. Our approach consisted of three main phases: analysis of the ground dataset in relation to the MODIS-derived phenology indices, and development of a sowing date detection technique.

4.1. Ground data pre-processing

In order to retrieve agricultural parcels time series, it was necessary, first, to create a mask of the pixel of interest. We have chosen to define the pixels of observation as being the gravity of study plots. We thus started from the status vector file containing digitized plots to create a new shapefile containing the gravity of the plots. We then rasterized the gravity to finally create the TIFF image pixels in the center of the plots, that we later used to extract the time series. We have chosen to extract the pixels in the center of the plots in order to limit the border effects that we have encountered in working on all of the pixels in a parcel. In fact, in passing from vector to raster, the delineation of plots becomes inaccurate, and we had degraded the accuracy of the information. In addition, the large pixel size (500 m) does not allow to remove pixels between two plots, especially in the case of parcels with narrow shapes.

4.2. Analysis of the phenological indices and comparison with ground data

In order to assess the consistency of the phenological indices presented in the MCD12Q2 product, we have initially performed a visual qualitative assessment of indices by superimposing them on EVI time series calculated from MCD43A4 products. Phenological indices and EVI values were extracted using algorithms developed in IDL. Our visual analysis was then based on a graphic observation (performed with R software) of the superimposition of the temporal profiles of EVI to the values of the phenological transition MCD12Q2 dates. We also conducted a statistical analysis of the MCD12Q2 data from an algorithm written in IDL, and that we have then formatted with the Excel software. Finally we studied the correlation between phenological variables of start of season and the sowing dates.

The exploration of the MCD12Q2 dataset was first made by analysis of the statistical distribution of the phenological transition dates. In order to assess the coherence of phenological indices of the MCD12Q2 product, we (i) conducted a visual qualitative assessment of indices by layering on EVI time series calculated from the MCD43A4 products, and (ii) studied the correlation between phenological variables of vegetation growth and end of season, and the dates of sowing and harvesting. The data processing was done with algorithms developed in IDL.

4.3. Development of an alternative method to map the phenological transition dates

For every pixel of interest representing the centre of gravity of the plots, we extracted the EVI temporal profile from the MOD13Q1 product. As this product has a spatial resolution of 250 m, we averaged the value of 4 pixels inside the pixel of interest whose resolution is 500 m. After that, we applied the filter Savitzky-Golay in order to smooth the temporal profile (Savitzky and Golay, 1964). From the smoothed profiles, we could estimate the start-up of the vegetation for each parcel dates. The methodology that we have chosen to extract these dates is to detect the date at which the smoothed curve of the EVI temporal profile reaches a threshold value. We tested this method for different threshold values by taking either (i) a fixed EVI value (from 0.1 to 0.9 with 0.1 step), or (ii) a value corresponding to a fraction of the filtered EVI range (from 0.1 to 0.9 with 0.1 step). We then analyzed the correlation between the dates obtained using those thresholds and the observed sowing dates at the field.

5. Results

5.1 Visual and statistical analysis of phenological indices MCD12Q2 and comparison with field data

The values of phenological indices of MCD12Q2 products were overlaid on temporal profiles of EVI calculated from NBAR MCD43A4 products. It turns out that this product presents a large number of missing data (Figure 2). Consequently, the indices provided by data MCD12Q2 are also presented in a disparate manner: for example on 1431 parcel only two cycles of vegetation are represented completely by four phenological season indicators : start (SOS) and end of season (EOS), beginning (SMAX) and end (EMAX) of maximum vegetation. Furthermore, the validity of some dates presented in phenological products MODIS can be questioned. MCD12Q2 phenological indices can be calculated only when the NBAR data themselves are available, however these data appear to be mostly missing.

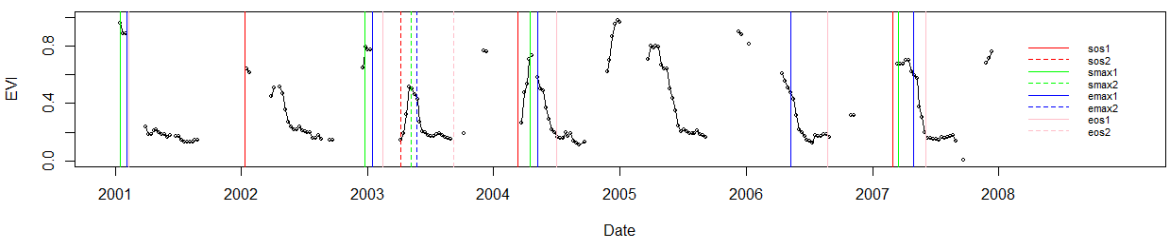

Figure 2. Example of 2001-2007 EVI time profile (MCD43A4) and the corresponding phenology transition dates (MCD12Q2), for field #1431. Vertical plain and dotted lines correspond to the first and second vegetation cycles, respectively.

A simple quantitative analysis of the status of 8 phenological indices from MCD12Q2 data confirmed the incompleteness of the data. For each phenological variable, we identified the percentage of pixels of observation that present data in one of the bands, in the two bands

or none of them (Figure 3). In addition to the fact that the image of 2005 contains no data, we find that the percentages of pixels containing no data varies between 20 and 60 (39 on average) and that some pixels do not contain data for the second cycle of vegetation, which is illogical since a first round of vegetation shall be identified before a second cycle is detected. In conclusion, the quality of the MCD12Q2 product is very heterogeneous, most vegetation patterns are not fully described by the four phenological transition dates. Anomalies in the MCD12Q2 product, due to vegetation indices data gaps were also observed on Mali data set (Vintrou et al. 2012), but in a smaller proportion making the product usable after a simple data selection, which is not the case in the present study

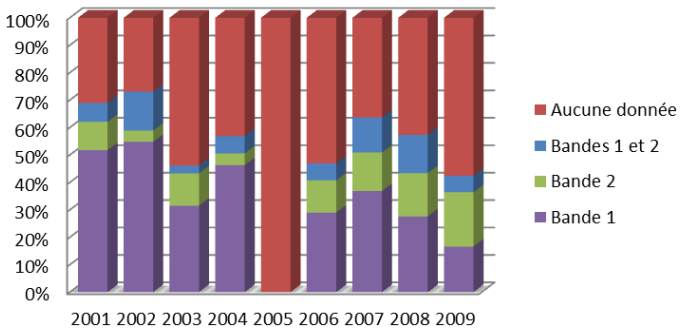

Figure 3. Statistical frequency distribution of the MCD12Q2 Start Of Season index. In red, the percentage of « no data » value.

4.2. Development of an alternative method to calculate the sowing dates

As MCD12Q2 and MCD43A4 products were not satisfactory for our study area, we decided to analyze the MOD13Q1 products to estimate the sowing dates. EVI temporal profiles from data MOD13Q1 production shows preliminary results (before smoothing) which appear to be satisfactory (Figure 4). MOD13Q1 EVI temporal profiles are more complete than those calculated from the NBAR MCD43A4 data available (Figure 2), and this,despite the smaller temporal window (8 days for MCD43A4 and 16 for MOD13Q1). Using the smoothed profiles, it becomes possible to clearly identify the presence of two cycles of vegetation in a crop year. On the other hand, the quality of phenological data MCD12Q2 is variable:

- some indicators seem to match the profiles of EVI (for example, year 2003- parcel 1431;) (Figure 4);
- most of the phenological indicators are inconsistent over MOD13Q1 profiles (e.g. for the first cycle of vegetation of 2001 on parcel 1396, or that of the year 2007 for the parcel 1431;) (Figure 4).

In order to take advantage of the EVI MOD13Q1 profiles we tried to extract phenological indicators by a simple method, focusing on the study of the sowing dates of the 2005-2006 season for which we have a large number of data to explore. We used the profile of EVI's first cycle of vegetation by extracting values between September 2005 and April 2006 (Figure 5), and calculated the correlation between the observed sowing dates and the EVI-profile dates obtained by EVI thresholding (Figure 5). Both absolute (from 0.1 to 0.9) and relative (of EVI amplitude) threshold values were tested. Only the results obtained with the absolute threshold values are presented here, as they give better results than the relative one (best R coefficient equals 0.8 for absolute values against 0.7 for relative values).

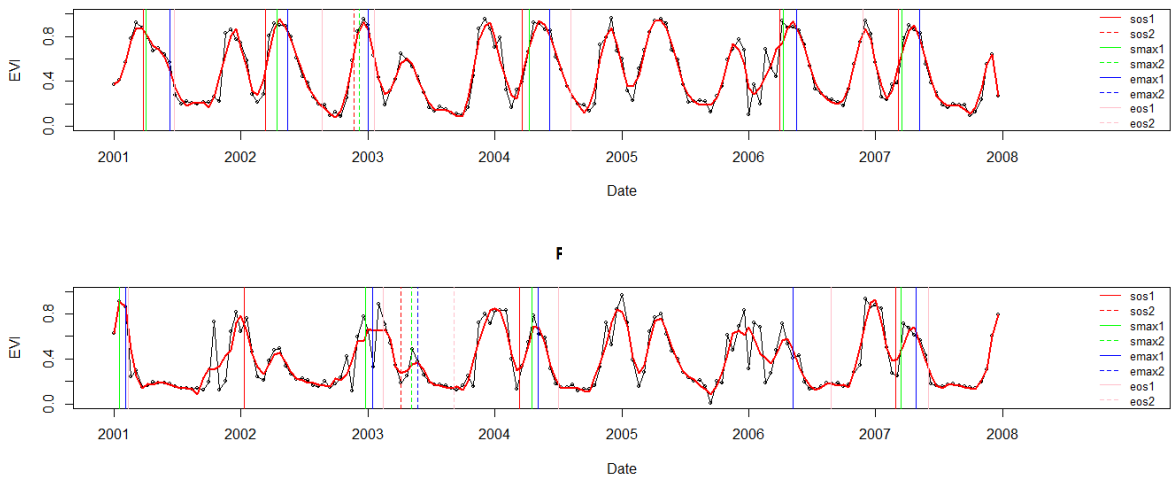

Figure 4. Example of EVI profiles for the 2001-2007 period, and for two plots. The black line represents MOD13Q1 EVI time profiles, the red line is the Savitzky-Golay filtered EVI profiles (3x3 window), and the vertical color lines correspond to the MCD12Q2 phenological dates. The fields #1396 (top figure) and #1431 (bottom figure) correspond to "soybean plus cotton" and "soybean plus corn" crops respectively.

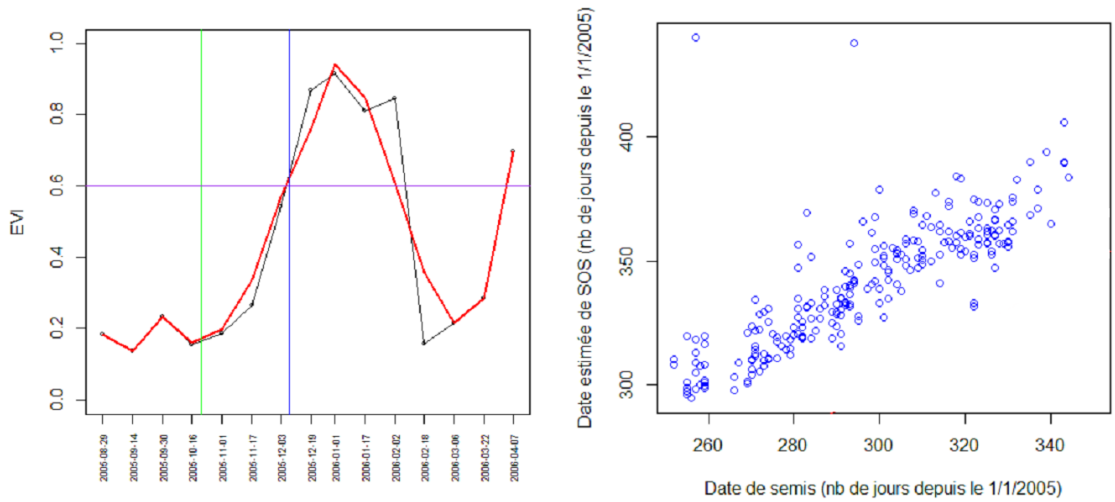

Figure 5. Example of sowing date data processing for a fixed EVI threshold of 0.6.

- Left figure: EVI 2005-2006 profile for field #1434. The black line represents MOD13Q1 EVI time profiles, the red line is the Savitzky-Golay filtered EVI time profile (3x3 window), the green vertical line corresponds to the observed sowing date, and the blue vertical line corresponds to the date at which the EVI threshold is reached.
- Right figure: Relation between the fields observed sowing dates and the 0.6 EVI threshold dates for the 2005-2006 data set (243 fields).

Figure 6 shows the Pearson correlation coefficients calculated for each threshold value. We observed a peak of correlation for threshold values between 0.5 and 0.7, with a maximum reached at 0.6 EVI value (R = 0.8). The low correlations obtained for thresholds between 0.1 and 0.4 (beginning of the growing cycle) is unexpected and can be explained by the fact that at the beginning of the crop cycle, vegetation indices irregularities occurred due to soil work or erratic vegetation regrowth.

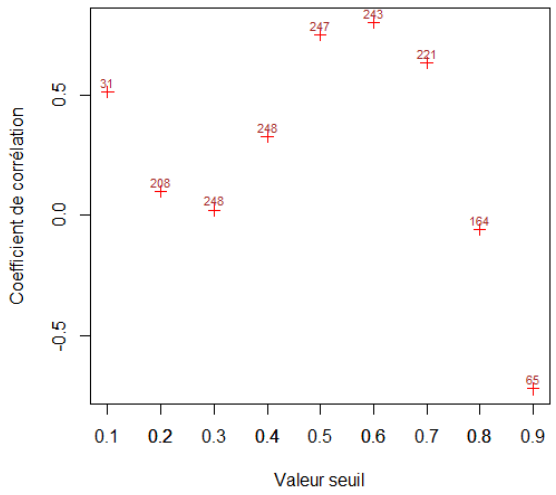

Figure 6. Pearson correlation coefficient calculated between observed sowing dates and EVI threshold dates, as a function of the EVI threshold values. The cropping year is 2005-2006 and the number of plots used in the correlation analysis is given as labels.

5. Discussion and conclusion

The acquisition and use of remote sensing data are subject to constraints, especially in tropical regions highly cloudy. The study of phenology of surface from satellite data can therefore lead to accurate results with data acquired with a great time repeatability. MODIS data therefore have high potential for analysis of phenology of tropical regions, since the sensor gets a global coverage of the Earth in one to two days, and corrected reflectance data are available. The MCD12Q2 data represent a valuable potential for the various applications of the study of surface phenology at regional and global scales. They would allow researchers on phenological or other disciplines saving a lot of time and energy. Indeed we have seen that phenological indices extracted from reflectance data or vegetation indices require the handling and processing of a large volume of data, the development of a rigorous methodology adapted to the region and the issues of the study, and an important work of validation of the results, which implies the availability of field data.

However, we have seen that MCD12Q2 data are not yet operational on Mato Grosso since a large number of pixels has no data, and that existing data are often at odds with the field data. From our research it would appear that the main limitation for obtaining actionable MCD12Q2 data lies in the availability of the NBAR data used for the calculation of the phenological indicators. It is very likely that the lack of MCD43A4 data is due to contamination of acquisitions by a too large and frequent cloud cover. Atmospheric and MODIS reflectances geometric correction must be pursued in order to have a consistent data set on which apply the methodology of Zhang et al. (2003) for the phenological data usable.

MCD12Q2 data validation work should also be pursued, however this is dependent on the collection of field data. This is what suggest Ganguly et al. (2010) in their presentation of the collection of the product MCD12Q2 5. In this study, the authors note also MCD12Q2 product seems to get better performance for early season and beginning of maximum of vegetation indices. Our statistical data MCD12Q2 study appears to confirm this observation.

It would be interesting to perform a comparison with other areas of study to determine if the problems we met (lack of availability of data, data among themselves and with field data inconsistency) are specific to the tropical environment. Given the current version of the phenological MODIS product, it seems more reliable to study the phenology of tropical surface passing through the intermediate data, such as the vegetation indices.

For the rest of the work we conducted, the perspectives of research include:

- analysis of MCD12Q2 data on other areas of study;
- analysis of the robustness of the 0.6 EVI threshold value for different years ;
- improvement of the methodology for the estimation of vegetation growth start dates, especially by using derivative functions as a way to select only the first vegetation peak;
- adaptation of this methodology (or a new development) to determine other phenology transition dates, relating for example to the duration of the period of maturity of crops or vegetation end date, that it would be possible to link to field data, such as yield or the date of the soybean crop.

References

Arvor, D. **Etude par télédétection de la dynamique du soja et de l'impact des précipitations sur les productions au Mato Grosso (Brésil)**. 2009. Thèse de doctorat, Université Rennes 2.

Arvor, D.; Jonathan, M.; Meirelles, M.S.P.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time series for crop mapping in the state of mato grosso, Brazil. **International Journal of Remote Sensing**, v. 32, n. 22, p. 7847-7871, 2011.

Curnel, Y.; Oger, R. Agrophenology indicators from remote sensing: state of the art. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, tome XXXVI-8/W48, Stresa (IT), p. 31-38, 2006.

Ganguly, S.; Friedl, M.A.; Tan, B.; Zhang, X.; Verma, M. Land surface phenology from MODIS : characterization of the collection 5 global land cover dynamics product, **Remote Sensing of Environment**, v. 114, n. 8, p. 1805-1816, 2010.

Huete, A,; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices, **Remote sensing of environment**, v. 83, n. 1, p. 195-213, 2002.

Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-1639, 1964.

Vintrou, E.; Bégué, A.; Baron, C.; Lo Seen, D.; Saad, A.; Traoré, S. Analysing MODIS phenometrics quality on cropped land in West Africa. In: *Sentinel-2 Preparatory Symposium*, Frascati (IT), 23-27 April 2012, 7 p., 2012.

Zhang, X.; Friedl, MA.; Schaaf, CB; Strahler, AH; Hodges, JCF; Gao, F; Reed, BC; Huete, A. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, n.84, p. 471-475, 2003.