Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil

Gillian L. Galford [a,b,*], John F. Mustard [a], Jerry Melillo [b], Aline Gendrin [a,c],
Carlos C. Cerri [d], Carlos E.P. Cerri [e]

[a] Geological Sciences, Brown University, United States
[b] The Ecosystems Center, MBL, United States
[c] Institut d'Astrophysique Spatiale, Orsay, France
[d] Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Brazil
[e] Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Brazil

Abstract

Since 2000, the southwestern Brazilian Amazon has undergone a rapid transformation from natural vegetation and pastures to row-crop agricultural with the potential to affect regional biogeochemistry. The goals of this research are to assess wavelet algorithms applied to MODIS time series to determine expansion of row-crops and intensification of the number of crops grown. MODIS provides data from February 2000 to present, a period of agricultural expansion and intensification in the southwestern Brazilian Amazon. We have selected a study area near Comodoro, Mato Grosso because of the rapid growth of row-crop agriculture and availability of ground truth data of agricultural land-use history. We used a 90% power wavelet transform to create a wavelet-smoothed time series for five years of MODIS EVI data. From this wavelet-smoothed time series we determine characteristic phenology of single and double crops. We estimate that over 3200 km^2 were converted from native vegetation and pasture to row-crop agriculture from 2000 to 2005 in our study area encompassing 40,000 km^2. We observe an increase of 2000 km^2 of agricultural intensification, where areas of single crops were converted to double crops during the study period.
© 2007 Elsevier Inc. All rights reserved.

1. Introduction

The southwestern Brazilian Amazon is one of the world's fastest growing agricultural frontiers. Historically, the clearing of forest and savanna ecosystems to create cattle pastures has been the primary land transformation (Skole & Tucker, 1993). This land-use pattern has recently changed. Today, pastures and areas of natural vegetation are being converted to large-scale croplands to grow cash crops, row crops including soybean, maize, and dry-land rice (Instituto Brasileiro de Geografia e Estatistica, 2006; Morton et al., 2006). From 1990 to 2000, soybean cover in the southwestern Brazilian Amazon doubled while production has nearly quadrupled due to farm mechanization (CONAB,

2004). A major frontier of row crops is in the state of Mato Grosso, home of some of the largest contiguous row-crop plantations in the world. Here, the area planted in soybean has increased on average 19.4% annually since 1999. By 2004 over 5 million hectares, or about 6% of Mato Grosso, was soybean plantations (CONAB, 2004).

Regional shifts in land-cover and land-use have numerous consequences relevant to both environment and agriculture, including changes in carbon and nitrogen storage, trace gas emissions, quality of surface water and biodiversity (Luizão et al., 1989; Melillo et al., 1996, 2001; Myers et al., 2000; Neill et al., 1997, 2001; Steudler et al., 1996). Determining the physical and temporal patterns of agricultural extensificaiton, or expansion, and intensification is the first step in understanding their implications, for example, long-term crop production, and environmental, agricultural and economic sustainability.

---

* Corresponding author. Geological Sciences, Brown University, United States.
E-mail address: gillian_galford@brown.edu (G.L. Galford).

Since the early 1970s, remote sensing studies have tracked the land-cover and land-use changes in the Brazilian Amazon, initially using Landsat data to identify areas of deforestation (Skole & Tucker, 1993). The iconic images of development in the state of Rondonia, off of highway BR364, showed the dramatic impact of deforestation. Recently, the Brazilian Instituto Nacional de Pesquisas Espaciais (INPE; National Institute for Space Research) has used Landsat sensors for monitoring deforestation and detecting fires for the purpose of enforcing environmental regulations (Instituto Nacional de Pesquisas Espaciais, 2006). As conversion from natural vegetation and pasture to row crops has become increasingly widespread, the focus has shifted to documentation of the land cover changes in type localities. For example, Brown et al. (2005) illustrated the large-scale development of soybean agriculture with temporal "snap-shots" of Vilhena, Rondônia with Landsat data in 1996 and 2001. Morton et al. (2006) document widespread, regional changes in land cover. This new focus on cropland detection is particularly important due to the large spatial scale of individual farms (i.e. a single farm of row-crop agriculture typically occupies more than 2000 ha) and regional agricultural intensification (Mueller, 2003).

Remotely-sensed green leaf phenology is one metric for distinguishing the type of land-cover and land-use change and is suitable for agricultural applications. Croplands present a more complex phenology than natural land cover due to their many peaks resulting from multiple crops planted sequentially within a growing season. Additionally, the uniform cover of green leaves in an agricultural field creates very high observed greenness, especially as compared to the bare soils left after harvest. This large dynamic range of cropland green vegetation through time depends highly on natural factors (e.g. magnitude and temporal variability of precipitation in the region), as well as, management decisions (e.g. time of planting, crop variety).

Phenology studies often utilize a curve-fitting algorithm for the observed data sets. A curve-fit simplifies parameterization necessary for identification of metrics such as start of season. Previous studies (e.g. Bradley et al., 2007; Zhang et al., 2003) have identified land cover based on specific properties of the observed green leaf phenology, such as start of season, dry season minimums, and amplitude of maximums. The simplest method for creating a smoothed time series is to use a multi-point smoothing function which may not remove high frequency noise. Other curve-fitting methods, (e.g. Bradley et al., 2007) rely on a harmonic curve-fit to the annual average phenology in order to characterize inter-annual variability of a time series. A sigmod curve-fitting algorithm can be applied to a time series of a single year (Fisher et al., 2006; Zhang et al., 2003) but utilizes *a priori* knowledge of the system's seasonality in order to detect the phenological peaks (i.e. the algorithm must be informed with expectation for when to find the phenological peak) (Fisher et al., 2006; Jönsson & Eklundh, 2004; Zhang et al., 2003).

All of these methods have proven powerful in the systems they have been tested in but will fail in the case of row-crop agriculture in Mato Grosso for three reasons: 1) they fail to remove high frequency noise caused by the long rainy season. A multi-point smoothing function maintains sensitivity to high frequency noise observed during the rainy season which poses problems when trying to detect the maximums defining the cropping system. Using such a smoothing procedure would require a finely-tuned crop detection algorithm that would have to be adjusted for the strong precipitation gradients in the region. 2) They cannot capture the inherent variability in the system. Using an average annual phenology to identify land cover (Bradley et al., 2007) does not work because it does not examine each year separately, a problem since observed annual phenology is not a function of the previous year. In this human environment, the change in phenology from year to year (e.g. when converting from natural vegetation to cropland or intensifying from single crops to double crops) can be tremendous, rendering the average annual phenology of the time series meaningless. The traditional Fourier transform expects periodicity whereas the change in crop behavior from single to double cropping systems in addition to management coupled with climate makes the time-series signal non-stationary, which is better handled by the wavelet transform (Sakamoto et al., 2005). 3) The stochastic nature of rainfall and the influence of human management affect the timing and spatial patterns of phenology peaks that make it difficult to precisely predict the timing. This system fails the criteria of *a priori* knowledge regarding the timing of phenology peaks necessary for some curve-fitting algorithms, such as the sigmod (Fisher et al., 2006; Zhang et al., 2003).

We look to the wavelet-based curve-fitting methodology (Wavelet based Filter for determining Crop Phenology, WFCP) presented by Sakamoto et al. (2005) in order to remove high frequency noise while remaining sensitive to annual changes in phenology. This is a necessary step in accepting WFCP as a generalized methodology. The true utility of such a method comes in being able to apply it to various study areas without sacrificing performance or requiring many changes. Our case study presents a robust test for the wavelet methodology—an area with high variability in phenology patterns with additional noise cause by the tropical rainy season. We are interested in the application of the WFCP in the southwestern Brazilian Amazon because of its curve-fitting capabilities for cropland phenology, as well as, the potential for it to be rapid and highly automated.

We implement a wavelet transform for time-series analysis to study these highly dynamic systems. Wavelet analysis provides an efficient method for extracting relevant information from large data sets such as hyperspectral image cubes, sea surface temperature, vegetated land-cover and seismological signals (e.g., Gendrin et al., 2006; Li & Kafatos, 2000; Mallat, 1998; Percival et al., 2004; Sakamoto et al., 2005, 2006; Torrence & Compo, 1998).

In agricultural applications, a wavelet-smoothed time series can be used to identify the start of growing season and the time of harvest with low error (11 to 14 days, respectively; Sakamoto et al., 2005). Wavelet analysis is capable of handling the range of agricultural patterns that occur through time as well as the spatial heterogeneity of fields that result from precipitation and management decisions because the transform is localized in time and frequency. Using a wavelet analysis for a study area in the Amazon is highly desirable because it removes the high

frequency noise caused by the frequent cloud-cover in a highly automated way.

MODIS data products offer a great opportunity for phenology-based land-cover and land-use change studies by combining characteristics of AVHRR and Landsat, including: moderate spatial resolution, frequent observations, enhanced spectral resolution and improved atmospheric calibration (Justice, 1998; Zhang et al., 2003). MODIS data products have provided global land-cover mapping annually to document land-cover change over time (Friedl et al., 2002; Hansen & DeFries, 2004). These data sets are informative at the global level, but lack relevant regional-scale details about land-cover and land-use classes and change. Recent regional-scale applications of MODIS data to cropland land-use include spectral unmixing to time series to detect subpixel land-cover in croplands (Lobell & Asner, 2004). Wardlow et al. (2007) demonstrate that MODIS vegetation indices in time series are statistically sufficient for distinguishing crop types across a broad region, such as the state of Kansas. In the Amazon, Anderson et al. (2005) have utilized MODIS data sets to document broad changes in land cover and land use.

Understanding the degree of extensification and intensification in croplands from remote sensing provides insight into the direction and magnitude of impacts on natural and agricultural environments. In the industrial-scale croplands that are beginning to dominate portions of the Amazon Basin, patterns of cropland extensification and intensification have biogeochemical consequences that affect the natural and cropland sustainability, including soil fertility for decades to come. *Our objective is two-fold: to understand 1) the massive transition from natural vegetation and pasture to large-scale row crops and 2) the intensification of cropping systems within existing croplands using MODIS data sets.* The purpose of this study is to detect cropping patterns for the *cerrado* region using a wavelet-smoothed time series. This study evaluates wavelet tools, as presented by Sakamoto et al. (2005), in a new environment and tests the limits of the wavelet model while modeling crop phenologies from MODIS time-series data

during the 2001–2006 growing seasons. We provide analysis and discussion of the effectiveness of wavelet analysis for the detection of single and double cropping systems and hereby demonstrate the utility of wavelet analysis on time-series data with application to a land-use and land-cover change case study in the southwestern Brazilian Amazon.

2. Methods and approach

An overview of the methodology used here is presented in Fig. 1. The general methodology is based on the "Wavelet based Filter for Crop Phenology" (WFCP, Sakamoto et al., 2005). There are four main steps: 1) Data processing, 2) Identification of land use, 3) Field verification and 4) Error Analysis.

2.1. Study area

We selected a region of rapid change in croplands in the state of Mato Grosso for this study (Upper left corner: 12° 15′ 23.61″ S, 59° 45′ 18.23″ W; Lower right corner: 13° 59′ 27.86″ S, 57°57′ 30.8″ W; Figs. 2, 3). The area is 40,100 km^2, has annual rainfall from 1800 to 2200 mm, and a dry season from July–September and rainy (growing) season from November to April. The soils are entisols with 15–25% clay. Dominant native vegetation types range from *cerradão* (woody savanna) and *cerrado* (open savanna), referred to from here forward as "cerrado". In this region, land-use transitions have two major pathways to row crops from *cerrado*: natural vegetation to pasture to row crops; and natural vegetation directly to row crops. Row crops are subject to a variety of management regimes—types and sequences of crops; types, timing and amounts of fertilizer and other chemicals; and tillage versus no-tillage.

2.2. Field data

Field work conducted in July 2005 provided ground-control points on a *fazenda* of 41 km^2 for which a detailed agricultural
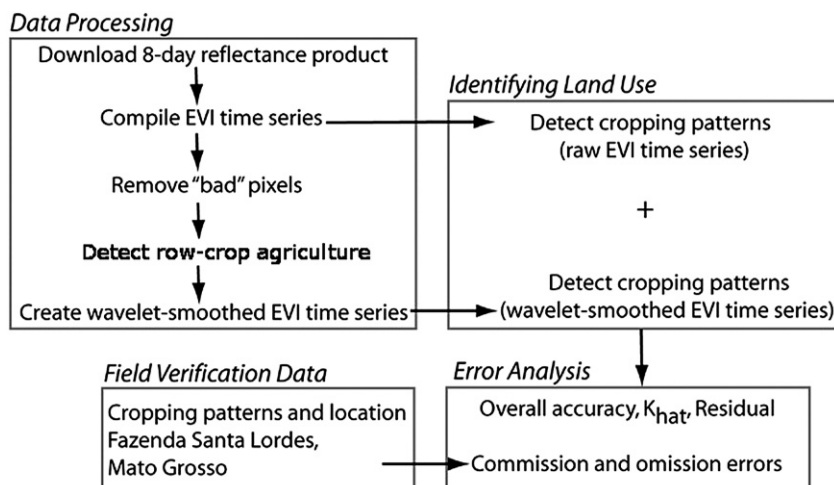

Fig. 1. Overview of methodology, divided into four parts: data processing, crop detection, field verification data and error analysis.
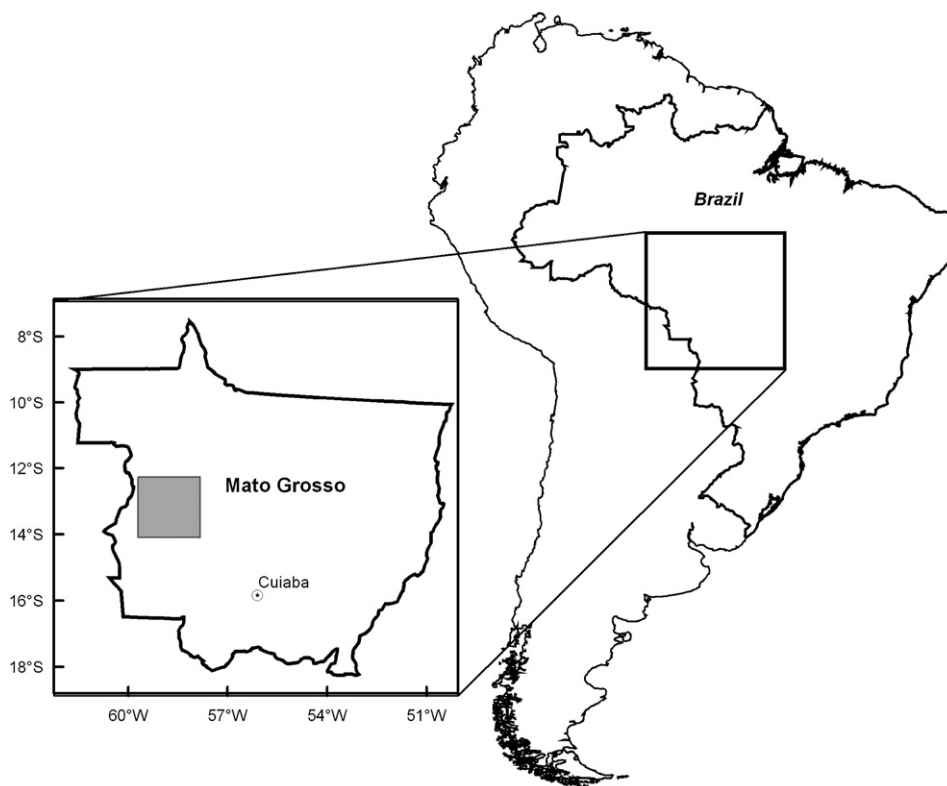
Fig. 2. Location map of the study area, shown by the gray box, in the southwestern Brazilian Amazon state of Mato Grosso.

history has been kept through this study period. Using a hand-held GPS we mapped three management units with different agricultural histories. The histories include information on the type of land cover or land use before row-crop agriculture, the timing of conversion to row-crop agriculture and the cropping patterns used for each year land use was row-crops. These histories were provided to us from records kept by the farm manager. Most of the native vegetation was converted in 2002 or 2003; one unit was previously pasture until conversion in 2002. The crops are generally single crops for the first two to three growing years and then change to double crops.

2.3. Creating a time series

Remotely-sensed data create a detailed classification of croplands by detecting important characteristics (parameters) of land-cover and land-use change from a smoothed Vegetation Index (VI) time series. For this work, we used MOD09 (V004) 8-day, 500 m surface reflectance composites data (Fig. 3). The study site was subset from the larger MODIS scene (h12v10).

We derived Enhanced Vegetation Index (EVI) products using the standard formulation (Huete et al., 2002): EVI was chosen because it has a greater dynamic range than the more commonly used NDVI and thus is better suited to capture the dynamic crop phenology in this region without reaching saturation (Huete et al., 2002). Combining the EVI images gives us an EVI time series for each pixel, although the time steps are not equally spaced. We used the date of observation flag included in the data product as the day of the year for each observation to create an unevenly-spaced time series for a given pixel. Using the date of

observation flags more accurately defines the timing and magnitude of green peaks. The 8-day product without the date of observation flags assumes evenly spaced observations when, in fact, observations can be up to 16 days apart or as few as 2 days apart. The accuracy of the shape of the input data affects the detection of cropping systems. Assuming evenly spaced data from with the original aggregated MODIS data (as with the 8-day product without the observation flags) misrepresents the data.

2.4. Data processing

Data processing for noisy and contaminated pixels consisted of 2 steps: 1) detecting of cloud-contaminated and extremely noisy pixels and 2) replacing bad data points through linear interpolation. Data processing treated each pixel as a one-dimension time series. For each time step, a point in the time series was identified as cloud-contaminated when band 3 (459–479 nm) reflectance values exceeded 10% (Sakamoto et al., 2006) and were subsequently removed. Extremely noisy data, generally caused by minor cloud contamination, were identified if they exceeded a 0.15 change threshold in EVI from the value at the previous time-step and were also removed.

We replaced the missing values through linear interpolation from observed data points. Since the observation dates varied by pixel and were unevenly spaced, we produced a daily time-step EVI time series to avoid an aliasing effect when creating a wavelet-smoothed time series. Then we resampled the daily interpolated data set to 7 day intervals to reduce the size of the data set and the processing time during further analysis.
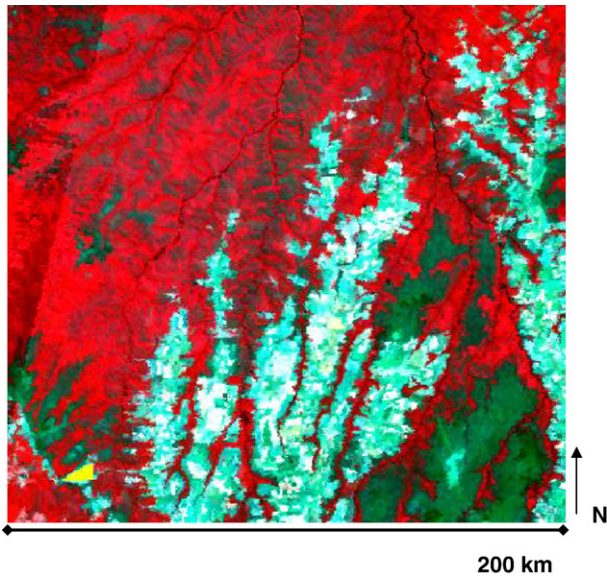
Fig. 3. False-color infrared MODIS image (Red=859 nm, Green=645 nm, Blue=555 nm) for the study area on 28 July 2005 (Upper left corner: 12° 15′ 23.61″ S, 59° 45′ 18.23″ W; Lower right corner: 13° 59′ 27.86″ S, 57°57′ 30.8″ W). Bright red areas represent dense *cerradão* woodland savanna native vegetation. Lighter reds to dark greens show the extent of *cerrado* native vegetation. Bright turquoise blues show row-crop agriculture, and very bright white areas are (bare) agricultural fields. Fazenda Santa Lordes is highlighted with a yellow polygon. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We first separate croplands from other land covers. Lands managed in croplands display a distinctly higher annual standard deviation compared to natural vegetation due to high vegetation density during the growing season and extremely low vegetation density following harvest. Woody *cerrado* land cover has very little phenological variation and maintains a mean value of 0.6 EVI, rarely exceeding 0.8 EVI (Fig. 4). The open *cerrado* phenology shows some seasonal variation as it decreases in greenness through the dry season and increases during the rainy season, with an annual mean of 0.25 EVI and a maximum around 0.6 EVI (Fig. 4). Both woody and open *cerrado* remain above 0.2 EVI through the year. Pasture phenology is similar to the *cerrado* with a slightly larger dynamic range and a mean of 0.5 EVI (Fig. 4). Croplands have maximum EVI values exceeding 0.8 and EVI minimums often reaching 0.1 or lower (Fig. 4).

Because these differences in amplitude of seasonal phenology in EVI we can use the standard deviation of an annual EVI time series to distinguish lands in croplands from native vegetation (Fig. 5). There is a bimodal distribution in the histogram of standard deviations for all pixels in the study area, separating croplands on the right tail (Fig. 5). There is a bimodal distribution of the standard deviation for each pixel. For each year, we used the standard deviation value that separated the two modes (the histogram minimum) as the detection point for croplands—all pixels with a standard deviation higher than this point were classes as croplands. The value of the detection point, or histogram minimum between modes, is included in Fig. 5 for each year of analysis. The mean detection point for all years was at the standard deviation of 0.149.

2.5. Creating a wavelet-smoothed EVI time series

Here we use a discrete wavelet transform. A wavelet function $\phi(t)$ is an oscillating function with a finite energy and null mean:

$$\int_{-\infty}^{+\infty} \phi(t)\mathrm{d}t = 0. \tag{1}$$

The wavelet transform $W(a,b)$ is defined by Eq. (2):

$$W(a,b)_i = \frac{1}{\sqrt{a}} \int \phi^* \left(\frac{t-b}{a}\right) s(t)\mathrm{d}t \tag{2}$$

where $s(t)$ is the analyzed input signal and ϕ^* is a mother wavelet, or a wavelet basis function. A number of different mother wavelets exist, including Daubechies, Derivative of a Gaussian (DOG) and Coiflet (Torrence & Compo, 1998). In this equation, the wavelet width is determined by the scaling parameter a while its center is determined by the parameter b. The variable t represents the time-step in the one-dimensional time series over which the integration is performed. The wavelet transform has the advantage of retaining information related to the width (scale) and the location (time) of the features present
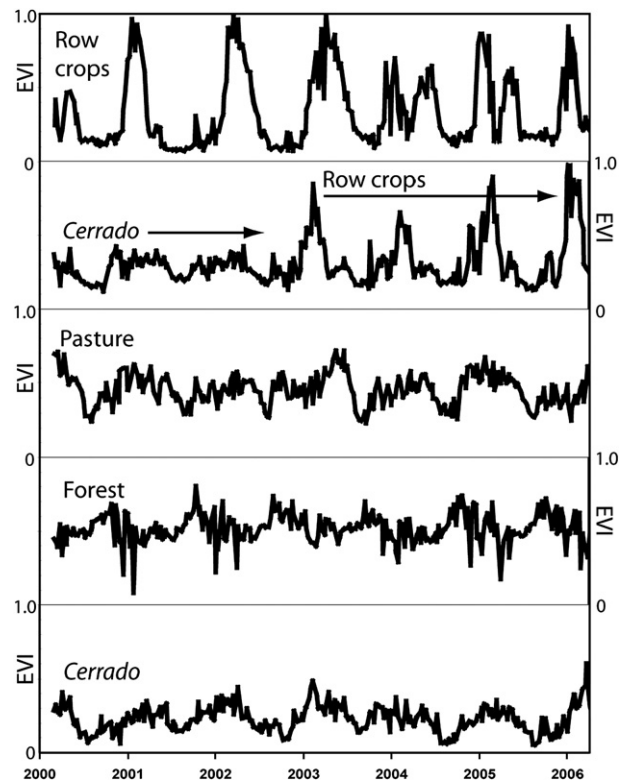

Fig. 4. EVI time series of representative land cover for 2000–2006 with cloud-contaminated points removed and filled by linear interpolation. Open c*errado* phenology has a low annual mean EVI and exhibits only minor seasonal fluctuations. Rainforest phenology has only slight changes in greenness between the wet and dry seasons. A pasture has higher annual mean and variance than *cerrado*, but a lower mean and higher standard deviation than forest. The time series second from the top shows an area in *cerrado* from 2000 to mid 2002, followed by a conversion prior to 2003 to cropland phenology. The top time series shows an area that exhibits single crop phenology in 2001, 2002, and 2003 and double crop patterns in 2004 and 2005.
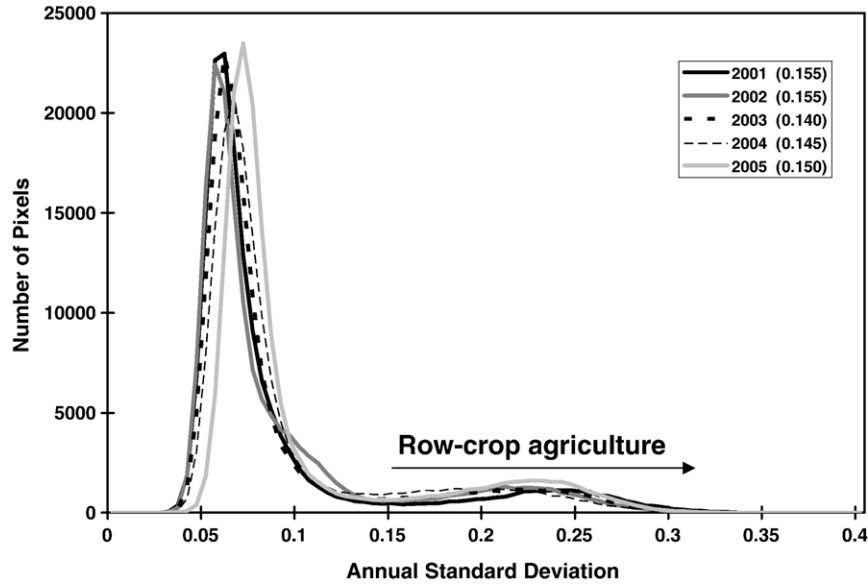
Fig. 5. This histogram show how the high standard deviation separates areas of row-crop agriculture from other land cover classes through the entire study scene. Row-crop agriculture was identified as areas that have an annual standard deviation greater than the local minimum (~0.15) identified specifically for that year. The actual standard deviation used for each year is printed in the legend.

in $s(t)$. This formulation (Eq. (3)) can be used to reconstruct a signal (Gendrin et al., 2006).

The wavelet-created time series W is a summation of wavelets over a number of different widths,

$$W = \sum_{i=1}^{x} W(a, b)_i \qquad (3)$$

where $W(a, b)_i$ is the wavelet transform created in Eq. (1). Wavelet transforms of decreasing width are summed from i to x, where x is the number wavelet transforms necessary to achieve the user defined number of coefficients retained from the input data. The width of a wavelet transform has half the width of the previous wavelet. It is the sum of the wavelets (W in Eq. (3)) that is referred from here on as the wavelet-smoothed time series.

The wavelet filtering begins by applying a smoothing function on the one-dimensional time series that is evenly-spaced, to avoid aliasing effects, after cloud-removal and interpolation of missing values. First, a discrete wavelet transform removes the residual high frequency noise. The smoothed EVI time series is then reconstructed with an inverse discrete wavelet transform.

Applying the wavelet to an EVI time series requires selecting parameters of mother wavelet, order, and power that define the wavelet behavior. We used the Coiflet mother wavelet with order 4 because the wavelet shape is as similar as possible to the peaks in agricultural phenology we are detecting. (See Sakamoto et al., 2005 for performance comparison of mother wavelets). Order is a measure of the wavelet's smoothness, where a higher order produces a smoother wavelet (Burke, 1994).

The wavelet requires a power threshold that corresponds to the number of coefficients determining how much of the input EVI time series is retained during the wavelet transform. A higher power or a greater number of coefficients retains more of the original data by forming a narrower wavelet that includes more fine-scale features but may also retain more noise. A lower power, or fewer coefficients, retains less high frequency data by applying a wider wavelet. A low power wavelet may capture trends through the entire time series but may loose phenological detail during a single year.

We conducted error analysis cropping patterns detected with the 70%, 80%, 85%, 90% (both 0.3 and 0.4 EVI detection thresholds; see Section 2.6 for further discussion of this threshold) and 95% (0.4 threshold) power wavelets. A random point generator selected 122 verification points within the row-crop agricultural zone. Reference data on cropping patterns for a given year in a given pixel were generated from the input EVI time series with bad pixels removed. These reference data are used to verify the cropping patterns detected from each wavelet-smoothed time series and the results tabulated to calculate overall accuracy and K_{hat}. Each point has five years worth of data and each year was treated individually, essentially multiplying the number of verification points by the number of years, giving a total of 610 verification points.

Table 1
The effects of wavelet power on accuracy are reported using overall accuracy and K_{hat} by comparing wavelet-detected cropping patterns to reference patterns generated by the user from the input EVI time series

Wavelet power (%)	Overall Accuracy	K_{hat}
70	81.0%	70.3%
80	83.4%	74.2%
85	90.0%	84.1%
90	87.2%	80.4%
90 (0.4 threshold)	88.5%	92.1%
95 (0.4 threshold)	44.8%	25.6%

Overall accuracy is the percent of points accurately identified in a class out of the total number of points sampled. K_{hat} values incorporate misclassifications while assessing classification accuracy.
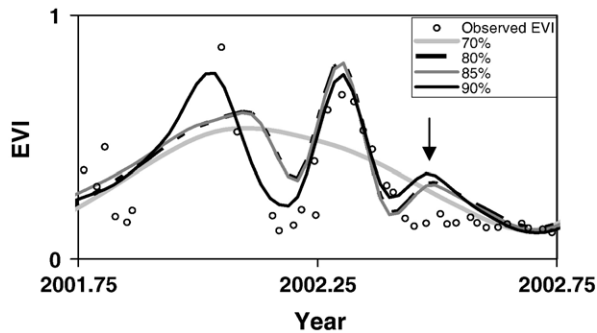
Fig. 6. Comparison of 70% power (9 coefficients), 80% power (15 coefficients), 85% power (19 coefficients) and 90% power (27 coefficients) wavelet-smoothed time series shows how wavelet power determines the amount of detail retained. Circles represent the EVI time series with cloud-contaminated values removed. Using a higher power may fit maxima in the data better but can also create false peaks, indicated here with an arrow. False peaks are removed from crop detection by an EVI threshold of 0.4.

This error analysis shows high overall accuracies and K_{hat} values for all wavelet powers except the 95% power wavelet (Table 1). From these results we conclude the 90% power wavelet-smoothed time series best captures cropping patterns and will be the focus of our further analysis. We employed the 90% power wavelet (27 coefficients) which minimized RMS error and had the lowest omission and commission errors in detecting cropping patterns. Qualitatively, we observed thousands of pixels where the strong overall performance of the 90% wavelet was apparent when compared to the other wavelet powers. Fig. 6 provides one such example. In areas of single crops, the 90% power wavelet captures the overall data trend well, getting closest to the high observed EVI values and the low values while reducing false detections of peaks. The agricultural system gets more complex where there are double crops. Resonance in the wavelet may create false peaks but they do not go above our threshold of 0.4 EVI. Throughout the time series, RMS error is low, on the order of magnitude of 0.1 for the 90% power wavelet-smoothed time series.

The power threshold is a variant on the method of Sakamoto et al. (2005) where multiple frequency thresholds defined the length of plausible growing seasons to remove noise with the wavelet. We chose to use the power variable to remove high frequency noise as it required no assumption of length of growing season. This gives more flexibility in fitting both very narrow peaks, as is often the case in both wide peaks found in single crops and very narrow peaks found in double cropping systems.

Application of the wavelet filter can create distortion around the edges of the EVI time series (Sakamoto et al., 2005). To avoid this problem, we augmented the input EVI time series to allow for spin-up, or conditioning of the wavelet. Conditioning the wavelet is a necessary step shown by Sakamoto et al. (2005). For our application, we augmented the one-dimensional EVI time series by replicating the first and last year worth of data ten times at the beginning and end, respectively, of the time series. After applying the wavelet transform, the extra years of data were removed from the wavelet-smoothed time series.

2.6. Phenology and land cover/land use

The cropping patterns, or the numbers of crops grown each year, were detected from the wavelet-smoothed EVI time series. Each crop is characterized by one maximum in the wavelet-smoothed EVI time series. To determine if a cropland pixel had a single or double cropping pattern, we detected the number of local maximums in one growing year of the wavelet-smoothed EVI time series. We defined a local maximum as having a higher EVI than the two points before and two points after that point. The wavelet-smoothed EVI time series divided into five growing years from August through July for 2000–2005 and are identified by their harvest year: 2001, 2002, 2003, 2004, and 2005.

Wavelets are very sensitive to small maximums in portions of the time series where EVI range is low. The wavelet response to these small local maximums slightly amplifies small real peaks, thereby creating false peaks in phenology. From farm histories, we know that the EVI for crops generally exceeds 0.4. We removed false detections by using a threshold of 0.4 EVI for the 90% power wavelet to minimize false detections to remove these minor false peaks from being detected as phenological peaks of cropland. In some areas we detect two or three real phenological maximums in the EVI time series. In such cases, the first maximum is minor and is likely caused by the early green-up of volunteer crops, weeds or other green cover at the beginning of the rainy season before crops are planted. The second and third maximums, where present, correspond to single and double crops. The early weedy growth or volunteer crops may be distinguished from crops as they do not exceed the 0.4 EVI threshold that prevents us from detecting early green-ups as crops. Utilizing this threshold works with the assumption that every area of row crops has a strong crop phenology (i.e. peaks above 0.4 EVI) but it may exclude very small, real, phenological maximums such as the case of a failing crop that did not exceed the threshold.

With these detection criteria, we can identify the cropping patterns and change in cropping patterns that characterize the intensification and extensification of cropland, as first shown by Sakamoto et al. (2006). For verification of the cropping patterns, we used the observed EVI time series with bad pixels removed. A random point generator selected 122 verification points within the cropland areas. For each point, there are five growing seasons (August–July) in the time series. Each growing season was treated individually, essentially multiplying

Table 2
Total area in row-crop agriculture * is reported here in square kilometers by harvest year

Year	Area (km^2)
2001	6255
2002	6799
2003	7543
2004	8532
2005	9535

* The area of row-crop agriculture was measured by the number or pixels having one or more crops in a year of the 90% wavelet-smoothed time series.

Fig. 7. The 90% power wavelet-smoothed time series results for detecting maximums representative of single and double crops are presented here. Cropping patterns in the study region show an increasing area is cultivated in double crops (black) instead of single crops (gray). The increase in double cropping (black) is notably centered in existing agricultural zones. Extent of row-crop agriculture over time is a result of the standard deviation threshold described in Fig. 3. The extent of row-crop agriculture (shown in black and gray) is observed to be spreading. Change is particularly notable on the edge of the cropland region, such as the areas circled in dotted lines.

the number of verification points to give a total of 610 verification points. We compared cropping pattern detected from the 90% wavelet-smoothed time series to the original data and, for each point, identified and tabulated misclassifications.

Further, comparison of our wavelet-smoothed time series to agricultural history for Fazenda St. Lordes allows us to assess the detection of cropping patterns. There are multiple different land-use histories corresponding to the management units within the *fazenda*. The agricultural history collected from farm records during a field visit in July 2005 includes the time of conversion as well as sequence and cropping patterns in subsequent years for each management units. We examine how well the process of cleaning the EVI time series and performing the wavelet-transform retains the character of the processes occurring on the ground by comparing the detected cropping patterns to the farm records.

2.7. Error analysis

We performed statistical analysis of the goodness of the curve fitting through residuals and Root Mean Square (RMS) error. We calculated the RMS error of the residual (difference between the raw EVI times series and the wavelet-smoothed time series). The RMS error gives a sense of the magnitude of error with the curve fit.

We analyzed error in the land-cover and land-use classes by comparing cropping patterns detected from the wavelet-smoothed time series to observed cropping patterns in the input data. By compiling an error matrix for the classes we could calculate overall accuracy, producer's accuracy and user's

accuracy as well as a K_{hat} value from KAPPA analysis (see Jensen, 1996).

To perform the statistical error assessments, we divided our data into four classes. The classes, based on data values, are: not cropland, single cropping system (one maximum), double cropping system (two maximums) and unclassified (more than two maximums detected). As a given pixel may change classes from one year to the next, we considered a pixel's class for one growing year a test point. We calculated omission and commission errors using two different sets of reference data, the raw EVI time series and spatially and temporally explicit farm history data. Overall accuracy is the total number of test points correctly classed by the total number of test pixels used. Omission error or producer's accuracy is the total number of correct pixels in a remotely-sensed class divided by the total number of pixels in that class from the reference data. Commission error, or user's accuracy, is the total number of correct pixels in a remotely-sensed class divided by the total number of pixels

Table 3
This table presents randomly selected unclassified pixels tabulated by their reference categories *

		Classification				
		Not RC	Single	Double	Unclassified	Total
Wavelet power	70%	0	10	80	9	99
	80%	0	12	71	17	100
	85%	1	10	89	0	100
	90%	0	19	73	8	100
	90% (0.4 threshold)	2	16	78	4	100
	Total	3	67	391	38	

* Except where noted, the wavelet-smoothed time series were analyzed with a 0.3 EVI detection threshold.

Table 4
Randomly generated points throughout the study area were used for error analysis of the land-use classifications: not row crops (Not RC), single cropping patterns (Single) and double cropping patterns (Double). The wavelet-detected results are compared to the reference data, in this case crop patterns detected by the user from the non-smoothed MODIS time series. Error matrix (A) shows the number of points correctly classified as well as the distribution of misclassified points. The producer's accuracy and user's accuracy (B) is low for Not RC Agriculture and Double crops but is rather high for Single crops

A		Reference data (non-smoothed MODIS cropping patterns)			
		Not RC	Single	Double	Total
Wavelet-detected cropping patterns	Not RC	100	0	0	100
	Single	1	87	12	100
	Double	2	26	168	196
	Total	103	113	180	396

B		Producer's accuracy	User's accuracy
Wavelet-detected cropping patterns	Not RC	97.1%	100.0%
	Single	77.0%	87.0%
	Double	93.3%	84.0%

Table 5
The overall accuracy of our algorithm was assessed by comparing our automated
crop detection techniques preformed on the wavelet-smoothed EVI time series
to the input EVI time-series data (non-smoothed MODIS cropping patterns
detected by the user or MODIS) and the farm history from Fazenda Santa Lordes

Reference data	Overall Accuracy	K_{hat}
MODIS	88.5%	92.1%
Fazenda	94.0%	85.7%

Overall accuracy is the percent of points accurately identified in a class out of the
total number of points sampled. K_{hat} values are also high for all wavelet powers.

in that remotely-sensed class. The K_{hat} statistic comes from KAPPA analysis for discrete multivariate accuracy assessment. K_{hat} incorporates information from the misclassifications recorded in the error matrix and gives a slightly different accuracy assessment than overall accuracy does. K_{hat} would equal zero if the classifications results were completely random. (Jensen, 1996).

For the raw EVI reference test case, we calculated overall accuracy by comparing the user-detected reference classes to the automated detection of classes from the wavelet-smoothed time series. We used one hundred test points per class. The test points were evenly distributed by year (20 test pixels per year per class) for the not cropland and unclassified classes since there was negligible change in the size of these classes over time. We weighted test points for single and double crop classes by the relative abundance in that year. We randomly located the test points for each year and class using a random sample. We then compared the wavelet-smoothed time-series crop detection results to the raw EVI input data (reference test information) that was subjected to the same criteria to create four classes. For the random test points across the entire scene, we use omission and commission errors to understand our accuracy within the classes as well as overall accuracy and the K_{hat} value.

We also used reference data (known cropping patterns) for Fazenda Santa Lordes to calculate omission and commission errors over the farm. We verify our results to an independent data source. One limitation to this method is the size of the fazenda. While this is a large fazenda, occupying 25 km^2, we are limited to a relatively small sample size (100 pixels) for statistical analysis.

Extensification and intensification are two measures of the extent of row crop agriculture. Extensification, or the increase in total row-crop agricultural area, is measured as the annual increase from one growing season to the next. Each year the area of extensification is calculated as the areas detected as row crops that were not previously detected. Intensification of row crops describes the change from a single to double cropping pattern from one year to the next. The concepts and metrics of extensification and intensification allow us to explain the patterns of agricultural development.

3. Results

3.1. Agricultural extensification

Cropland in the study area increased from 6255 km^2 in the 2001 growing season to 9535 km^2 in the 2005 growing season, as calculated from cropland detection based on annual standard deviation (Table 2). This represents a 34% increase in row-crop agriculture to cover a total area of 24% of the study area. Increases in land cover of row crops were largely at the edges of existing croplands (Fig. 7).

3.2. Detection of cropping patterns

Statistical analysis of detection errors shows overall good results from the wavelet-smoothed time series. The error matrix shows that majority of pixels considered unclassified (more than two crops detected) were actually false-detections of double crops (Table 3). For the purpose of tabulating single and double cropping patterns, we considered double crops to be all pixels with two or more maximums, although this may introduce more error. From the error matrix results we can asses our accuracy (Table 4A). The omission and commission errors (Table 4B) are derived from the error matrix for each classification (Not row crops, 'Not RC', includes native vegetation and pastures; Single cropping patterns; and double cropping patterns). All omission errors are less than 10%, except for Single crops, which have a producer's accuracy of 77.0%. User's accuracy is above 84.0% for all classes analyzed (Table 4B). The wavelet-smoothed time series gives us an
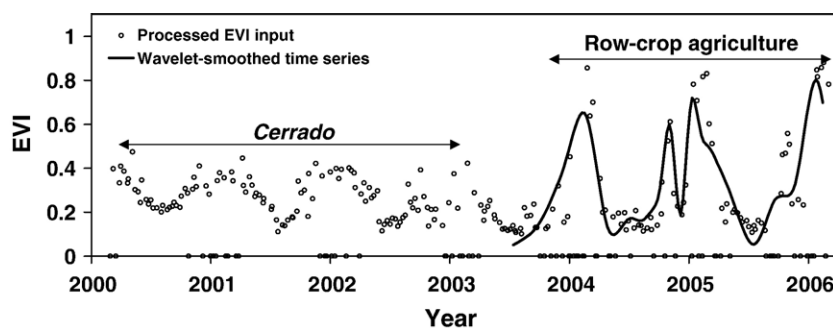

Fig. 8. A field site at Fazenda Santa. Lordes is represented in this time series. The EVI daily time series (2000–2006) after processing for clouds and extremely noisy pixels is plotted with small circles. The solid line shows the 90% power wavelet-smoothed time series. Land cover is converted from *cerrado* to row-crop agriculture in the middle of 2003. In 2004, a single rice crop was grown. In 2005, two soybean crops were grown.

Table 6
An error matrix shows the agreement and disagreement between reference data (Fazenda Santa Lordes farm history) and the 90% wavelet-detected cropping patterns (*A*). The producer's accuracy and user's accuracy (B) are shown for each land-use classification

A		Reference (farm history)			
		Not RC	Single	Double	Total
Wavelet-detected	Not RC	71	4	0	75
	Single	0	18	2	20
	Double	0	0	5	5
	Total	71	22	7	100

B		Producer's accuracy	User's accuracy
Wavelet-detected	Not RC	100.0%	94.7%
	Single	81.8%	90.0%
	Double	71.4%	100.0%

Table 7
The cropping patterns for each year are presented here

Area (km^2)		Area (km^2) by cropping pattern	
		Single	Double
Year	2001	3124	3131
	2002	3251	3548
	2003	4063	3480
	2004	2790	5742
	2005	3643	5892

These are the results from the 90% power wavelet-smoothed time series using a 0.4 EVI crop detection threshold, as it performs with the lowest misclassifications (Table 5). The intensification of double crops after 2003 is particularly notable.

overall accuracy of 88.5% for the entire study region and the corresponding K_{hat} value is 92.1% (Table 5).

The wavelet transform captures land cover and land use at a site on Fazenda Santa Lordes where *cerrado* was converted to row crops in 2003, with the first crop grown in the 2003–2004 season (Fig. 8). From farm history, we know the first year of cropping was a single soybean crop in 2003–2004 and a double crop in 2004–2005. Statistical error analysis for wavelet-derived classes compared with land-use records for Fazenda Santa Lordes shows high accuracy (Table 6). Producer's accuracy is high (low omission errors) the classifications of not row crops (100.0%) and single crops (81.8%) but low (higher omission errors) for double crops (71.4%). User's accuracy is

high (low commission error) for all classes: not row crops is 94.7%; single crop accuracy is 90.0% and double crops accuracy is 100.0% (greater than 99.99%). The commission and omission errors may not be representative of the entire scene as there were only 5 pixels in this category. The K_{hat} value is 85.7% and the overall accuracy is 94% for land-use classifications on the fazenda (Table 5).

Single and double cropping patterns show a dynamic relationship (Fig. 9). Both cropping patterns have a net increase (Table 7). Single crops increase from 3124 to 3643 km^2 (a 14% increase) and double crops increase from 2283 to 4443 km^2 (a 49% increase) during the study period (Fig. 9). We see a decrease in single crops between the 2003 and 2004 growing years while double crops continue to increase. The increase in single crops from 2002 to 2003 may represent the extensification of croplands into areas that were previously native vegetation. After the first growing year with a single crop many of these areas may have been converted to a double crop, creating a dynamic relationship between the cropping patterns. Agricultural intensification, or increase in the number of crops grown per area, follows a pattern similar to agricultural extensification,
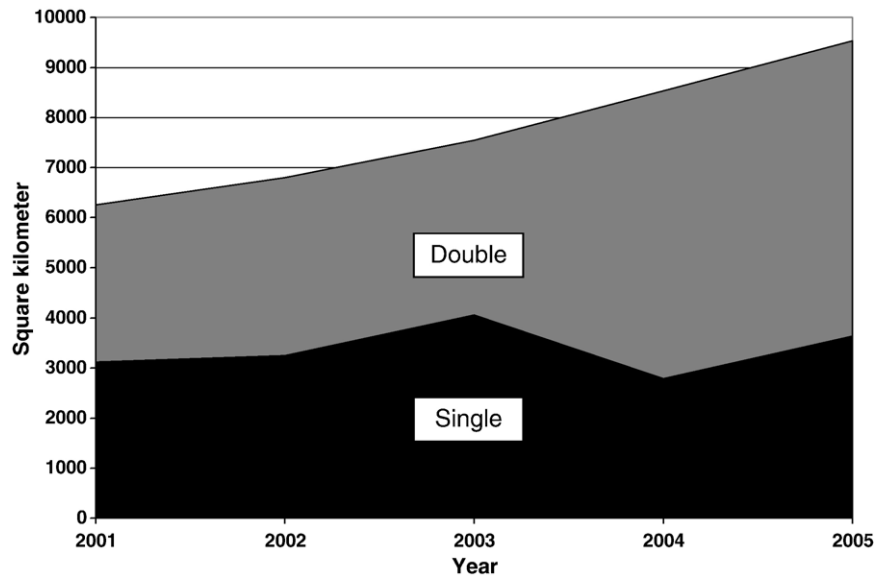

Fig. 9. The 90% power wavelet with a 0.4 EVI threshold for detecting maxima preformed with this lowest misclassifications (Table 2). These results were used to track the area in single and double cropping patterns by year. The area in croplands increases through the time series. The intensification of double crops after 2003 is particularly notable.

radiating outwards from the older mechanized agricultural areas to the periphery with time (Fig. 7).

4. Discussion

A major frontier croplands is found in the state of Mato Grosso, Brazil where natural ecosystems of Amazon rainforest and *cerrado* (savanna) are giving way to some of the largest contiguous row-crop plantations in the world. Agricultural census provides detailed information on cultivated area and crop yields, but is not spatially explicit beyond the level of *municipio* nor does it tell us the sequence and duration of crops. Census data is compiled annually, but does not elucidate the number of crops grown per parcel, or other land-use practices that impact local biogeochemistry. Remote sensing tools allow us to detect agricultural phenology and derive parameters from which we can construct timeline of land-use and land-management on a per pixel basis. These results have many possible future applications such as for understanding changes in carbon and nitrogen cycling, important from both agricultural and environmental sustainability perspectives, or for estimating crop production rates for economic analysis.

Our application of the wavelet transform to an EVI time series (WFCP) captures crop phenological behavior with low error. This was a test of the WFCP methodology under new environmental conditions, namely climate which creates high frequency noise from clouds, and different agricultural phenology than the areas the model was created and validated (Sakamoto et al., 2005, 2006). Occasionally, the wavelet transform exaggerates small phenology peaks from weedy growth or creates a false peak due to wavelet resonance. Small peaks such as these should be classified with double crops. Although the classification of false maximums or weedy growth as double crops is troublesome from a curve-fitting perspective, it is acceptable for the purposes of detecting cropping patterns. Using the 90% power wavelet, many of these false maximums actually fall below the detection threshold for maximums of 0.4 EVI and are not counted, contributing to our high overall accuracy.

We have observed distinct spatial patterns in the wavelet-smoothed time-series results. New areas of croplands are nucleated around existing areas of croplands (Fig. 7). Small areas of *cerrado* between large areas of croplands are often filled in with croplands. We observe areas of single crops becoming areas of double crops over time, as seen in Fig. 7. This intensification appears spatially constrained to the center of the zones of agriculture and likely reflects the evolution of farming practices with cropland age. Our knowledge of agricultural practices supports this—often, in the first years of cultivation, a single soybean crop is grown. After two to three years, double cropping practices emerge, where there may be two soybean crops grown, soybean and secondary cash crop such as corn, or soybean and a soil-conditioning crop such as millet. Increases in crops in the 2002–2004 time period may be related to economic factors, such as the high global market price (Morton et al., 2006). These results show an increase of crops grown per area across the entire study area, either as the development of new croplands or as the intensification of existing lands from a single crop system to a double crop system.

5. Conclusions

The goal of this study was to apply time-series analysis to detect rapid changes in land-cover and land-use choices (single and double cropping patterns). The challenge of this study was to detect crop patterns within croplands. We have tested the wavelet transform to filter noisy EVI time-series data. First, we detect the areas of row crops by applying an annual standard deviation threshold to discriminate row crops from other land-cover types. This threshold was selected annually from the local minimum in a bimodal histogram of standard deviation. Identifying areas of row crops on a year-to-year basis allows us to analyze cropping patterns for a pixel only after it has been converted to row crops, thereby reducing processing time. After selecting areas in row crops, we then created wavelet-smoothed time series with the 90% power wavelet. Local maximums, or phenological peaks, were counted as a crop if the time series exceeded 0.4 EVI; this threshold was a means of removing false peaks sometimes created in the wavelet-smoothed time series.

We selected this study area to test this model in an area of rapid development of row crops. During the five-year study period, we found an increase in croplands of 3281 km^2, an area larger than the state of Rhode Island. Intensification of row crops is also evident, with increases in row crops coming first in single crops and, subsequently, in double crops, such as in 2003 and 2004. We expected this pattern where there is an extensification in row crops that would typically be grown in a single crop pattern for the first growing season. In subsequent years agricultural intensification, or a shift from growing one crop to two crops, could explain the increase in double crops. Spatially, the extensification of row crops is on the edges of existing areas of agriculture and intensification occurs within the existing areas of croplands.

These results show that there is a large increase in cropland in our study area and it is important to understand how this drastic change in land cover will (and does) impact carbon and nitrogen cycling. Distinguishing crop types, such as soybean and corn, is important as different crops have different implications for carbon and nitrogen cycling. Soybean plants fix nitrogen, but most of the fixed nitrogen leaves the system at harvest. Without proper management, over time, the loss of carbon and nitrogen decreases the soil fertility and may have other implications for land-use sustainability and management. Secondary crops, such as corn, may require large inputs of nitrogen fertilizers that increase nitrous oxide emissions. Addition of nitrogen fertilizers also impact local water quality. Knowing the number and type of crops being used allows us to proceed with spatially explicit models of biogeochemical changes associated with this agricultural development and intensification.

In this study we demonstrate the stability of the wavelet approach over many years of an EVI time series. Thereby we can apply the WFCP methodology to a wider study region to study regional development patterns of row crops. One future

study includes parameterizing the wavelet-smoothed time series to identify crop types. We foresee the ability to use other types of time series in this wavelet method of analysis (WFCP), such as fractions of green vegetation from Spectral Mixture Analysis.

Acknowledgements

References

Anderson, L. O., Shimabukuro, Y. E, Defries, R. S., & Morton, D. (2005). Assessment of land cover and land use changes in the Brazilian Amazon using multi-temporal fraction images derived from Terra MODIS: Examples from the state of Matos Grosso. *IEEE Transactions on Geoscience and Remote Sensing*, *2*(3), 315–318.

Bradley, B., Jacob, R., Hermance, J., & Mustard, J. F. (2007). A curve-fitting technique to derive inter-annual phonologies from time series of noisy NDVI satellite data. *Remote Sensing of Environment*, *106*, 137–145.

Brown, C. J., Koeppe, M., Coles, B., & Price, K. P. (2005). Soybean production and conversion of tropical forest in the Brazilian Amazon: The case of Vilhena, Rondonia. *Ambio*, *34*(6), 462–469.

Burke, B. (1994). The mathematical microscope: Waves, wavelets, and beyond. In M. Bartusiak, B. Burke, A. Chaikin, A. Greenwood, T. A. Heppenheimer, M. Hoffman, D. Holzman, E. J. Maggio, & A. S. Moffat (Eds.), *A positron name Priscilla: Scientific discovery at the Fronteir* (pp. 196–235). Washington DC: National Academy Press.

CONAB (Companhia Nacional de Abastecimento) (2004). Homepage: www. conab.gov.br

Fisher, J. I., Mustard, J. F., & Vadeboncoeur, M. A. (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment*, *100*, 265–279.

Friedl, A. F., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of the Environment*, *83*, 287–302.

Gendrin, A., Langevin, Y., Bibring, J. P., & Forni, O. (2006). A new method to investigate hyperspectral image cubes: An application of the wavelet transform. *Journal of Geophysical Research*, *111*(E10004). doi:10.1029/2004JE002339

Hansen, M. C., & DeFries, R. S. (2004). Detecting long-term global forest chance using continuous fields of tree-cover maps from 8-km Advanced Very High Resolution Radiometer (AVHRR) data for the years 1982–1999. *Ecosystems*, *7*, 695–716.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*, 195–213.

Instituto Brasileiro de Geografia e Estatistica (IGBE) (2006). *Systematic survey of agricultural production.* IGBE website accessed October 23, 2006.

Instituto Nacional de Pesquisas Espaciais (INPE) (2006). Homepage: http://www.inpe.br/

Jensen, J. R. (1996). *Introductory digital image processing: A remote sensing perspective* (pp. 197–253)., 2nd ed. Upper Saddle River: Prentice Hall.

Jönsson, P., & Eklundh, L. (2004). TIMESAT—A program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, *30*, 833–845.

Justice, C. O. (1998). The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geosciences and Remote Sensing*, *36*(4), 1228–1249.

Li, Z., & Kafatos, M. (2000). Interannual variability of vegetation in the United States and its relation to El Niño/Southern Oscillation. *Remote Sensing of Environment*, *71*, 239–247.

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

Luizão, F., Matson, P., Livingston, G., Luizão, R., & Vitousek, P. (1989). Nitrous oxide fluxes following tropical land clearing. *Global Biogeochemical Cycles*, *3*(3), 281–285.

Mallat, S. (1998). *A wavelet tour of signal processing.* San Diego: Academic Press 637 pp.

Melillo, J. M, Houghton, R. A., Kicklighter, D. W., & McGuire, A. D. (1996). Tropical deforestation and the global carbon budget. *Annual Review of Energy and the Environment*, *21*, 293–310.

Melillo, J. M., Steudler, P. A., Fiefel, B. J., Neill, C., Garcia, D., Piccolo, M. C., et al. (2001). Nitrous oxide emissions from forests and pastures of various ages in the Brazilian Amazon. *Journal of Geophysical Research*, *106*(D24), 34,179–34,188.

Morton, D. C., DeFries, R. S., Shimbukuro, Y. E., Anderson, L. O., Arai, E., de Bon Espirito-Santo, R., et al. (2006). Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. *PNAS*, *103*(39), 14637–14641.

Mueller, C. (2003). *Expansion and modernization of agriculture in the cerrado—The case of soybeans in Brazil's Center-West. Department of economics working paper*, *Vol. 306*. Brasilia: University of Brasilia.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., de Fonseca, G. A. B, & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858.

Neill, C., Deegan, L., Thomas, S. M., & Cerri, C. C. (2001). Deforestation for pasture alters nitrogen and phosphorus in small Amazonian streams. *Ecological Applications*, *11*(6), 1817–1828.

Neill, C., Melillo, J. M., Steudler, P. A., Cerri, C. C., Moraes, J. F. L., Piccolo, M. C., et al. (1997). Soil organic matter stocks following forest clearing for pasture in the southwestern Brazilian Amazon. *Ecological Applications*, *7*(4), 1216–1225.

Percival, D. B., Wang, M., & Overland, J. E. (2004). An introduction to wavelet analysis with applications to vegetation time series. *Community Ecology*, *5*(1), 19–30.

Sakamoto, T., Nguyen, N. V., Ohno, H., Ishitsuka, N., & Yokozawa, M. (2006). Spatio-temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers. *Remote Sensing of Environment*, *100*, 1–16.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, *96*, 366–374.

Skole, D., & Tucker, C. (1993). Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. *Science*, *260*(5116), 1905–1910.

Steudler, P. A., Melillo, J. M., Feigel, B. J., Neill, C., Piccolo, M. C., & Cerri, C. C. (1996). Consequence of forest-to-pasture conversion on CH_4 fluxes in the Brazilian Amazon Basin. *Journal of Geophysical Research*, *101*(D13), 18,547–18,554.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, *108*, 290–310.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*, 471–475.