

YSSP Interim Report

IR-14-017

Satellite time series analysis for land use/cover change detection

Victor Maus, vwmaus1@gmail.com

Approved by

Aline Mosnier, Ian McCallum
YSSP Supervisors

Michael Obersteiner
Program Director, Ecosystems Services and Management Program

October 1, 2013

YSSP Interim Reports present work of the International Institute for Applied Systems Analysis carried out by participants of its Young Scientists Summer Program (YSSP) during their three-month stay at IIASA and receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

Contents

Introduction.....	1
Methods	3
Results and discussions	5
Conclusions.....	8
References	10
Appendix A. Main functions used to time series analysis.....	12

Abstract

Currently, Brazilian land use data comes from the national agricultural census and land cover data comes from global data sets with sparse temporal coverage. This no longer meets the needs of the earth system modeling community. Long-term satellite image datasets with high temporal frequency yield a sequence of data points in a time series that can be used to detect and monitor land use and land cover changes. The vegetation phenological cycles are reflected in the satellite time series, allowing the classification of land cover types in time segments. This research aims at developing an automatic methodology to yield information about land use and land cover trajectories. To construct land use/cover trajectories maps, Dynamic Time Warping (DTW) is used to extract information from the MODIS 2-band Enhanced Vegetation Index (EVI2) time series. Validation tests were made in the areas of Mato Grosso state, Brazil. The preliminary results for the proposed methods are promising when compared with the official TerraClass land use maps in the Amazon Biome, finding 78.2% and 85.0% global accuracy for 2008 and 2010, respectively. Exploratory DTW results show significant potential to detect land use and cover changes.

Acknowledgments

I would like to thank my YSSP supervisors, Aline Mosnier and Ian McCallum, for the patient guidance, encouragement and advice they provided throughout the summer. I would also like to thank all the members of staff who helped me during my time at IIASA, especially Tanja and Aleks. I would like to thank Florian Kraxner and my Ph.D supervisors Gilberto Câmara and Fernando Ramos, who encouraged me to participate in the summer program. In particular I wish to thank also the Brazilian CGEE funding. I would like to express my acknowledgements to all the YSSPers for all the conversations and moments together.

About the Authors

Victor Maus has a Bachelor of Engineering degree in Environmental Engineering (2009) and received his MSc degree in Computational Modeling from the Federal University of Juiz de Fora in 2011. He is currently a third year PhD student in Earth System Science of the National Institute for Space Research, Brazil. His main fields of scientific interest include mathematical modeling, numerical simulation and geoinformatics. Victor has been professor at Federal University of Pampa since 2012 and member of the INPE's team in the REDD-PAC Project.

Satellite time series analysis for land use/cover change detection

Victor Maus

Introduction

In the 1960s and 1970s government policies and subsidies to develop, populate and integrate the Brazilian Amazon region into the rest of the country, induced extensive and predatory use of natural resources in that region (ARAUÚJO and LEÉNA, 2011). The Brazilian Amazonian rain forest originally occupied an area of 4,100,000 km², from which 734,298 km² have been deforested since the 1970s. Recently, the annual deforestation is decreasing and last year reached the lowest value since the beginning of monitoring, 4,571 km² (INPE, 2013).

Understanding the dynamics of land use and land cover change even after deforestation can help us to understand the drivers of deforestation, and could be helpful to design better policies to stop deforestation and accompany local actors towards more sustainable land management. The time series of land use and land cover maps can be useful, to answer related questions like: What are the main land use trajectories after deforestation? What land use trajectories can lead to land abandonment or land degradation? What is the effectiveness of efforts to deforestation reduction?

Currently, there is a lack of time series maps which would demonstrate the land use and land cover dynamics in Brazil. Brazilian land use data comes from the national agricultural census and land cover data comes from global data sets with sparse temporal coverage. The Brazilian Institute of Geography and Statistics (IBGE) conducts the agricultural census which provides information about areas under cultivation in the national territory for each municipality (IBGE, 2006). For the Amazon rain forest the Brazilian National Institute for Space Research (INPE), through the PRODES Project, monitors the forest area and estimates the annual rate of deforestation since 1989, but without information about land use after deforestation. INPE also develops the TerraClass Project. This Project provides detailed land use information in the Amazon but is only for two years, 2008 and 2010, which is insufficient to evaluate issues related to different land trajectories.

Potentially, we can use long-term satellite images with high temporal frequency to extract land use and land cover trajectories. The sequence of images allow us to compose time series for each pixel that show the time variation of surface attributes such as the reflectance of a specific band of the electromagnetic spectrum or derived indexes (Figure 1). Vegetation cover studies typically use time series of vegetation indexes e.g., Normalized Differences Vegetation Index (NDVI) computed through the reflectance in red and near infrared bands (ROUSE et al., 1973, 1974); Enhanced Vegetation Index (EVI) computed through the reflectance in blue, red and near infrared bands (JUSTICE et al., 1998); and recently the 2-band EVI (EVI2), an adaptation of EVI without the blue band (JIANG et al., 2008).

This vegetation index reflects the seasonal cycles of land cover and its changes over time that are typically correlated with land use changes. The Figure 2 shows a sample of land use changes reflected in the vegetation index time series. In this sample, the area initially was covered by forest that was removed in 2005 and converted to pasture until 2010. After that a land use change occurred, to single cropping in 2011 and to double cropping in 2012, i.e. it is two crops in the same space during a single growing season.

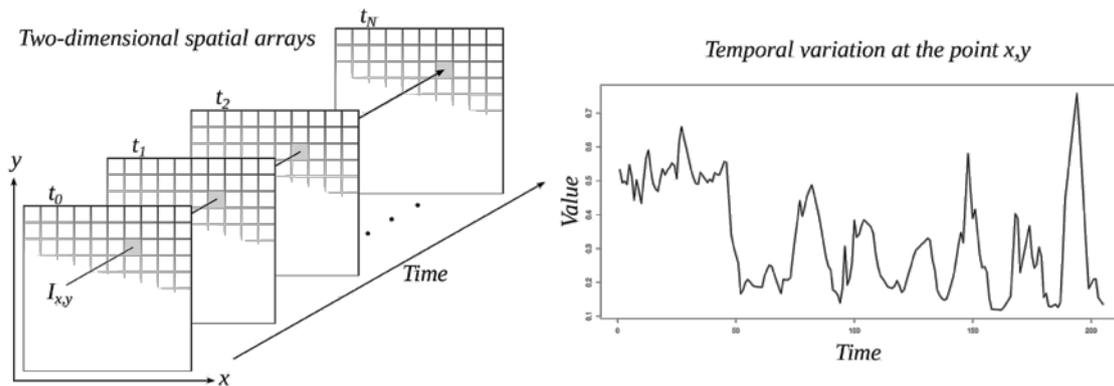


Figure 1. Satellite time series composition.

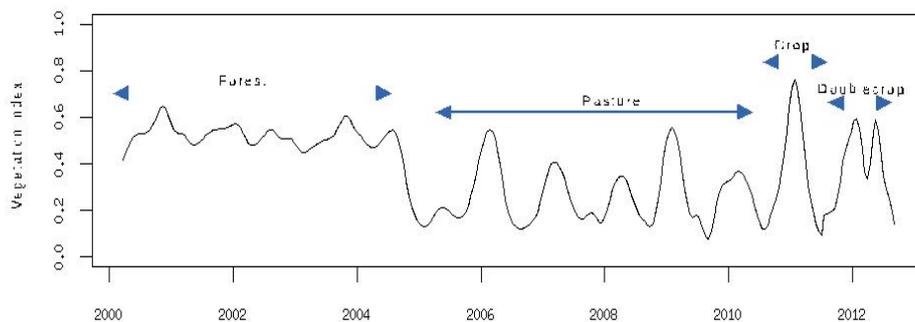


Figure 2. Phenological vegetation cycles reflected in vegetation index time series.

Some studies have shown success in extracting land use/cover information from vegetation time series. Jakubauskas et al., (2002), successfully used 1 km NDVI time series from National Oceanic and Atmospheric Administration (NOAA) - Advanced Very High Resolution Radiometer (AVHRR) to classify corn, wheat, milo and alfalfa. To map rice areas in South and Southeast Asia, Xiao et al., (2005), combine 500 m NDVI, EVI and Land Surface Water Index (LSWI) from Moderate Resolution Imaging Spectroradiometer (MODIS). Sakamoto et al., (2009), used the 250 m MODIS EVI product to analyze expansion of inland aquaculture and triple rice-cropping areas in a coastal area of the Vietnamese Mekong Delta. Conrad et al., (2011) used 250 m MODIS NDVI to differ cotton, rice, winter wheat and rotations with fallow and rice. Wardlow et al., (2007) used 250 m MODIS NDVI and EVI to classify winter wheat and rotations with corn, soybeans, sorghum and fallow.

The cited works evaluate time series in yearly time segments i.e. breaking the sequence into segments of one year, from which they extract phenological parameters (e.g. dates of beginning and end, standard deviation, mean, number of peaks, etc). This segmentation can aid in extraction of important information due to variability of phenological cycles, e.g. in the case of plants that require cycles longer than one year. Furthermore, the weather and/or human interventions can change the dates and other features of phenological events. The classification methods based on parameter extraction do not use all information about the shape of cycles which can help to improve the results. Some techniques of dynamic programming based on the ability to approximately match sequences of values can be used to find patterns in a data stream even considering the time variability, as shown in phenological cycles.

The main contribution of this research was to develop an automatic method based on dynamic programming to yield sequences of land use and land cover maps without time segmentation and using shape information from the entire cycle. We used the classical data mining method Dynamic Time Warping (DTW) to match typical vegetation patterns in long-term EVI2 time series.

Originally, the DTW was successfully applied to automate speech recognition in spite of wide variations in timing and pronunciation (VELICHKO and ZAGORUYKO, 1970; SAKOE and CHIBA, 1978) and in problems of finding patterns in time series data (BERNDT and Clifford, 1994). The DTW algorithm compares two sequences of temporally dependent data, computes the dissimilarity (dtw distance), and finds an optimal alignment between both, under certain restrictions. It also warps the data sequences to match each other (RABINER and JUANG, 1993). Some adaptations in the DTW algorithm have allowed researchers to find all subsequences within a long data stream that are similar to one or more given queries sequences, Figure 3 (MÜLLER, 2007).

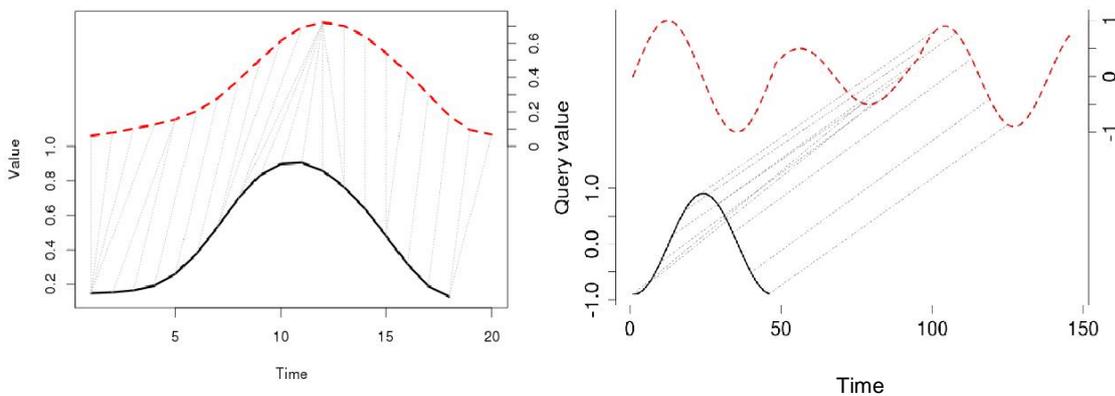


Figure 3. Optimal alignment between data sequences.

Methods

In this study we used the 2-band Enhanced Vegetation Index (EVI2) that enhances the vegetation variation and corrects noise and saturation problems shown in NDIV and EVI (FREITAS et al., 2011). We first download the EVI2 computed by Laboratory of Remote Sensing in Agriculture and Forestry (LAF/INPE) which is based on MODIS MOD13 Q1 product 250 m 16-day composite images. Below, the flowchart summarizes the proposed method, Figure 4.

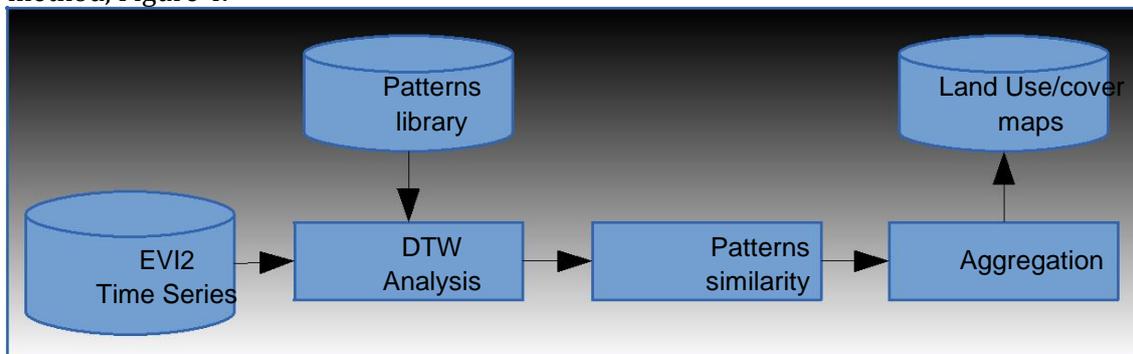


Figure 4. Methodology flowchart.

The second step was build a pattern library. As showed in Figure 5, each vegetation land cover has a related temporal pattern, which was included in the patterns library. The classes considered in this work were: forest, clearcut, pasture, single cropping, double cropping, and fallow (Figure 5). Forest has slight changes in greenness between the wet and dry seasons. Clearcut has a fast decrease in vegetation index value. Pasture outline has higher standard deviation than forest and lower standard deviation than single and double cropping, which differ to each other by the number of annual peaks.

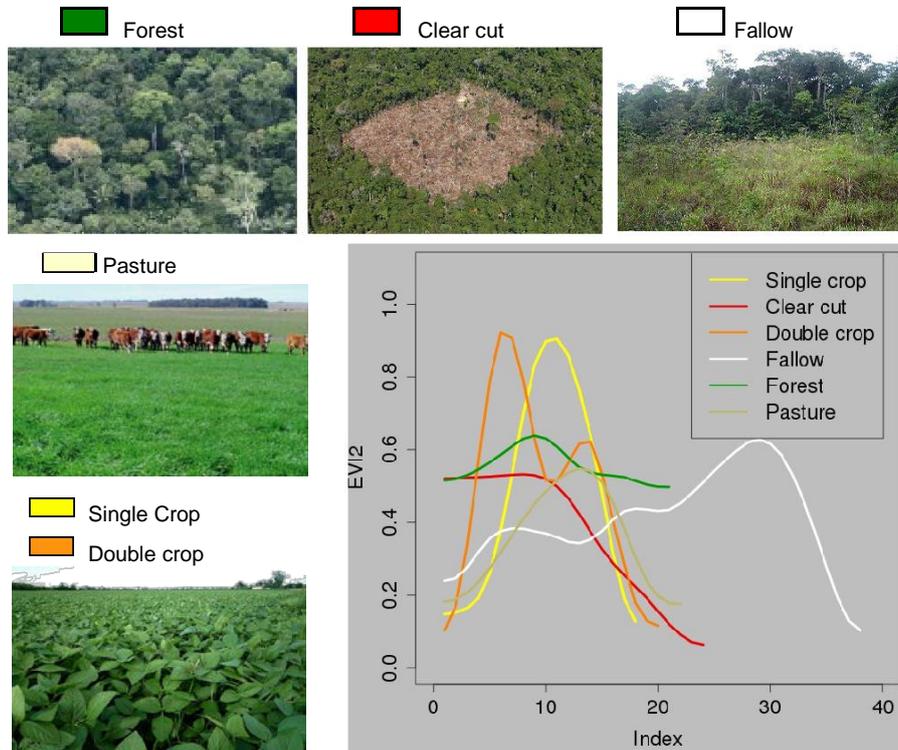


Figure 5. Classes and EVI2 temporal patterns.

Then, we used the DTW algorithm to compute the dissimilarity (dtw distance), and find an optimal alignment between EVI2 time series. The queries to DTW method were all sequences stored in the patterns library. The algorithm returns the DTW distance from all subsequences in the long-term time series to all patterns stored in the pattern library. This way more than one pattern can be aligned to one subsequence, but with different DTW distances that allow to find what is the most similar pattern to one subsequence. To compare DTW classification with other maps, we aggregated the results by agricultural year from August to July. Then, we picked out the pattern with the lowest dtw distance for each agricultural year i.e the pattern with the best alignment during the period was the class used to build the final land use and cover maps.

We performed simulations in a test area of 8.325 ha in Mato Grosso state, Brazil (Figure 6). This state had a lot of land use changes over the last decade and also has the largest deforested area of Amazon rain forest (INPE, 2012). Results were then, compared to the TerraClass maps (CRA/INPE, 2012) for 2008 and 2010 for the following classes: forest, pasture, and cropland. TerraClass does not differentiate single and double cropping but it encompasses four classes of pasture. Therefore, to do the accuracy calculations we made some class aggregation (Table 1). Single and double cropping are combined into cropland, pasture includes the four pasture classes from TerraClass and pasture and fallow classes from the DTW simulations. TerraClass has 30 m spatial resolution,

therefore to compare the maps generated in this work, TerraClass maps were aggregated to 250 m spatial resolution.



Figure 6. Study area Mato Grosso, Brazil.

Table 1. Equivalence among classes to accuracy calculations.

Aggregated classes	TerraClass	DTW Classification
Forest	Forest	Forest
Cropland	Annual agriculture	Single cropping Double cropping
Pasture	Clear pasture Dirty pasture Degraded pasture Pasture with regeneration	Pasture Fallow

The algorithms were written in R language, using the packages: “aRT” that provide an environment to work with spatial data stored in TerraLib database; “dtw” that provide the classical DTW algorithm and some variations; “multicore” that provide optimized functions to parallel computing; and “pastecs” to find local minimums in a data stream. The Geographical Information System (GIS) TerraView was used to compose and visualize maps.

Results and discussions

According to PRODES Project (INPE, 2013), in the studied area the first deforestation event occurred in 2002, that agreed with the clear cut identified by the automatic method developed in this work. The Figure 7 shows the DTW classification, a Landsat image and a EVI2 time series from a sample pixel inside this area. Before 2002 the EVI2 index is greater than 0.5 with low standard deviation, after that, occurred the clear cut and the biomass burning (see Landsat image, Figure 7), EVI2 where then reduced to about 0 (zero). In the following years the EVI2 index shows a biomass regrowth similar to a fallow temporal shape from 2003 to 2004. We observed that fallow is common after clear cut events in the study area.

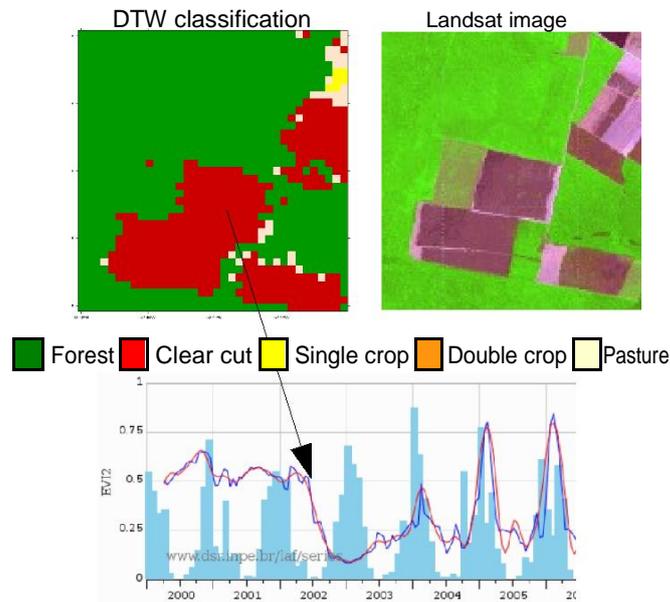


Figure 7. Clear cut detection.

In many places in Mato Grosso, Brazil, the single cropping is being replaced by double cropping. The Figure 8 shows a sample of this transition from 2007 to 2008. The EVI2 index indicates a single cropping from 2005 to 2007 and a switch to double cropping in 2008.

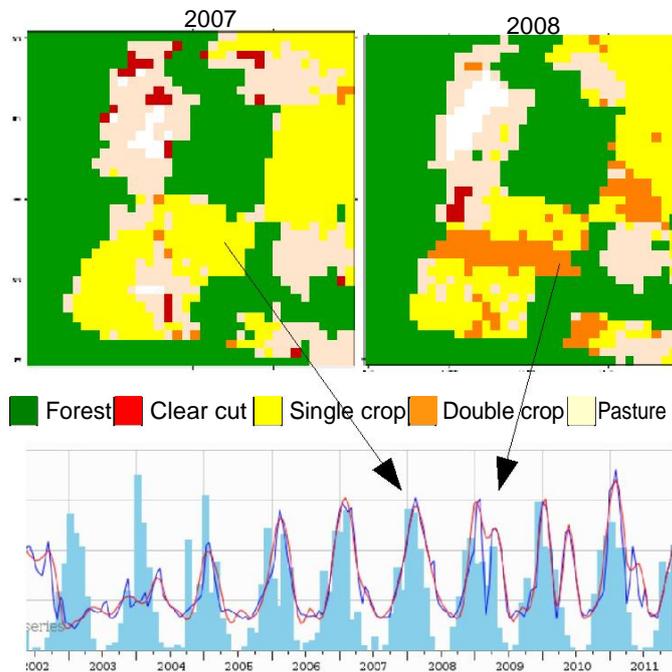


Figure 8. Transition from single cropping to double cropping.

Preliminary tests of DTW classification shows a satisfactory agreement with official TerraClass maps from 2008 and 2010 with global accuracy of 78.2% and 85.0%, and Kappa coefficients of 0.62 and 0.72, for each year. The main classification error occurs

in boundaries between different land covers (Figure 9) i.e. where MODIS pixels contain a mixture in surface reflectance. Furthermore we can see a land cover class appearing in some pixels isolated within large homogenous areas that are probably mistakes in the automatic classification.

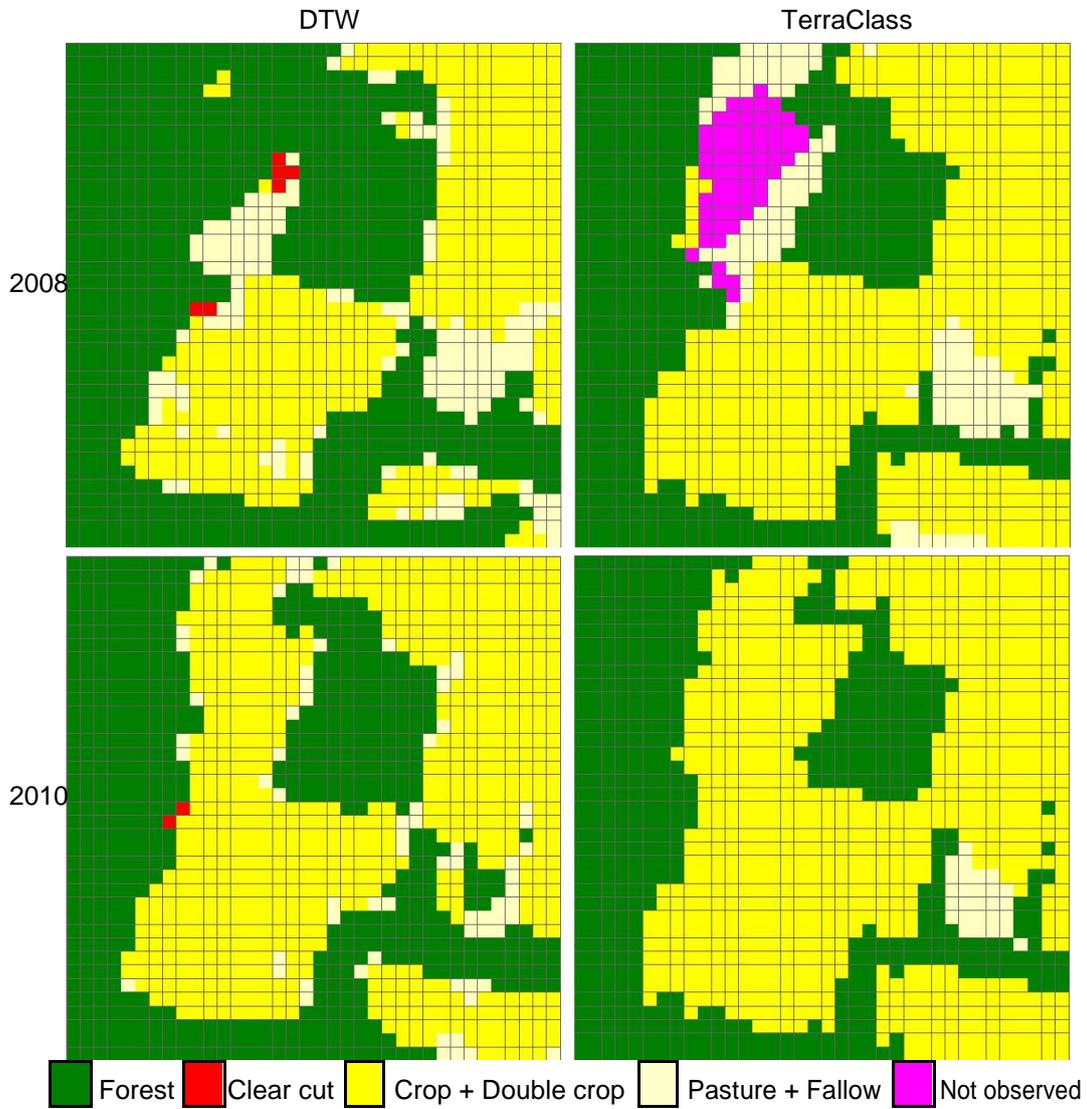


Figure 9. TerraClass and DTW maps from 2008 and 2010.

The maps yielded by DTW also show some impossible transitions, such as the transition to forest after a clearcut event: *forest*→*clear cut*→*forest*. A sample of this mistake can be visualized from 2004 to 2006 in the area pointed in Figure 10. The EVI2 in that area shows a clearcut pattern from 2005 followed by a biomass regrowth. The EVI2 value reaches about 0.5 which is about the value of the EVI in the forest pattern (Figure 5). This leads to classify the pixel as forest in 2006 while it is in fact a different type of vegetation.

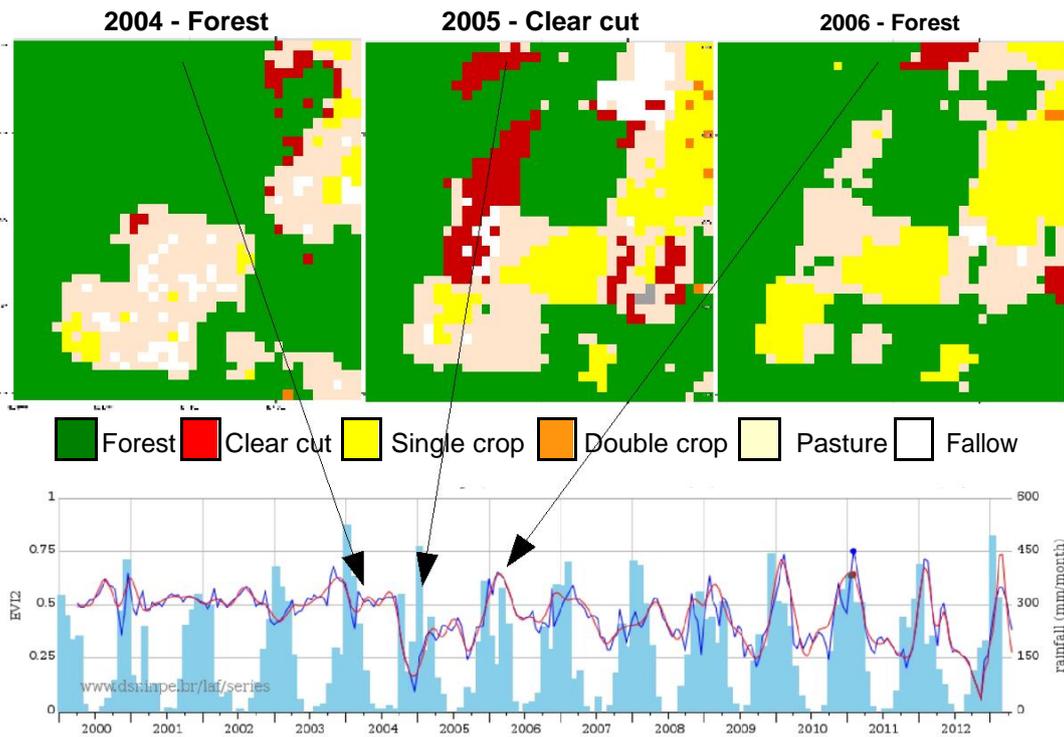


Figure 10. Transitions from clear cut to forest.

Conclusions

High levels of uncertainty surrounding past, current and future trajectories of land use, in particular in tropical forest regions, are seen as a major stumbling block in terms of reducing deforestation and the associated carbon emissions and increasing food and energy security. Currently, Brazilian land use data comes from the national agricultural census and land cover data comes from global data sets with sparse temporal coverage. This no longer meets the needs of the earth system modeling community. Long-term satellite image datasets with high temporal frequency yield a sequence of data points in a time series that can be used to detect and monitor land use and land cover changes. The vegetation phenological cycles are reflected in the satellite time series, allowing the classification of land cover types in time segments, leading to the extraction of land use change.

This study has resulted in a novel approach to create spatial and temporal datasets of land use through a data mining technique known as Exploratory Dynamic Time Warping (DTW). DTW results show significant potential to detect land use and land cover changes, which is useful not only to provide land use and land cover maps but also to understand land cover changes potentially leading to the establishment of more informed policies. The methodology was able to accurately identify forest, clear cut events, pasture, single and double cropping.

Validation tests were made in the areas of Mato Grosso state, Brazil. The preliminary results for the proposed methods are promising when compared with the official

TerraClass land use maps in the Amazon Biome, finding 78.2% and 85.0% overall accuracy for 2008 and 2010, respectively. This is not only an improvement in terms of overall accuracy, but this method provides us with a continuous measure of land use change (not limited to the periodic campaigns of official data). To improve the results of the DTW classifications, it is essential to include post- processing steps with rules for land use and cover transitions and spatial filtering, in addition to improving the algorithm for the detection of fallow fields.

The findings of this approach have large potential for radically changing the way land use is monitored currently, having wide ranging policy implications. This method offers a methodology for tapping into the ever expanding archive of terrestrial satellite imagery, offering potentially a huge return on investment.

References

- Araújo, R. & Léna, P. (2011). **Desenvolvimento Sustentável e Sociedade na Amazônia**. Museu Paraense Emílio Goeldi.
- Berndt D.; Clifford J. Using Dynamic Time Warping to Find Patterns in Time Series. **In Proc. of KDD Workshop**, 1994.
- Conrad, C.; Colditz, R. R.; Dech, S.; Klein, D.; Vlek, P. L. G. (2011). Temporal segmentation of MODIS time series for improving crop classification in central Asian irrigation systems. **International Journal of Remote Sensing**, v. 32, n. 23, p. 8763-8778.
- CRA/INPE (2012). **Projeto TerraClass**. Available at <http://www.inpe.br/cra>. Accessed in 01 July 2013.
- Freitas, R. M. de; Arai, E.; Adami, M.; Ferreira, A. S.; Sato, F. Y.; Shimabukuro, Y. E.; Rosa, R. R.; Anderson, L. O.; Rudorff, B. F. T. (2011). Virtual laboratory of remote sensing time series: visualization of MODIS EVI2 data set over south America. **Journal of Computational Interdisciplinary Sciences**, v. 2, n. 1, p. 57-68.
- IBGE (2006). **Censo Agropecuário 2006**, Instituto Brasileiro de Geografia e Estatística. Available at <http://ibge.gov.br/>. Accessed in 20 September 2013.
- INPE (2012). **Monitoramento da Floresta Amazônica Brasileira por Satélite** (Monitoring the Brazilian Amazon Forest by Satellite). Available at <http://www.obt.inpe.br/prodes>. Accessed in 20 September 2013.
- INPE (2013). **PRODES Amazon deforestation database**, São Jose dos Campos, INPE. Available online at: <http://www.obt.inpe.br/prodes>. Accessed in 10 July 2013.
- Jakubausks, M. E.; Legates, D. R.; Kastens, J. H. (2002). Crop identification using harmonic analysis of time -series AVHRR NDVI data. **Computers and Electronics in Agriculture**, v. 37, n. 1â3, p. 127-139.
- Jiang, Z.; Huete, A. R.; Didan, K.; Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. **Remote Sensing of Environment**, v. 112, n. 10, p. 3833-3845.
- Justice, C.; Vermote, E.; Townshend, J.; Defries, R.; Roy, D.; Hall, D.; Salomonson, V.; Privette, J.; Riggs, G.; Strahler, A.; Lucht, W.; Myneni, R.; Knyazikhin, Y.; Running, S.; Nemani, R.; Wan, Z.; Huete, A.; van Leeuwen, W.; Wolfe, R.; Giglio, L.; Muller, J.; Lewis, P. & Barnsley, M. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research Geoscience and Remote Sensing, **IEEE Transactions on**, v. 36, n. 4, 1228-1249.
- Müller, M. (2007). **Information retrieval for music and motion**. Springer.

Rabiner, L.; Juang, B. (1993). **Fundamentals of speech recognition**. Prentice-Hal International, Inc.

Rouse, J. W., Haas, R. H., ScheU, J. A., and Deering, D. W. (1973). **Monitoring vegetation systems in the great plains with ERTS**, Third ERTS Symposium, NASA SP-351 1:309-317.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C. (1974). **Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation**. NASA/GSFC Type III Final Report, Greenbelt, Md. 371 pp.c.

Sakamoto, T.; Van, P. C.; Kotera; Nguyen, K. D. & Yokozawa, M. (2009). Analysis of rapid expansion of inland aquaculture and triple rice-cropping areas in a coastal area of the Vietnamese Mekong Delta using MODIS time-series imagery. **Landscape and Urban Planning**, v. 92, n. 1, 34-46.

Sakoe, H.; Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. **Acoustics, Speech and Signal Processing, IEEE Transactions on**, v. 26, n. 1, p. 43-49.

Velichko, V.; Zagoruyko, N. (1970). Automatic recognition of 200 words. **International Journal of Man-Machine Studies**, v. 2, n. 3, p. 223-234.

Wardlow, B.; Egbert, S.; Kastens, J. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. **Remote Sensing of Environment**, v. 108, n. 3, p. 290-310.

Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frohling, S.; Li, C.; Salas, W.; Moore, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. **Remote Sensing of Environment**, v. 95, p. 480-492.

Appendix A. Main functions used to time series analysis

R packages dependencies: aRT; dtw; multicore; pastecs.

Kthbacktrack: Backtrack the steps taken. This function was adapted from dtw package. Functions called in this function are described in dtw package documentation.

```
kthbacktrack2 <- function(gcm) {
  dir<-gcm$stepPattern;
  npat <- attr(dir,"npat");
  n <- nrow(gcm$costMatrix);
  m <- ncol(gcm$costMatrix);
  i <- n;
  j <- gcm$jmin
  ## drop rows with (0,0) deltas
  nullrows <- dir[,2]==0 & dir[,3] ==0 ;
  tmp <- dir[!nullrows,];
  ## Pre-compute steps
  stepsCache <- list();
  for(k in 1:npat) {
    stepsCache[[k]] <- .extractpattern(tmp,k);
  }
  ## mapping lists
  ii<-c(i);
  jj<-c(j);
  repeat {
    ## cross fingers for termination
    if(i==1){# && j==1) {
      break;
    }
    ## direction taken
    s<-gcm$directionMatrix[i,j];
    if(is.na(s)) {
      break;
    }
    ## undo the steps
    steps<-stepsCache[[s]];
    ns<-nrow(steps);
    ## In some rare cases (eg symmetricP0), ns will be 1
    ## R indexing rules make k==0 a no-op anyway
    for(k in 1:ns) {
      ## take note of current cell, prepending to mapping lists
      if(i-steps[k,1] > 0) { # Modified from original function
        ii <- c(i-steps[k,1],ii); # Modified from original function
        jj <- c(j-steps[k,2],jj); # Modified from original function
      }
      # Modified from original function
      ## All sub-steps are visited & appended; we have dropped (0,0) deltas
    }
    i <- ii[1]#i-steps[ns,1]; # Modified from original function
    j <- jj[1]#j-steps[ns,2]; # Modified from original function
  }
  out<-list();
  out$index1<-ii;
  out$index2<-jj;
  return(out);
}
```

RecruitLocalmins: Find the first point of subsequences candidates to align a pattern.

```
RecruitLocalmins <- function(x,span,threshold){
  # x - a vector with last line DTW CostMatrix.
  # span - Neighborhood of minimum point in the CostMatrix. Default = 10.
```

```

# threshold – maximum acceptable DTW distance to similarity. Default: threshold = 2.
# Find turning points
turnp <- turnpoints(x)$tpo
# Find local minimum
dfturnp <- diff(x[turnp])
localmins <- turnp[dfturnp > 0]
index <- array(NA)
k <- 1
repeat{
  kmin <- localmins[which.min(x[localmins])]
  if ( (x[kmin] > threshold) || (k > length(localmins)) ) break
  index[k] <- kmin
  k <- k + 1
  if ( (kmin-10) < 1 ) x[1:(kmin+10)] <- 9999
  else x[(kmin-10):(kmin+10)] <- 9999
}
if ( (x[length(x)] - x[length(x)-1]) <= 0 ) && (x[length(x)] < threshold) )
  index[k] <- length(x)
return(index)
}

```

FindPattern: Function to compute DTW distance and align all subsequences within long-term data stream, that are similar to the patterns in the library.

```

FindPattern <- function(TS.df, TS.dates, threshold = 2, span = 10){
  # Parameters:
  # TS.df: a data frame with all time series  $Y = \{y(t)_1, y(t)_2, \dots, y(t)_n\}$   $t=1,2,\dots,m$ 
  # where n and m are amount of time series and amount of measures, respectively.
  # TS.dates: a vector with the dates for each measures.
  # threshold – maximum acceptable DTW distance to similarity. Default: threshold = 2.
  # span – Neighborhood of minimum point in the CostMatrix. Default: span = 10.
  ### Get patterns/signatures
  TPFilelist <- dir(TPPATH,pattern=".csv")
  TPatterns <- mclapply(paste(TPPATH,TPFilelist,sep=""),
    read.table, as.is=TRUE, header=TRUE, sep=";")
  # Set patterns/signatures names with file names
  names(TPatterns) <- unlist(strsplit(TPFilelist, split="\\.csv"))
  cat("\n Finding for",length(TPFilelist),
    "patterns in",dim(TS.df)[1],"time series\n Please wait...")
  # Loop for all time series
  DTWDistance <- mclapply( 1:(dim(TS.df)[1]), function(j) {
    # Get time serie j
    y <- as.numeric(TS.df[j,])
    # Loop for all temporal signatures (X)
    out.patterns.distance <- mclapply(seq_along(TPatterns), function(i){
      # Get pattern i
      x <- as.numeric(TPatterns[[i]][,2])
      out <- list()
      # Algorithm to compute all subsequences within
      # the Time Series Y similar to Pattern/Signature X
      # Step 1. Compute Accumulated Cost Matrix (Sequence Y against
      # pattern/signature X with DTW Package)
      alignment <- dtw(x,y,step=rabinerJuangStepPattern(6,"c"),
        keep=TRUE,open.begin=TRUE,open.end=TRUE,
        window.type=sakoeChibaWindow, window.size=5)
      # Step 2. Determine the distances for k-th paths
      d <- alignment$costMatrix[alignment$N,1:alignment$M]
      [is.na(alignment$costMatrix[alignment$N,1:alignment$M])]
      Nna <- length(alignment$costMatrix[alignment$N,1:alignment$M]
        [is.na(alignment$costMatrix[alignment$N,1:alignment$M])])
      # Step 3. Determine ranked list of indexes to minimal
      # distances within a threshold (Index for last point of a
      # Pattern on the Time Series)
    },

```

```

index_b <- RecruitLocalmins(d, span, threshold)
if ( is.na(index_b[1]) ) return(NULL)
index_b <- index_b + Nna
# Step 4. Compute the starting indexes
# Index for first point of a Pattern on the Time Serie
index_a <- unlist(lapply(seq_along(index_b), function(k){
  alignment$jmin <- index_b[k]
  mapping <- kthbacktrack(alignment)
  return(mapping$index2[1])
})) # End index_a loop
# Get distance (Paths cost)
distance <- unlist(lapply(index_b, function(b){
  return(d[b-Nna])
}))
out$a <- index_a
out$b <- index_b
out$distance <- distance
return(out)
})# End all patterns
names(out.patterns.distance) <- names(TPatterns)
return(out.patterns.distance)
})# End all patterns
return(DTWDistance)
}#End function

```

GetTS: function to Download EVI2 data from LAF/INPE server. Given a pair of coordinates the function returns a list of northeast and southwest coordinates, Dates, Raw EVI2 and Wavelet Fitted EVI2.

```

GetTS <- function (lat, lon, l=1){
  # Parameters:
  # lat: latitude
  # lon: longitude
  ### returns a list of:
  ### northeast and southwest coordinates,
  ### Dates, Raw EVI2 and Wavelet
  ### Fitted EVI2.
  out <- list(box=NULL,raw=NULL,filtered=NULL);
  count = 0;
  conAttempts = 10;
  address <- paste("http://www.dsr.inpe.br/laf/download/getSeries.php?lat=",lat,"&lng=",lon,sep="")

  while( count <= conAttempts )
  {
    count <- count + 1;
    doc <- xmlRoot(xmlTreeParse(address));
    rootName <- xmlSApply(doc, xmlName);
    if( class(doc)[1]!="XMLNode" || length(rootName) <= 1 )
    {
      if( count > conAttempts )
      {
        cat("\n ",conAttempts," connect attempts to dsr.inpe.br server failed..");
        cat("\n Can not getting EVI2 time series for lat:",lat," lon:",lon);
        cat("\n Try to connect again (y/n)? ");
        flag <- scan(what="character", nmax=1);
        if( flag == "y" || flag == "Y")
        {
          count <- 0;
        }else{
          return(out);
        }
      }else{
        errorValue <- xmlValue(doc);
        warning(errorValue, " lat: ",lat, " lon: ",lon, immediate. = TRUE, call = FALSE);
      }
    }
  }
}

```

```

warning("\n",count," : Unable to access time series from
http://www.dsr.inpe.br/laf/download/getSeries.php?lat=",lat,"&lng=",lon, immediate. = TRUE, call = FALSE);
cat("\nTrying to reconnect dsr.inpe.br server...");
}
}else{
cat("\n",l," : Successful time series lat:",lat," lon:", lon, sep="");
count <- conAttempts + 1;
}
}#End While
# Get box coordinate
boxCoord <- xmlSApply(doc[["box"]], function(b){
xmlSApply(b, xmlValue);
});
# Get raw data
rawEVI2 <- xmlSApply(doc[["data"]], function(b){
xmlSApply(b[["raw"]], xmlValue);
});
# Get filtered data
filEVI2 <- xmlSApply(doc[["data"]], function(b){
xmlSApply(b[["filtered"]], xmlValue);
});
# Get date data
date <- xmlSApply(doc[["data"]], function(b){
xmlSApply(b[["date"]], xmlValue);
});
Coords <- list(northeast=c(lat=as.numeric(boxCoord[1]),lon=as.numeric(boxCoord[2])),
southwest=c(lat=as.numeric(boxCoord[3]),lon=as.numeric(boxCoord[4])))
out <- list(box=Coords, date=as.Date(date), raw=as.numeric(rawEVI2), filtered=as.numeric(filEVI2));
return(out);
}

```

GetListTS: function to Download EVI2 data from LAF/INPE server for a list of points. Given a list of coordinates the function returns for each pixel a list of: northeast and southwest coordinates, Dates, Raw EVI2 and Wavelet Fitted EVI2.

```

#
# R function to Download EVI2 data from LAF/INPE server for a list of points
#
GetListTS <- function(PixelCoord, BASENAME="pto"){
# Parameters:
# lat: latitude
# lon: longitude
### returns for each a list of:
### northeast and southwest coordinates,
### Dates, Raw EVI2 and Wavelet
### Fitted EVI2.
cat("\nGetting EVI2 time series from LAF/INPE\nPlease wait...\n")
t1 <- Sys.time()
date.df <- data.frame()
raw.df <- data.frame()
filtered.df <- data.frame()
Coord.df <- data.frame(northeast_lon=NA,northeast_lat=NA,
southwest_lon=NA,southwest_lat=NA)
SampleNames <- array();

for(i in 1:length(PixelCoord))
{
pto <- GetTS(PixelCoord[[i]][['lat']],PixelCoord[[i]][['lon']],i)
SampleNames[i] <- paste(BASENAME,i,sep="")
Coord.df[i,] <- cbind(pto$box$northeast["lon"],
pto$box$northeast["lat"],
pto$box$southwest["lon"],
pto$box$southwest["lat"]
)
}
}

```

```

        date.df <- rbind(date.df, format(as.Date(pto$date, "%Y-%m-%d"), "%Y-%m-%d"))
        raw.df <- rbind(raw.df, pto$raw)
        filtered.df <- rbind(filtered.df, pto$filtered)
    }
    date.df <- format(as.Date(names(date.df), "X.%Y.%m.%d."), "%Y-%m-%d")
    vec_names <- format(as.Date(date.df, "%Y-%m-%d"), "%Y_%m_%d")
    names(date.df) <- as.character(vec_names);
    names(raw.df) <- as.character(vec_names);
    names(filtered.df) <- as.character(vec_names);
    t2 <- Sys.time();
    cat(i,"time series downloaded in ",as.double(difftime(t2,t1,tz,units="secs")), "secs\n")
    return( list(coord=Coord.df, date=date.df, raw=raw.df, filtered=filtered.df) )
}

```