

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327085565>

Big earth observation time series analysis for monitoring Brazilian agriculture

Article in ISPRS Journal of Photogrammetry and Remote Sensing · August 2018

DOI: 10.1016/j.isprsjprs.2018.08.007

CITATIONS

3

READS

440

12 authors, including:



Michelle C. A. Picoli

National Institute for Space Research, Brazil

44 PUBLICATIONS 127 CITATIONS

[SEE PROFILE](#)



Ieda Sanches

National Institute for Space Research, Brazil

57 PUBLICATIONS 296 CITATIONS

[SEE PROFILE](#)



Rolf Simoes

National Institute for Space Research, Brazil

9 PUBLICATIONS 5 CITATIONS

[SEE PROFILE](#)



Alexandre Xavier Ywata Carvalho

Caixa Econômica Federal

75 PUBLICATIONS 358 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Rede Amazônia Sustentável [View project](#)

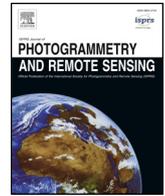


e-sensing: big earth observation data analytics for land use and land cover change information (www.esensing.org) [View project](#)



Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Big earth observation time series analysis for monitoring Brazilian agriculture

Michelle Cristina Araujo Picoli^{a,*}, Gilberto Camara^a, Ieda Sanches^a, Rolf Simões^a, Alexandre Carvalho^b, Adeline Maciel^a, Alexandre Coutinho^c, Julio Esquerdo^c, João Antunes^c, Rodrigo Anzolin Begotti^a, Damien Arvor^d, Claudio Almeida^a

^a National Institute for Space Research (INPE), São José dos Campos, Brazil

^b Institute of Applied Economic Research (IPEA), Brasília, Brazil

^c Embrapa Agricultural Informatics, Brazilian Agricultural Research Corporation (Embrapa), Campinas, Brazil

^d Université de Rennes, CNRS, LETG - UMR 6554, F-35000 Rennes, France

ARTICLE INFO

Keywords:

Big earth observation data
Land use science
Satellite image time series
Crop expansion
Brazilian Amazonia biome
Brazilian Cerrado biome
Tropical deforestation

ABSTRACT

This paper presents innovative methods for using satellite image time series to produce land use and land cover classification over large areas in Brazil from 2001 to 2016. We used Moderate Resolution Imaging Spectroradiometer (MODIS) time series data to classify natural and human-transformed land areas in the state of Mato Grosso, Brazil's agricultural frontier. Our hypothesis is that building high-dimensional spaces using all values of the time series, coupled with advanced statistical learning methods, is a robust and efficient approach for land cover classification of large data sets. We used the full depth of satellite image time series to create large dimensional spaces for statistical classification. The data consist of MODIS MOD13Q1 time series with 23 samples per year per pixel, and 4 bands (Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), near-infrared (nir) and mid-infrared (mir)). By taking a series of labelled time series, we fed a 92-dimensional attribute space into a support vector machine model. Using a 5-fold cross validation, we obtained an overall accuracy of 94% for discriminating among nine land cover classes: forest, cerrado, pasture, soybean-fallow, fallow-cotton, soybean-cotton, soybean-corn, soybean-millet, and soybean-sunflower. Producer and user accuracies for all classes were close to or better than 90%. The results highlight important trends in agricultural intensification in Mato Grosso. Double crop systems are now the most common production system in the state, sparing land from agricultural production. Pasture expansion and intensification has been less studied than crop expansion, although it has a stronger impact on deforestation and greenhouse gas (GHG) emissions. Our results point to a significant increase in the stocking rate in Mato Grosso and to the possible abandonment of pasture areas opened in the state's frontier. The detailed land cover maps contribute to an assessment of the interplay between production and protection in the Brazilian Amazon and Cerrado biomes.

1. Introduction

Since the 1980s, Brazil has become one of the world's largest agricultural exporters. Brazil is the world's largest producer of sugarcane, coffee, and orange juice, and the second largest producer of soybeans, beef, and chicken meat. Brazilian crop and livestock producers face a major challenge. While producing food for a growing world demand, Brazilian agriculture has to contribute to the country's commitments to reduce its deforestation rates and greenhouse gas (GHG) emissions

(Garnett, 2015). In the coming decades, Brazil needs to equalize economic gains and sustainable agricultural practices. A way to achieve this aim consists of (1) promoting new agricultural practices to increase productivity and (2) implementing efficient command-and-control and market-oriented policies to control deforestation (Nepstad et al., 2014; Arvor et al., 2016).

Brazil's government has acted to reduce deforestation and the resulting emissions. Combining rapid assessment of new forest cuts with strong law enforcement, Brazil cut tropical deforestation by 80% from

* Corresponding author.

E-mail addresses: michelle.picoli@inpe.br (M.C.A. Picoli), gilberto.camara@inpe.br (G. Camara), ieda.sanches@inpe.br (I. Sanches), rolf.simoese@inpe.br (R. Simões), alexandre.ywata@ipea.gov.br (A. Carvalho), adeline.maciela@inpe.br (A. Maciel), alex.coutinho@embrapa.br (A. Coutinho), julio.esquerdo@embrapa.br (J. Esquerdo), joao.antunes@embrapa.br (J. Antunes), rodrigo.begotti@inpe.br (R.A. Begotti), claudio.almeida@inpe.br (C. Almeida).

<https://doi.org/10.1016/j.isprsjprs.2018.08.007>

Received 2 October 2017; Received in revised form 7 August 2018; Accepted 7 August 2018

0924-2716/ © 2018 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

2005 to 2010 (Assunção et al., 2015). These initiatives were complemented by actions in the private sector, such as the Soy Moratorium. The moratorium is an agreement signed by the major soybean traders pledging not to buy soy grown in Amazon forest areas cleared after July 2008. During 2004 and 2005, 30% of the soybean expansion in this region occurred through deforestation. In 2014, only 1% of the new soy expansion in the Amazon biome resulted from the direct conversion of forest to cropland (Gibbs et al., 2015). Despite these advances, the environmental impacts of crop production and cattle-raising in Brazil's Amazonia and Cerrado biomes continue to raise concerns (Nepstad et al., 2014). To develop adequate public policies that balance production with protection, Brazil needs comprehensive information on land change dynamics.

Previous studies of Brazilian agricultural dynamics have focused on the state of Mato Grosso, one of the world's fast expanding agricultural frontiers. Spera et al. (2014) used satellite remote sensing to examine patterns of cropland expansion in Mato Grosso from 2001 to 2011. They used the Enhanced Vegetation Index (EVI) from the Moderate-Resolution Imaging Spectroradiometer (MODIS) time series, coupled with a decision-tree algorithm. Data from crop-specific growing season lengths and maximum EVI thresholds were used to classify large-scale croplands into five classes: soy, cotton, soy-corn, soy-cotton, and irrigated crops. Their paper describes how crop expansion depends on land attributes such as soil, climate, and topography. The authors found that most of the suitable areas for cropland expansion in Mato Grosso had been occupied by 2006. As a consequence, farmers increased double crop systems to make up for the scarcity of high quality agricultural land. Since their paper deals with how land quality affects farmers' decision-making, it does not include accuracy assessments for the classification results.

Arvor et al. (2011) used MODIS EVI time series to identify five crop classes: soybean, corn, cotton, soybean-corn, and soybean-cotton. They assume that corn is only planted in consortium with soybeans. The authors collected ground data sets at 50 farms in Mato Grosso in 2005–2006 and 2006–2007. The study used a two-step classification method, first creating a cropland mask and then discriminating the crop varieties of interest inside the mask. To create the mask, they assumed that crop EVI profiles are identifiable as having one of two cycles with high maximum values and low minimum values. To classify crop types inside the cropland mask, they used the Jeffries-Matsushita distance to rank the 23 dates of each MODIS EVI series. The authors used the best subset of these dates as inputs to a supervised classifier, followed by post-processing using segmentation to produce more homogeneous results. Their reported accuracy is 85% for the agricultural mask and 74% for the crop classification, using validation data not included in the training set.

To describe the spatial dynamics of crop production in Mato Grosso from 2001 to 2014, Kastens et al. (2017) used MODIS Normalized

Difference Vegetation Index (NDVI) time series. They take ground reference data from 2009 to 2016 to train and validate a random forest classification model. The reported accuracy was 79% for distinguishing five crop classes (soybean-fallow, fallow-cotton, soybean-cotton, soybean-crop, and pasture/cerrado). The soybean-crop class includes corn, millet, sorghum, and sunflower, which the authors stated they could not distinguish well.

Studies covering the whole Amazonia biome focus on deforestation and its relation to pasturelands. The PRODES (Programa de Cálculo do Desflorestamento da Amazônia) developed by Brazil's National Institute for Space Research (INPE) maps clear cuts in the Amazon forest yearly, producing a forest/non-forest mask (INPE, 2014). Hansen et al. (2013) produced global maps of forest-cover change using LANDSAT-class data. Parente et al. (2017) present maps of pastureland areas in Brazil using LANDSAT-8 images. INPE, together with the Brazilian Agriculture Research Corporation (EMBRAPA), produced TerraClass, a map of land cover change in the Amazon biome (Almeida et al., 2016). TerraClass produces a cropland mask, using Landsat-5 Thematic Mapper (TM) and MODIS data, and does not distinguish between different crops. These efforts are relevant and produce important data sets, but none provides a complete assessment of overall land cover change and its relation to different crop systems.

There are no previous studies in the literature that map both the dynamics of crop expansion and the land changes due to pasture expansion in Brazil's agricultural frontiers. To address this challenge, we have developed new methods to produce consistent multi-year maps of the different types of land cover in Brazil. These maps provide information on crop production systems and pasture expansion into natural vegetation. The results enable an informed assessment of the interplay between production and protection in the Brazilian Amazon and Cerrado biomes.

This work presents innovative methods for using satellite image time series to produce land use and land cover classification over large areas in Brazil from 2001 to 2016. We use the full depth of the MODIS time series data to classify natural and human-transformed land areas in the state of Mato Grosso, Brazil's agricultural frontier. Our hypothesis is that building high-dimensional spaces using all values of the time series, coupled with advanced statistical learning methods, is a robust and efficient approach for land cover classification of large data sets.

2. Background

2.1. Overview

Since remote sensing satellites revisit the same place, we can calibrate their images so that measures of the same place at different times are comparable (Fig. 1(a)). These observations can be organized so that each measure from the sensor maps to a three-dimensional array in

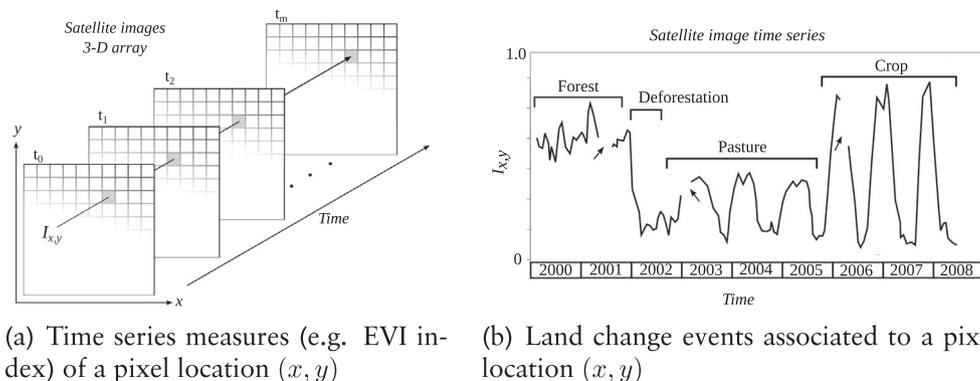


Fig. 1. A 3-dimensional array of satellite data (a) and events describing change at a particular location (b).

Source: Maus et al. (2016).

space-time. From a data analysis perspective, each pixel location (x, y) at consecutive times, t_1, \dots, t_m , makes up a satellite image time series (SITS), such as the one in Fig. 1(b). From these time series, we can extract land-use and land-cover change information. In Fig. 1(b), after the forest was cut in 2002, the area was used for cattle raising (pasture) for three years, during 2002–2008, then turned into cropland.

Using time series derived from big Earth Observation (EO) data sets is one of the leading research trends in Land Use Science (Verbesselt et al., 2010; Kennedy et al., 2014; Pasquarella et al., 2016). Multiyear time series of land surface attributes allow a broader view of land cover change. Time series of remote sensing data show that land cover changes do not always occur in a progressive and gradual way; they may highlight periods of rapid and abrupt change followed by quick recovery (Lambin and Helmut, 2006). Applications of SITS include mapping for detecting forest disturbance (Kennedy et al., 2010), ecological dynamics (Pasquarella et al., 2016), agricultural intensification (Galford et al., 2008), and phenological change detection (Verbesselt et al., 2010).

One of the more promising uses of satellite time series is land use and land cover classification and change detection. The growing demand for agricultural (crops, livestock) and natural (timber, ore) resources has caused major environmental impacts, especially in the tropics. From 1980 to 2000, 55% of the new agricultural land in the tropics came from intact forests, and another 28% came from disturbed forests (Gibbs et al., 2010). Algorithms for land cover classification include those using SITS such as BFAST (Verbesselt et al., 2010) and TIMESAT (Jönsson and Eklundh, 2004) and methods based on Dynamic Time Warping (DTW) (Petitjean et al., 2012; Maus et al., 2016). Despite these advances, classification of satellite time series in large areas remains a challenging task (Pasquarella et al., 2016).

Most studies on the use of SITS for land cover classification published in the literature use variations of classical remote sensing image classification methods. In their review of land use and land cover classification using SITS, Gomez et al. (2016) highlight 12 papers. These works are all similar. For multiyear studies, researchers first derived “best-fit” yearly composites and then classified each composite image. An example is the work by Hansen et al. (2013) who processed 650,000 LANDSAT images from 2000 to 2010 to produce maps of global forest loss. Their method classifies each two-dimensional yearly image composite one by one. A pixel-based classification algorithm processed each image to detect forest cover. Comparing the results for 2000 and 2010, the authors produced an account of global forest loss during the 2000–2010 decade. Camara et al. (2016) denote these works as taking a *space-first, time-later* approach.

Space-first, time-later methods do not use the full potential of remote sensing time series. The benefits of SITS increase when the temporal resolution of the big data set captures the most important changes. In these cases, the temporal autocorrelation of the data will be stronger than the spatial autocorrelation. Given data with adequate repeatability, a pixel is more related to its temporal neighbours than to its spatial ones. In these cases, *time-first, space-later* methods may lead to better results than the *space-first, time-later* approach (Camara et al., 2016).

2.2. Combining SITS with statistical learning methods

This work combines SITS with statistical learning. In a broad sense, statistical learning refers to a class of algorithms used for classification and regression analysis (Hastie et al., 2009). These methods include linear and quadratic discrimination analysis, support vector machines, random forests, and neural networks. In a typical classification problem, we have measures that capture class attributes. Based on these measures, referred to as training data, the task is to select a predictive model that allows us to infer the classes of a larger data set.

There has been much recent interest in using classifiers such as support vector machines (Mountrakis et al., 2011) and random forests

(Belgiu and Dragut, 2016). Most often, researchers use a *space-first, time-later* approach, in which the dimension of the decision space is limited to the number of spectral bands or their transformations. Sometimes, the decision space is extended with temporal attributes. To do this, researchers filter the raw data to get smoother time series (Brown et al., 2013; Kastens et al., 2017). Then, using software such as TIMESAT (Jönsson and Eklundh, 2004), they derive a small set of phenological parameters from vegetation indexes, like the beginning, peak, and length of the growing season (Estel et al., 2015; Pelletier et al., 2016). These approaches do not use the power of advanced statistical learning techniques to work on high-dimensional spaces with big training data sets (James et al., 2013). They have one thing in common: raw time series data is considered too noisy to be used directly. This questions the impact of the noise removal and homogenization steps since it may reduce the information present in the SITS.

An alternative approach, proposed in this paper, is to use the full depth of SITS to create larger dimensional spaces. We tested different methods of extracting attributes from time series data, including those reported by Maus et al. (2016), Pelletier et al. (2016), and Kastens et al. (2017). Our conclusion is that part of the information in raw time series is lost after filtering or statistical approximation. By choosing a statistical classifier which is robust with respect to noise, one should be able to achieve better results than using the current approaches. Thus, the method we developed has a deceptive simplicity: *use all the data available in the time series samples*. The idea is to have as many temporal attributes as possible, increasing the dimension of the classification space. In this work, we used the MODIS MOD13Q1 product with 23 samples per year per pixel and 4 bands (NVDI, EVI, near-infrared (nir) and mid-infrared (mir)). By taking a series of labelled time series, we fed a 92-dimensional attribute space into the statistical inference model.

Table 1 shows a performance comparison of the classifiers for the state of Mato Grosso data set. This assessment was carried out using cross-validation to estimate the expected prediction error. We ran five trials. In each trial, 80% of the samples were used to train the classifier, and 20% were set aside for testing. A simple average of the five predictions gives us an estimation of the expected prediction error. Among the models tested were a support vector machine (SVM): a classifier that uses linear and non-linear mapping of the input vectors into high-dimensional spaces, building hyperplanes that allow distinguishing between the data classes; random forest (RFOR): an ensemble learning method for classification that works by building a multitude of decision trees at training time; and linear discriminant analysis (LDA): a method that finds a linear combination of features that characterizes or separates the desired classes. The SVM classifier has a better discriminating power than RFOR and LDA. This result cannot be generalized, however (Shao and Lunetta, 2012) compared the algorithms: support vector machine, neural network, and CART to test classification performance using MODIS time-series data, also concluded the SVM was superior to the other algorithms in the land use classification.

As an example, Fig. 2 shows a plot of the NDVI values of 370 time series for the land cover class “Pasture” based on ground samples. Each thin line is one time series. The darker lines are the median and first and third quartile values. By visualizing the data, the challenge of distinguishing noise from natural variation becomes clear. The data shows natural variability due to different climate regimes and shows noise associated with cloud cover. To avoid losing information, we used the raw data to train a SVM, a classifier which is robust with respect to

Table 1
Experiments with the SVM, RFOR and LDA models, for model selection.

	SVM	RFOR	LDA
Accuracy	0.93	0.92	0.91
Kappa	0.92	0.91	0.90

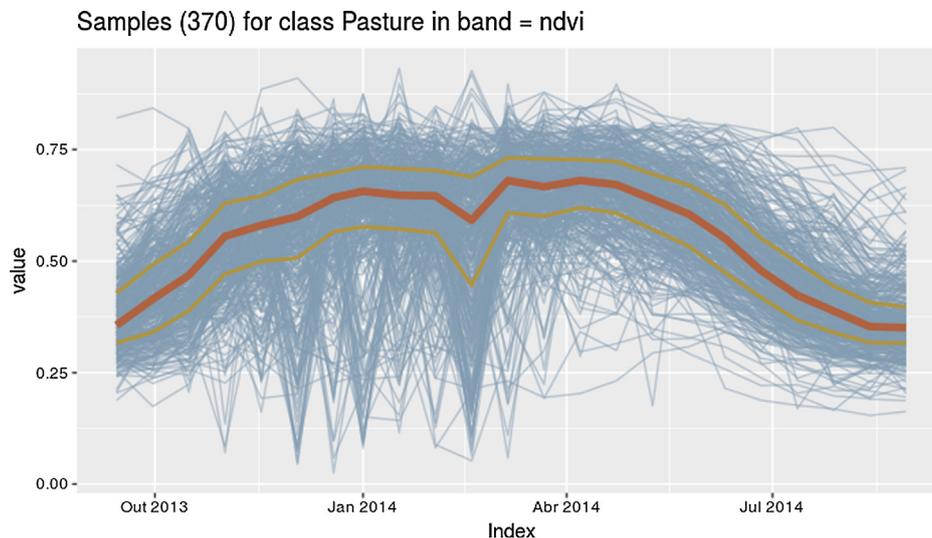


Fig. 2. Time series of 370 ground samples for land cover class "Pasture" in the state of Mato Grosso, Brazil.
Source: authors.

noisy data sets (Hastie et al., 2009).

The SVM is a classifier which considers that the boundary between two classes is non-linear. In its simplest form, an SVM implements a linear classifier by defining boundaries in an n -dimensional space to distinguish two classes. SVMs build hyperplanes that represent the largest separation between the two classes. The hyperplanes maximize the distance from the planes to the nearest data point on each side. The training samples that define the hyperplane of maximum margin are called support vectors. There are many cases where the classes cannot be correctly distinguished by linear hyperplanes. In these situations, the SVM algorithm uses non-linear mappings to project the input vectors to a very high-dimension feature space. In this new feature space, the SVM builds a linear decision surface (Cortes and Vapnik, 1995). SVM implementations include polynomial and radial kernels to deal with non-linear class boundaries.

2.3. Computational infrastructure

Progress on big EO data analytics depends on researchers developing and sharing new methods. Thus, an architecture for big EO data analytics should meet the needs of the researchers. Results should be shared with the scientific community, enabling collaborative work. Researchers should be able to replicate best practices and build their own infrastructure. To achieve these goals, our architecture uses the following building blocks:

1. The SciDB open source array database (Stonebraker et al., 2013) that allows easy mapping of big EO data to its data structure.
2. R as the tool for big data analytics, so that researchers can thus scale up their methods, reuse previous work, and collaborate with their peers.
3. The R packages SITS (Simoes et al., 2017) and dtwSat (Maus et al., 2017), for big EO analytics on SITS.
4. A set of web services for big EO data, adapted to the needs of SITS (Vinhas et al., 2016).

Array databases split large volumes of data into distributed servers in a "nothing shared" way: a big array is broken into "chunks" that are distributed among different servers. Array database management system (DBMS) such as SciDB (Stonebraker et al., 2013) reduce the impedance mismatch between the data model (raster), the storage model (arrays), and the analysis functions. Entire collections of image data fit into single spatiotemporal arrays. Using array DBMS with

statistical computing is a natural solution for EO applications. SciDB has an R interface that allows researchers to parallelize complex analyses and run algorithms on large remote sensing data sets (Fig. 3). This solution is a suitable compromise between the needs for massive parallel data processing and maximum flexibility in algorithm design.

In terms of hardware, our architecture has five servers with two six-core CPUs, operating at 2.4 GHz with a 15 MB cache. Each server has 96 GB of RAM and 16 TB of data storage. This gives 60 cores that can work in parallel in a nothing shared data storage. The array database SciDB includes the full set of MODIS MOD13Q1 images for South America at a 250 m resolution, with 13,800 images associated with 317 billion data series. The case study described in this paper covers the state of Mato Grosso, Brazil, an area of 900,000 km², which has about 20 billion measures. The full processing of all time series to classify 16 years of data in Mato Grosso takes about 6 h using the R-SciDB interface. Given these results, we argue that using SciDB combined with R is an adequate solution for big EO data analytics.

3. Methods

3.1. Study area

The state of Mato Grosso has an area of 903,357 km² and is the third largest state of Brazil. It includes three of Brazil's biomes: Amazon, Cerrado, and Pantanal (Fig. 4). The Cerrado biome covers 40% of the whole territory and is an important biome related to animals species (more than 1,500 species), birds (837 species), amphibians (150 species), and reptiles (120 species). The Pantanal biome, which occupies 7% of the state, is rich in bio-diversity, and is a UNESCO World Natural Heritage and Biosphere Reserve. In the Amazon biome in Mato Grosso, there are two types of forest: Amazon Forest and Seasonal Forest, which together occupy about 53% of the territory of Mato Grosso.

3.2. Data

We used the MOD13Q1 product from National Aeronautics and Space Administration from 2001 to 2016, provided every 16 days at 250-meter spatial resolution in the sinusoidal projection (Didan, 2015)¹. To do the analysis, we selected the indices NDVI and EVI, and the

¹ Since the 2004 MODIS image presented high amount of noise in the MIR band, results from 2004 were not used in the analysis.

Fig. 3. Remote execution of R scripts in SciDB.

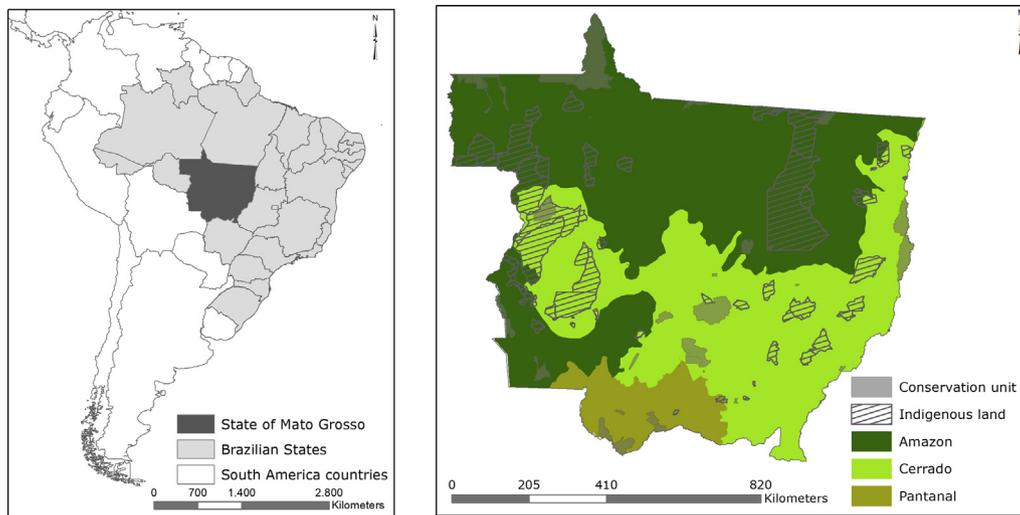
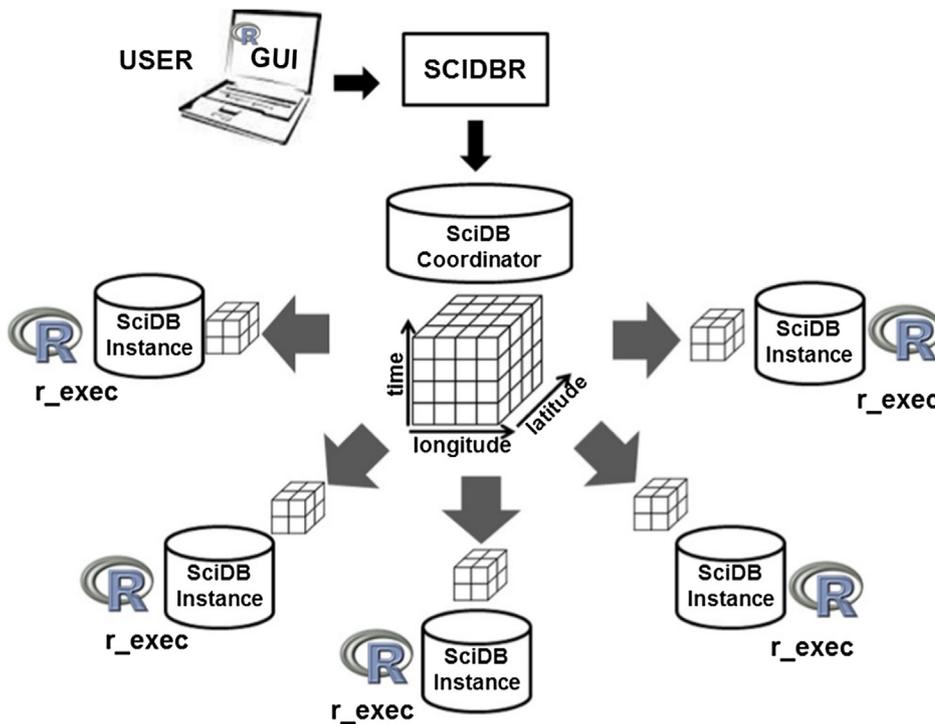


Fig. 4. Left: Location of Mato Grosso relative to South America continent and Brazil. Right: Biomes, conservation units and indigenous lands in Mato Grosso.

Table 2
Ground samples used as training data for Mato Grosso.

Class label	Count	Freq
Cerrado	400	18.9%
Fallow-Cotton	34	1.6%
Forest	138	6.5%
Pasture	370	17.5%
Soy-Corn	398	18.8%
Soy-Cotton	399	18.9%
Soy-Fallow	88	4.2%
Soy-Millet	235	11.1%
Soy-Sunflower	53	2.5%

original bands nir and mir. We defined nine land cover classes: (1) forest, (2) cerrado, (3) pasture, (4) soybean-fallow (single crop), (5) fallow-cotton (single crop), (6) soybean-cotton (double crop), (7)

soybean-corn (double crop), (8) soybean-millet (double crop), and (9) soybean-sunflower (double crop). According to the Brazilian Institute of Geography and Statistics (IBGE), crop classes (4)–(9) accounted for more than 93% of Mato Grosso agricultural lands in 2015. Crop and pasture ground data was collected through field observations and farmer interviews (Kastens et al., 2017; Sanches et al., 2018). Samples for the cerrado and forest classes were provided through field work and high resolution images. Ground samples for soybean-fallow class were provided through field work, based on previous work (Arvor et al., 2011). Table 2 lists the distribution of the ground samples.

To get an overall view of the temporal signatures of the ground samples, we used a generalized additive model (GAM) to estimate the joint distribution of the set samples for each class (Maus et al., 2016). The GAM estimates use a smoothing function that approximates the idealized temporal patterns (Fig. 5). One can observe that the temporal signatures of the soy-corn, soy-millet and soy-sunflower classes are

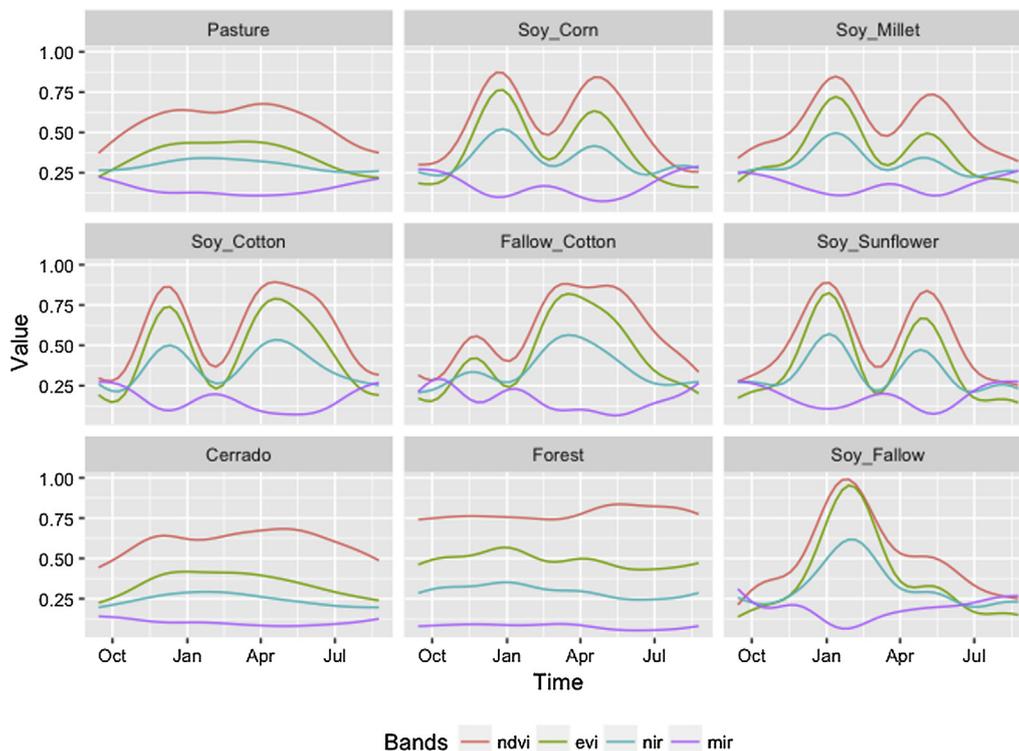


Fig. 5. Estimated temporal patterns of NDVI, EVI, NIR and MIR bands for the selected land cover classes.

similar, leading to some possible confusion. As our experiments show, these are the classes which are harder to distinguish (see Section 4.1).

3.3. Time series classification

The MODIS images were inserted into the SciDB database, and the three-dimensional array of satellite data was created. In order to retrieve the time series for the set of ground sample points, from which we knew latitude, longitude, and land cover over a specific time interval, we used the web time series service (WTSS), available in the SITS R package. These time series were used to train the SVM model.

The SVM is a generalization of the separating hyperplane classifier (Hastie et al., 2009). This generalization combines the notions of the optimal separating hyperplanes, soft margins, and the enlargement of the input attribute space with a nonlinear mapping to a feature space. In a separating hyperplane formulation, the classifier works only with separable sample sets. It finds a linear boundary in the input attribute space that not only divides two classes but maximizes its margin between them. A boundary is a hyperplane that divides the entire attribute space into two parts. Fig. 6 presents an overview of the methodology used to classify the SITS.

3.4. Post-processing masks

We applied three masks to the final classified maps. The sugarcane masks from 2003 to 2016 come from the Canasat project (Rudorff et al., 2010). This project maps sugarcane areas in the South-Central region of Brazil using LANDSAT images (Adami et al., 2012). Sparovek et al. (2015) provided the urban area mask. The water mask comes from Pekel et al. (2016), who used three million LANDSAT satellite images to quantify changes in global surface water over the past 32 years (1984–2015).

3.5. Accuracy

To estimate the classification accuracy, we ran a 5-fold cross-

validation procedure (Wiens et al., 2008). In this validation, we ran five different assessments. For each assessment, 80% of the samples were used for training and 20% for prediction. The accuracy of all five classifications is averaged to produce a single estimation. Using a 5-fold validation has some advantages compared to other validation methods. The goal of cross-validation is to find out how well a given statistical learning procedure can be expected to perform using independent data (James et al., 2013). Increasing the fold reduces the bias of the estimate of model performance using independent data, at the cost of increasing its variance. Given the number of samples for each class (see Table 2), we consider a 5-fold cross-validation to be adequate for our training set.

4. Results and discussion

4.1. Classification accuracy

The 5-fold cross-validation estimated an overall accuracy of 94% and the Kappa index was 0.92. Producer and user accuracies for all classes were close to or better than 90% (Table 3). This confirms the applicability of the proposed method to classify agricultural areas. As expected, the matrix shows some confusion between the soybean-corn and soybean-millet classes. Since corn and millet have similar physical characteristics, they can be spectrally confused (Fig. 5). Both are grasses, with lanceolate leaves; the height of corn can reach up to 3.5 m, whereas millet varies between 1.5 and 3 m, and can reach more than 5 m. The results show a good discrimination between the different crops, which improves on previous work (see Table 4). This methodology brought advance in the agricultural crop mappings accuracy. Until then, the highest overall accuracy achieved in a mapping in Mato Grosso, which distinguished different crops, using MODIS time series had been 80% (Chen et al., 2018). Galford et al. (2008) study was overall accuracy equal 94%, however the authors used only three classes: single crop, double crop, and not row crops.

Measured deforestation in Mato Grosso from 2005 to 2016 was 4.1 million hectares, 12% of the total forest area, considering both the Amazon and Cerrado biomes. The areas classified as forest were

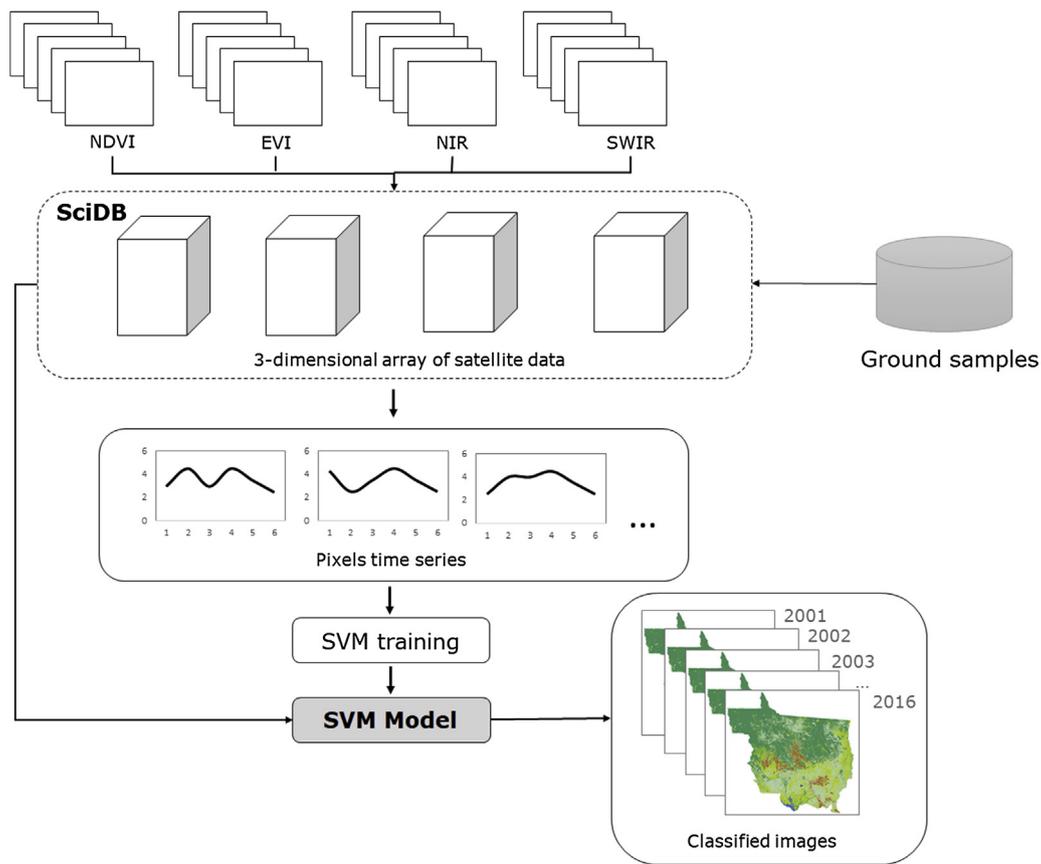


Fig. 6. Time series classification methodology.

compared with the Hansen et al. (2013) mapping for the year 2000. These authors used LANDSAT images to map the percent of tree crown cover densities. Trees were defined as all vegetation taller than 5 m in height. In order to separate the forest areas, we examined the areas with more than 25% tree cover on the Hansen et al. (2013) map. We found that 99% of the pixels classified as forest match the pixels indicated by Hansen et al. (2013) as having more than 25% tree cover. For the cerrado class, 62% of the pixels match the pixels indicated by Hansen et al. (2013) as having more than 25% tree cover. This difference occurs because our cerrado class includes both wooded and wooded-herbaceous physiognomies.

The pixels labelled as pasture were compared to the pasture mapping done by Parente et al. (2017), who produced a pasture mask for Brazil in 2015 using LANDSAT-8 images and a RFOR classifier. The difference between the total pasture area in our results and that mapped by Parente et al. (2017) for the state of Mato Grosso was 4%. The correlation between the individual pasture pixels in both maps was

89%. Part of this difference can be explained by the fact that the map by Parente et al. (2017) uses additional masks to exclude indigenous areas and national parks. An additional factor is that Parente et al. (2017) used LANDSAT images, whereas we used MODIS. We note that user accuracy for the pasture class on the map made by Parente et al. (2017) is 83%, whereas the user accuracy for the pasture class using SVM classification is 96%. Further detailed studies are required to assess the quality of these approaches and to improve pasture assessments in the Amazonia and Cerrado biomes.

Fig. 7 shows two of the resulting maps of the spatial distribution of land cover classes, for the years 2005 and 2016. The full data set, including all resulting maps and the ground sample data as well as the software used to produce the maps, is openly available on the internet. Please see the end of the paper for more detailed information.

Table 3

Confusion matrix of MODIS time series images, obtained by 5-fold cross validation of classification of field data, and values of producer's accuracy (PA) and user's accuracy (UA) for each class.

	1	2	3	4	5	6	7	8	9	UA
1 Cerrado	393	0	0	12	0	0	0	0	0	0.97
2 Fallow-Cotton	0	33	0	0	1	2	0	0	0	0.92
3 Forest	1	0	136	0	0	0	0	0	0	0.99
4 Pasture	6	0	1	357	3	1	0	5	0	0.96
5 Soy-Corn	0	1	1	1	352	18	0	26	4	0.87
6 Soy-Cotton	0	0	0	0	13	376	0	4	0	0.96
7 Soy-Fallow	0	0	0	0	0	0	88	0	0	1.00
8 Soy-Millet	0	0	0	0	25	2	0	199	2	0.87
9 Soy-Sunflower	0	0	0	0	4	0	0	1	47	0.90
PA	0.98	0.97	0.99	0.96	0.88	0.94	1.00	0.85	0.89	

Table 4
Summary of selected approaches for cropland mapping using MODIS time-series data.

Method	Description of examples	Accuracy	References
Traditional classifiers and machine learning classifiers	Maximum likelihood classifier based on MODIS EVI time-series data to map crop types and cropping system in Mato Grosso, Brazil	Soybean: 66.67% (UA) 64% (PA) Soybean-Non commercial crop: 56.03% (UA) 65% (PA) Soybean-Maize: 72.63% (UA) 69% (PA) Soybean-Cotton: 93.41% (UA) 85% (PA) Cotton: 85.29% (UA) 87% (PA) Overall accuracy: 74% Kappa index: 0.67	Arvor et al. (2011)
	Decision tree classifier based on MODIS EVI time-series data to classify agricultural land use data in Mato Grosso, Brazil	Forest: 93.85 (UA) 89.71 (PA) Pasture: 93.55 (UA) 94.16 (PA) Cropland: 89.02 (UA) 91.25 (PA) Overall accuracy: 92.38 Kappa: 0.88	Macedo et al. (2012)
	Decision tree classifier based on MODIS NDVI and EVI time-series data to classify agricultural land use data in Mato Grosso, Brazil	Soy-Commercial*: 79% (UA) 84% (PA) Soy-Cover**: 80% (UA) 83% (PA) Soy-Cotton: 76% (UA) 43% (PA) Cotton: 77% (UA) 66% (PA) Pasture: 76% (UA) 58% (PA) Overall accuracy: 79.3% Kappa: 0.66	Brown et al. (2013)
	Random forest classifier based on MODIS NDVI time-series data to classify agricultural land use data in Mato Grosso, Brazil	Pasture/Cerrado: 89% (UA) 92% (PA) Soy-Single: 74% (UA) 75% (PA) Cotton: 70% (UA) 49% (PA) Soy-Double: 78% (UA) 83% (PA) Soy-Cotton: 85% (UA) 59% (PA) Overall Accuracy: 79% Kappa: 0.71	Kastens et al. (2017)
	Decision tree classifier based on MODIS NDVI time-series data to classify agricultural land use data in Mato Grosso, Brazil	Soy-Maize: 78% (UA) 79% (PA) Soy-Cotton: 88% (UA) 70% (PA) Soy-Fallow: 30% (UA) 61% (PA) Soy-Pasture: 81% (UA) 74% (PA) Fallow-Cotton: 78% (UA) 66% (PA) Single crop: 87% (UA) 80% (PA) Overall accuracy: 73%	Chen et al. (2018)
	Support vector machine classifier based on MODIS NDVI, EVI, nir and mir time-series data to classify agricultural land use data in Mato Grosso, Brazil	Cerrado: 0.97 (UA) 0.98 (PA) Fallow-Cotton: 0.92 (UA) 0.97 (PA) Forest: 0.99 (UA) 0.99 (PA) Pasture: 0.96 (UA) 0.96 (PA) Soy-Corn: 0.87 (UA) 0.88 (PA) Soy-Cotton: 0.96 (UA) 0.94 (PA) Soy-Fallow: 1 (UA) 1 (PA) Soy-Millet: 0.87 (UA) 0.85 (PA) Soy-Sunflower: 0.90 (UA) 0.89 (PA) Overall accuracy: 94%	by authors
Data transform algorithms	Wavelet analysis based on MODIS EVI time-series data to map cropland distribution in Mato Grosso, Brazil	Not row crops***: 94.7% (UA) 100.0% (PA) Single crop: 90.0% (UA) 81.8% (PA) Double crop: 100.0% (UA) 71.4% (PA) Overall accuracy: 94% Khat: 85.7%	Galford et al. (2008)

* Double crop

** Single crop

*** Row crops = soybean, maize, and dryland rice

4.2. Cropland expansion and intensification

We compared our crop classification to the IBGE official crop statistics (IBGE, 2017). IBGE holds yearly sample surveys of agricultural production at the municipal level, the so-called PAM (“Pesquisa Agrícola Municipal”). At the state level, the soybean, cotton, corn, and sunflower areas mapped by our work had a correlation of 98%, 96%,

73%, and 80%, respectively, with the state level results of the IBGE PAM (Fig. 8). Compared to the IBGE PAM, the classification over-estimated the soybean and corn areas and underestimated the cotton and sunflower areas. These differences may have been caused by the spatial resolution of the MODIS images (250 m), which generates spectral mixing due to different land uses within a single pixel (Friedl et al., 2002; Zhong et al., 2016). However, the lack of a reliable

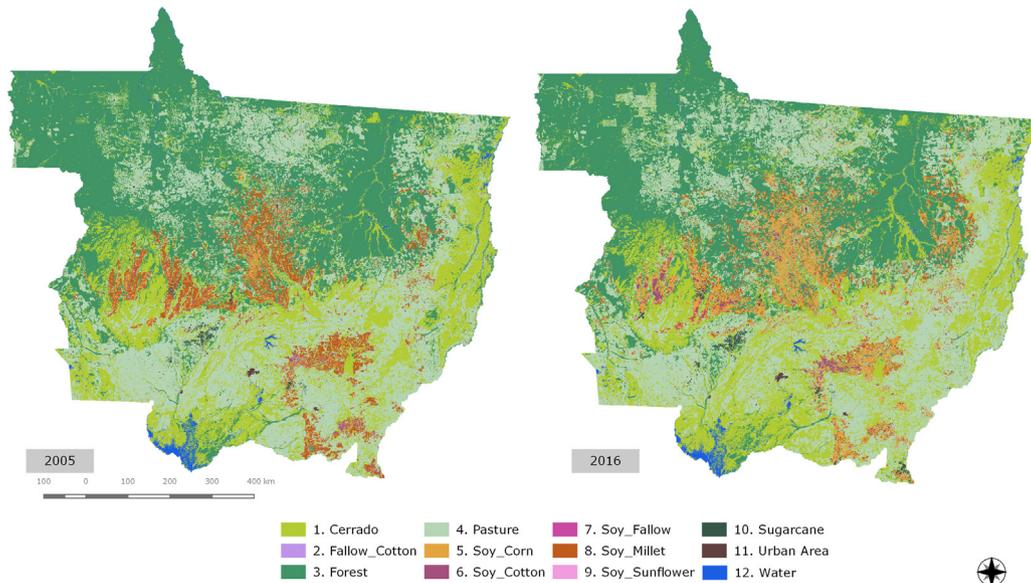


Fig. 7. SVM classification for state of Mato Grosso in 2005 and 2016, with sugarcane, urban area and water masks.

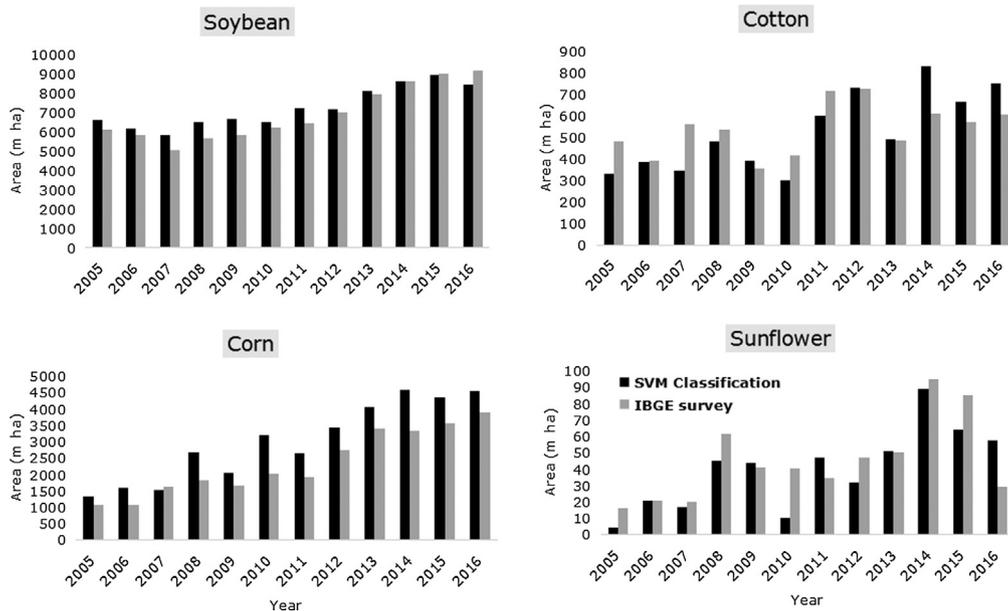


Fig. 8. Total area of soybean, cotton, corn and sunflower in state of Mato Grosso estimated by SVM classification and the IBGE cropland survey.

reference data set precludes an objective assessment. The IBGE PAM results are based on surveys and not on samples. Thus, they contain uncertainties as well and should not be taken as absolute references. To produce the PAM, IBGE staff do not go into the field. They contact large producers and also rely on subjective estimates of the local IBGE staff. Therefore, comparing our results with data from the PAM does not entail an accuracy estimate of our work. Correlation between the sum of agricultural areas classified in this study and the estimates by IBGE for the harvests from 2005 to 2016 are equal to 98%. Thus, we consider that the proposed methodology is effective for mapping agricultural crops in Mato Grosso.

In Mato Grosso, the cropland area increased by 1.83 million hectares (26.5%) from 2005 to 2016. The greatest expansion occurred during 2007–2008 and 2012–2013, with a growth rate of 11.9% and 11.6%, respectively. The expansion of agricultural areas occurred mainly around the BR163 highway, which crosses Mato Grosso in a north-south direction. At the edge of this road, it is possible to observe the expansion of agriculture in the northern direction within the

Amazon biome.

Today, municipalities such as Querencia and Tabaporã, where there was almost no presence of agriculture in the early 2000s, are prolific producers of soybeans. Arvor et al. (2012) also observed the same expansion trend around the BR163 highway. This area has the highest soybean yields in Mato Grosso due to its soil, topography, and climate (Spera et al., 2014). The area also has the largest proportion of double cropping due to its longer rainy season (Arvor et al., 2013). Furthermore, the Brazilian government is asphaltting the BR163 highway up to its connection with the Mirituba and Santarem harbours in the state of Para on the Amazon River, which would decrease the transportation costs for soybean exports.

The soybean class also decreased in area from 2005 to 2006 and from 2006 to 2007 due to the economic crisis, when the Brazilian currency was devalued compared with the US dollar. The unfavourable exchange rate affected soybean production from 2005 to 2007 (Arvor et al., 2012). However, soybean cropland increased significantly by 0.93 million hectares (12.9%) between 2012 and 2013. According to

the Brazilian National Supply Company (CONAB, 2013), this growth is due to better prices for soybean in the international market and the repercussions of these prices in the domestic market. New commercial arrangements, such as advance commercialization, also contributed to this increase.

4.3. Cropland and pasture intensification

Due to the increased demand for food and biofuels, producers in the state of Mato Grosso intensified agricultural production by adopting double crop systems. The area cultivated with double crop systems, involving soybeans (first cycle) + some other crop (second cycle) or some other crop (first cycle) + cotton (second cycle), increased from 6.58 to 8.43 million hectares during 2005 to 2016, an increase of 28%. Double crop systems are currently predominant in Mato Grosso. The area devoted to corn also increased by replacing millet, and corn is the crop of choice for planting in consortium with soybeans. Millet lost an area of 1.87 million hectares (61%) to corn. Due to improvements in corn varieties and the increase in Brazil's corn exports, corn has replaced millet as a more profitable option. In the municipality of Campo Verde (located in the Cerrado biome, in the south-east part of the state of Mato Grosso), it is possible to observe this transition from single to double crop systems, with the replacement of fallow-cotton by soybean-cotton, from 2005 to 2015 (Fig. 9). The double crop system is more profitable for the producer, and represents a better use of agricultural areas, allowing an increase in production while at the same time reducing the pressure of expansion over native vegetation, and protect the soil surface from the hot temperatures and isolation typical of tropical climates. Double crop also enables producers to adopt no tillage practices (apart from cotton) which are better from an ecological point of view. Previous authors (Kastens et al., 2017; Spera et al., 2014; Arvor et al., 2012) have already pointed out the increase in double crop production associated with soybeans.

Pasture area in Mato Grosso between 2005 and 2015 declined by 4.6 million hectares, from 28.1 to 23.5 million hectares. According to IBGE, the cattle head in the state has increased from 26.7 million in 2005 to 29.3 million in 2015, a growth of 10% (IBGE, 2017). In Fig. 10, we show that the stocking rate in Mato Grosso has grown steadily. The cattle heads grew by 10%, while pasture decreased by 16% between 2005 and 2015. In general, there is a trend towards pasture intensification coupled with abandonment of frontier areas, especially those in the northern-most part of the state.

Our results for 2016 point to a total of 28.9 million hectares of pasture. This increase in 2016 can be explained by the increase in deforestation that occurred in 2015. After the Soy Moratorium was

introduced in 2008, the deforested areas have generally been converted to pasture areas. The deforested area in Mato Grosso has been suffering oscillations since the beginning of the mapping in 1989, according to PRODES. These oscillations can also be observed in pasture areas. Between 2009 and 2014, the deforestation rate in the state of Mato Grosso has remained around 1000 km² per year. In 2015 there was an increase to 1600 km² per year. This increase directly reflected the increase in pasture areas in 2016, which in turn reflected a decrease in the stocking rate.

4.4. Overall discussion

The results highlight important trends in agricultural intensification in Mato Grosso. Double crop systems are now the most common production system in the state, thus increasing the potential for sparing land from agricultural production. As pointed out by other authors (Spera et al., 2014; Gibbs et al., 2015; Kastens et al., 2017) the impact of crop production on deforestation has decreased since 2005. Arguably, this is due to a combination of factors, including the Soy Moratorium, increased law enforcement, and the occupation of the best farming areas in the state's Amazon biome (Spera et al., 2014; Arvor et al., 2016). A less-studied issue is the increase in pasture productivity. Pasture expansion and intensification has been less studied than crop expansion, although it has a stronger impact on deforestation and GHG emissions. Our data points to a significant increase in the stocking rate in the state of Mato Grosso and to the possible abandonment of pasture areas opened in the state's frontier. As for analysing the deforestation (Isabelle and Damien, 2016; Arvor et al., 2016) and cropland intensification dynamics (VanWey et al., 2013), further studies coupling fieldwork, mapping and economic models are required to better understand the underlying driving forces for the cattle-raising sector.

Our results point out the conflicting forces at play in the agricultural expansion in Mato Grosso. In some segments (such as crop production), there is a consolidation underway. The best producing areas have been occupied, and the emphasis now is on increasing productivity by the adoption of double crop systems. In the case of cattle-raising, one can see mixed signs. On one hand, there is a modest, but significant, increase in the stocking rate (until 2015); however, there is still expansion going on in the northern frontiers of the state, which need to be better studied. Many factors could be at play, including land speculation, and indirect land use change due to crop expansion. This situation poses important challenges. The large-scale mapping that we produced for Mato Grosso needs to be expanded to the whole Amazon and Cerrado biomes, and also needs to be supported by economic analysis. There is a need for continuous improvement of land cover classification

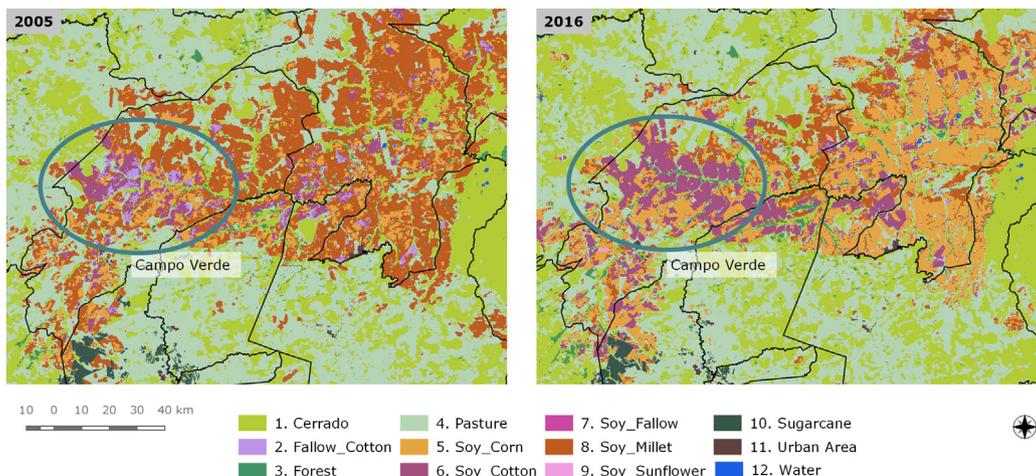


Fig. 9. Change in the production system in the municipality of Campo Verde - MT from single crop in 2005 to double crop in 2016. In the gray highlight it is possible to observe the change from the fallow cotton class in 2005 to the soybean cotton class in 2016.

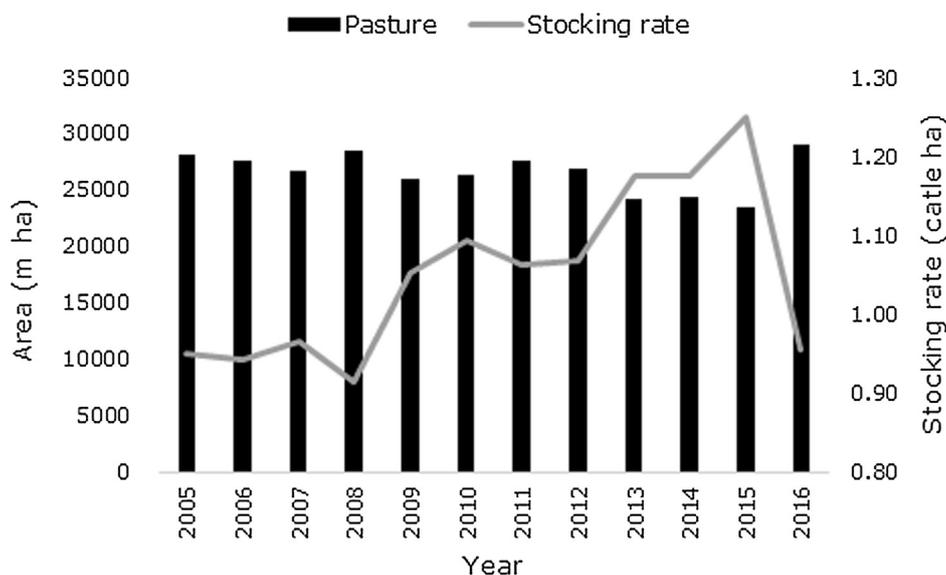


Fig. 10. Pasture area, provide by SVM classification, and stoking rate, calculate by IBGE cattle head value and pasture areas, in Mato Grosso state, from 2005 to 2016.

using remote sensing time series, by using LANDSAT-class satellites to increase the spatial resolution and classification accuracy.

For future classifications, we suggest new tests using new images data sets (e.g., Sentinel, LANDSAT), and stratification of the study area by geography before the classification as well as testing the inclusion of ancillary data, such as climate data or digital elevation models.

5. Conclusion

This paper provides the first detailed mapping of cropland and pasture intensification in the state of Mato Grosso that includes more detailed information about different crops. It was possible to disaggregate agricultural classes such as millet and corn and separate types of vegetation such as forest and cerrado. The results allow a direct estimation of the direct land use change, expansion and intensification of agricultural crops, and expansion of pasture monitoring. We were able to contribute to an assessment of the interplay between production and protection in the Brazilian Amazon and Cerrado biomes.

The methods presented in this paper are innovative in their use of SITS coupled with advanced statistical learning. The idea of increasing the dimension of the classification space by using the full depth of the time series is, to our knowledge, a new one. The good validation results we have obtained demonstrate the power of the proposed method. Two important qualities of the method are its reproducibility and directness. We hope that our results encourage further work on the processing of SITS using open source software.

Data and software

The detailed maps for the state of Mato Grosso from 2001 to 2016 at full MODIS resolution as well as the ground samples used as training data have been deposited at the PANGAEA Earth Sciences Data Repository. The DOI to access the data is <https://doi.pangaea.de/10.1594/PANGAEA.881291>.

Acknowledgements

This work was funded by the São Paulo Research Foundation (FAPESP) through an eScience Program grant 2014/08398-6. We also thank FAPESP for grants 2016/23750-3 (MP) and 2016/16968-2 (RB) and the Coordination for the Improvement of Higher Education (CAPES) and National Council for Scientific and Technological Development (CNPq) for grants 312151/2014-4 (GC) and 140684/

2016-6 (RS). This work is also supported by the International Climate Initiative of the Germany Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (IKI) under grant 17-III-084-Global-A-RESTORE+ (“RESTORE+: Addressing Landscape Restoration on Degraded Land in Indonesia and Brazil”). The work also received support from the Science for Nature and People Partnership (SNAPP), National Center for Ecological Analysis and Synthesis at the University of California Santa Barbara. We thank Ricardo Cartaxo, who provided insight and expertise to support this paper, and Luiz Assis and Eduardo Llapa for their technical support.

References

- Adami, M., Rudorff, B.F.T., Freitas, R.M., Aguiar, D.A., Sugawara, L.M., Mello, M.P., 2012. Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in Brazil. *Sustainability* 4 (4), 574–585.
- Almeida, C., Coutinho, A., Esquerdo, J., Adami, M., Venturieri, A., Diniz, C., Dessay, N., Durieux, L., Gomes, A., 2016. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazonica* 46 (09), 291–302.
- Arvor, D., Daugeard, M., Tritsch, I., de Mello-Théry, N., Théry, H., Dubreuil, V., 2016. Combining socio-economic development with environmental governance in the Brazilian Amazon: the Mato Grosso agricultural frontier at a tipping point. *Environ. Dev. Sustain.* 20 (1), 1–22.
- Arvor, D., Dubreuil, V., Simões, M., Bégué, A., 2013. Mapping and spatial analysis of the soybean agricultural frontier in Mato Grosso, Brazil, using remote sensing data. *Geo J.* 78 (5), 833–850.
- Arvor, D., Jonathan, M., Meirelles, M., Dubreuil, V., Durieux, L., 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* 32 (22), 7847–7871.
- Arvor, D., Meirelles, M., Dubreuil, V., Bégué, A., Shimabukuro, Y.E., 2012. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Appl. Geography* 32 (2), 702–713.
- Assunção, J., Gandour, C., Rocha, R., 2015. Deforestation slowdown in the Brazilian Amazon: prices or policies? *Environ. Dev. Econ.* 20 (6), 697–722.
- Belgiu, M., Dragut, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogrammetry Remote Sensing* 114, 24–31.
- Brown, J.C., Kastens, J.H., Coutinho, A.C., Victoria, D.d.C., Bishop, C.R., 2013. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sens. Environ.* 130, 39–50.
- Camara, G., Assis, L.F., Ribeiro, G., Ferreira, K.R., Llapa, E., Vinhas, L., 2016. Big earth observation data analytics: matching requirements to system architectures. In: *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*. ACM, Burlington, CA, USA, pp. 1–6.
- Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L.V., Sanches, I.D., da Silva, R.F.B., Huang, J., Luiz, A.J.B., de Oliveira, M.A.F., 2018. Mapping croplands, cropping patterns, and crop types using modis time-series data. *Int. J. Appl. Earth Obs. Geoinf.* 69, 133–147.
- CONAB, 2013. Follow up of Brazilian crop: grains (ninth survey). Tech. Rep., National Supply Company (CONAB).
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20 (3), 273–297.

- Didan, K., 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC. (Available at: <https://doi.org/10.5067/modis/mod13q1.006>).
- Estel, S., Kuemmerle, T., Alcantara, C., Levers, C., Prishchepov, A., Hostert, P., 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. *Remote Sens. Environ.* 163, 312–325.
- Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., Woodcock, C., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* 83 (12), 287–302.
- Galford, G.L., Mustard, J.F., Melillo, J., Gendrin, A., Cerri, C.C., Cerri, C.E., 2008. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens. Environ.* 112 (2), 576–587.
- Garnett, T., 2015. Where are the best opportunities for reducing greenhouse gas emissions in the food system (including the food chain)? *Food Policy* 36, S23–S32.
- Gibbs, H.K., Rausch, L.B., Munger, J., et al., 2015. Brazil's soy moratorium. *Science* 347 (6220), 377–378.
- Gibbs, H.K., Ruesch, A.S., Achard, F., Clayton, M.K., Holmgren, P., Ramankutty, N., Foley, J.A., 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Nat. Acad. Sci.* 107 (38), 16732–16737.
- Gomez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. *ISPRS J. Photogrammetry Remote Sensing* 116, 55–72.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342 (6160), 850–853.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer, New York.
- IBGE, 2017. Brazilian Institute of Geography and Statistics. Municipal Agricultural Production. (Available at: <http://www2.sidra.ibge.gov.br/>).
- INPE, 2014. Amazon Deforestation Monitoring Project - Prodes. URL www.obt.inpe.br/prodes.
- Isabelle, T., Damien, A., 2016. Transition in environmental governance in the Brazilian Amazon: emergence of a new pattern of socio-economic development and deforestation. *Land Use Policy* 59, 446–455.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning: with Applications in R.* Springer, New York, EUA.
- Jönsson, P., Eklundh, L., 2004. TIMESAT? a program for analyzing time-series of satellite sensor data. *Comput. Geosciences* 30 (8), 833–845.
- Kastens, J., Brown, J., Coutinho, A., Bishop, C., Esquerdo, J., 2017. Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PLOS ONE* 12 (4), e0176168.
- Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., Healey, S.P., Helmer, E.H., Hostert, P., Lyons, M.B., Meigs, G.W., Pflugmacher, D., Phinn, S.R., Powell, S.L., Scarth, P., Sen, S., Schroeder, T.A., Schneider, A., Sonnenschein, R., Vogelmann, J.E., Wulder, M.A., Zhu, Z., 2014. Bringing an ecological view of change to landsat-based remote sensing. *Front. Ecol. Environ.* 12 (6), 339–346.
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series. *Remote Sens. Environ.* 114 (12), 2897–2910.
- Lambin, E.F., Helmut, G., 2006. *Land-Use and Land-Cover Change. Global Change – The IGBP Series.* Springer.
- Macedo, M.N., DeFries, R.S., Morton, D.C., Stickler, C.M., Galford, G.L., Shimabukuro, Y.E., 2012. Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. *Proc. Nat. Acad. Sci.* 109 (4), 1341–1346. <https://doi.org/10.1073/pnas.1111374109>.
- Maus, V., Appel, M., Giorgino, T., 2017. Time-weighted dynamic time warping for satellite image time series analysis. R package version 0.2.4.9.
- Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F.M., de Queiroz, G.R., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 9 (8), 3729–3739.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. *ISPRS J. Photogrammetry Remote Sensing* 66 (3), 247–259.
- Nepstad, D., McGrath, D., Stickler, C., Alencar, A., Azevedo, A., Swette, B., Bezerra, T., DiGiano, M., Shimada, J., Seroa da Motta, R., Armijo, E., Castello, L., Brando, P., Hansen, M.C., McGrath-Horn, M., Carvalho, O., Hess, L., 2014. Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. *Science* 344 (6188), 1118–1123.
- Parente, L., Ferreira, L., Faria, A., Nogueira, S., Araújo, F., Teixeira, L., Hagen, S., 2017. Monitoring the Brazilian pasturelands: a new mapping approach based on the Landsat 8 spectral and temporal domains. *Int. J. Appl. Earth Obs. Geoinf.* 62, 135–143.
- Pasquarella, V.J., Holden, C.E., Kaufman, L., Woodcock, C.E., 2016. From imagery to ecology: leveraging time series of all available landsat observations to map and monitor ecosystem state and dynamics. *Remote Sens. Ecol. Conservation* 2 (3), 152–170.
- Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 418–422.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 187, 156–168.
- Petitjean, F., Inglada, J., Gancarski, P., 2012. Satellite image time series analysis under time warping. *IEEE Trans. Geoscience Remote Sens.* 50 (8), 3081–3095.
- Rudorff, B.F.T., Aguiar, D.A., Silva, W.F., Sugawara, L.M., Adami, M., Moreira, M.A., 2010. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. *Remote Sens.* 2 (4), 1057–1076.
- Sanches, I.D., Feitosa, R.Q., Diaz, P.M.A., Soares, M.D., Luiz, A.J.B., Schultz, B., Maurano, L.E.P., 2018. Campo verde database: seeking to improve agricultural remote sensing of tropical areas. *IEEE Geoscience Remote Sens. Lett.* 15 (3), 369–373.
- Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and {CART} algorithms for the land-cover classification using limited training data points. *ISPRS J. Photogrammetry Remote Sensing* 70, 78–87.
- Simoes, R., Camara, G., Carvalho, A., Maus, V., Assis, L., Maciel, A., Queiroz, G., 2017. *SITS: Satellite Image Time Series Analysis. R package version 0.9.30.*
- Sparovek, G., Barreto, A.G.O.P., Matsumoto, M., Berndes, G., 2015. Effects of governance on availability of land for agriculture and conservation in Brazil. *Environ. Sci. Technol.* 49, 10285–10293.
- Spera, S.A., Cohn, A.S., VanWey, L.K., Mustard, J.F., Rudorff, B.F., Risso, J., Adami, M., 2014. Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. *Environ. Res. Lett.* 9 (6), 064010.
- Stonebraker, M., Brown, P., Zhang, D., Becla, J., 2013. Scidb: A database management system for applications with complex analytics. *Comput. Sci. Eng.* 15 (3), 54–62.
- VanWey, L., Spera, S., de Sa, R., Mahr, D., Mustard, J., 2013. Socioeconomic development and agricultural intensification in mato grosso. *Philos. Trans. R. Soc. B. Biol. Sci.* 368 (1619), 20120168.
- Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing Environ.* 114 (1), 106–115.
- Vinhas, L., Ribeiro, G., Ferreira, K.R., Camara, G., 2016. Web services for big earth observation data. In: *Proceedings of the 17th Brazilian Symposium on Geoinformatics. INPE, Campos do Jordão, SP, Brazil*, pp. 26–35.
- Wiens, T.S., Dale, B.C., Boyce, M.S., Kershaw, G.P., 2008. Three way k-fold cross-validation of resource selection functions. *Ecol. Model.* 212 (3), 244–255.
- Zhong, L., Hu, L., Yu, L., Gong, P., Biging, G.S., 2016. Automated mapping of soybean and corn using phenology. *ISPRS J. Photogrammetry Remote Sensing* 119, 151–164.