A Time Series Mining Approach for Agricultural Area Detection

João Paulo da Silva, Jurandir Zullo Jr, and Luciana Alvim Santos Romani

Abstract—Acquiring meaningful data to be employed in building training sets for classification models is a costly task, both in terms of difficult to find suitable samples as well as their quantity. In this sense, Active Learning (AL) improves the training set building by providing an efficient way to select only essential data to be attached to the training set, consequently reducing its size and even enhancing model's accuracy, when compared to random sample selection. In this paper, we proposed a framework for time series classification in order to monitor sugarcane area in São Paulo, Brazil. The AL approach consisted of selecting seasonal time series information from less than 1 percent of each class' pixels to build the training set and evaluate this selection by an expert user supported by distance measurements, repeating this process until both distance measurement thresholds were satisfied. In most years, the classification results presented about 90 percent of correlation with official estimates based on both traditional and satellite image analysis methods. This framework can then help Land Use Change (LUC) monitoring as it produced similar results compared to other methods that demands more human and financial resources to be adopted.

Index Terms—Environment, pixel classification, remote sensing, time series analysis

1 Introduction

Brazilian southeastern and central regions economy is based on pillars in which agriculture is one of the most important. Historically, the state of São Paulo is the major national sugarcane (*Saccharum* spp.) producer, being that large investments were made in the 2000s to increase the production aiming to attend internal and external markets demand. So, between 2001 and 2010, sugarcane area increased from $2.5^{*}10^{6}$ ha to $5^{*}10^{6}$ ha whereas the production increased from $207^{*}10^{6}$ tons to $361^{*}10^{6}$ tons [1]. The sugarcane's cycle in these regions begins in April and lasts until March of the next year, comprising the growing, maturation and harvest phases [2].

The possible climate change, one of the most important discussions nowadays, could bring several consequences to agricultural production, leading to change its current geography [3], [4], [5], for example. Beyond its relevance in fuel and food industries, and although it's still a recent research theme, studies indicate that C4 plants like sugarcane could take advantages of climate changes, precisely of atmosphere's CO_2 concentration growth. This advantage is associated with the increasing of photosynthetic efficiency (and consequently of the biomass production) [6], increasing of non-structural carbohydrates concentration (reflecting in a more efficient production) [7] and reduction of stomatal conductance (that increases the efficiency of plant water use, allowing sugarcane to keep the production with lower water consumption) [8], [9]. With possible changes in sugarcane climate risk zoning and production geography due to economic and climate factors [10], [5], it is important to monitor it efficiently to assess its real impact to environment.

The use of remote sensing data for environmental monitoring improves the assessment of human activities' impacts and also helps decision-makers to reach best practices in crop conduction and planning. In Brazil, the INPE/CANASAT (https://bit.ly/1PYXtAj) project [11], developed by National Institute for Space Research (acronym INPE, in Portuguese), used Landsat-5 data to monitor the sugarcane planted area and its changes between 2003 and 2008. It was concluded that sugarcane area grew $1.88^{*}10^{6}$ ha being that 50 percent of this expansion was over pasture and annual crops areas. Although it had been the main initiative of using remote sensed data for sugarcane monitoring in Brazil, it was unfortunately discontinued.

In spite of Landsat's medium spatial resolution (30 m), its 16-day temporal frequency may lead to data missing due to the incapacity of optical sensors to collect ground data through clouds during rainy seasons. So, a strategy that could be adopted to minimize this problem is to use a sensor like MODIS, which has a higher temporal frequency (1-2 days) of data acquisition. In this case, it is important to combine MODIS spectral bands to generate data that could be used to assess land use (LU) with confidence [12]. Even with the advantages of time series over individual images, their use is still representing a challenge for LU monitoring because they need to be pre-processed in order to produce more accurate results [13].

- *J. P. Silva is with the Campinas State University, Campinas, SP 13083-875, Brazil. E-mail: joao.silva@feagri.unicamp.br.*
- *J. Zullo, Jr., is with the Campinas State University, Campinas, SP 13083-970, Brazil. E-mail: jurandir@cpa.unicamp.br.*
- *L. A. S. Romani is with the Brazilian Agricultural Research Corporation, Campinas, SP 13083-886, Brazil. E-mail: luciana.romani@embrapa.br.*

Manuscript received 31 Jan. 2018; revised 8 Sept. 2018; accepted 17 Apr. 2019. Date of publication 29 Apr. 2019; date of current version 29 Aug. 2020. (Corresponding author: J. P. da Silva.)
Recommended for acceptance by M. Datcu, J. Le Moigne, P. Soille, P. G. Marchetti, and G.-S. Xia.
Digital Object Identifier no. 10.1109/TBDATA.2019.2913402

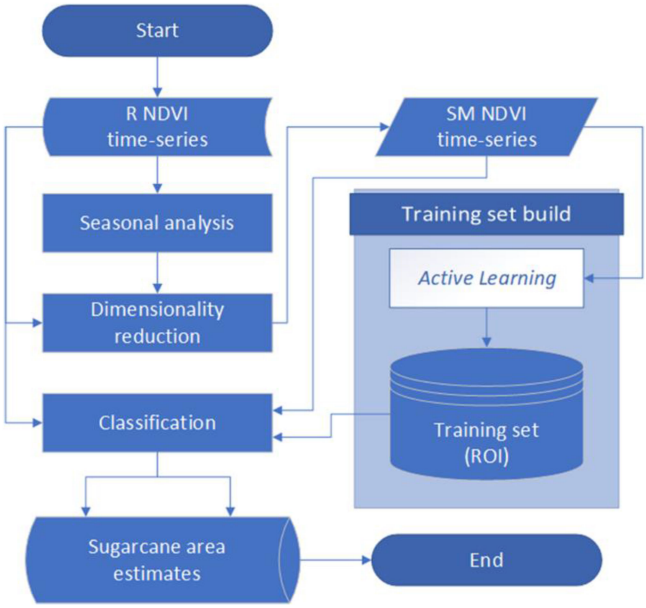

Fig. 1. Development workflow, which includes data acquisition and organization, pre-processing, training set build and time series classification.

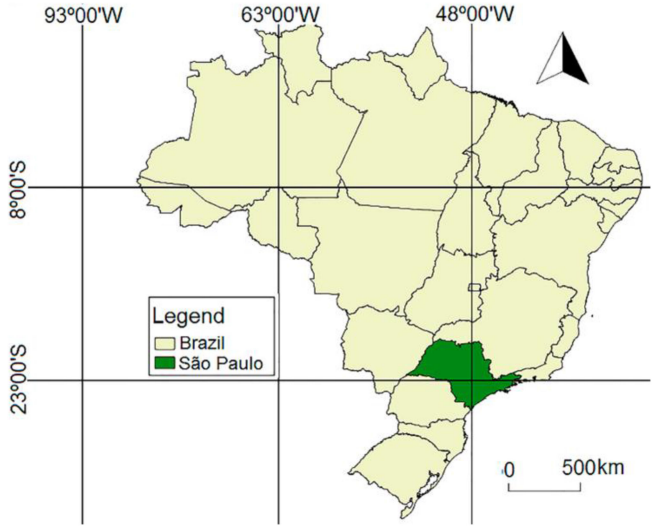

Fig. 2. São Paulo state and its localization in Brazil.

Some methodologies of LU monitoring do not present scalability, since they need more human resources to be executed, as can be seen in [11], where image interpretation and validation are executed with high human intervention. This condition limits the expansion of monitoring projects when human and financial resources are scarce. In addition, remote sensing data is becoming more available on quantity and quality, demanding efficient methods to explore all its potential.

Data mining classification techniques, for example, can be very useful in all this context of the use of remote sensing data in real applications such as the agricultural monitoring. Some studies proposed the use of Support Vector Machine (SVM) algorithm to classify remote data for applications like phenology [14] and phytosanitary assessment [15], [16] and crop areas monitoring [17], [18], whereas other [19], [20] focused on creating frameworks for analyzes efficiency improvement.

In classification, a key point is to build reliable training sets. Whereas it is traditionally a costly task [21], [22], it also must be done in order to express all relevant characteristics about data to enable algorithms differing between classes with confidence.

This way, [23] presents a review about AL, an iterative resampling approach [24] based on query strategies with the training set building process so that it can result in a 90 percent smaller set without accuracy loss [25]. In supervised classification, which depends on the quality of referenced (training) data to reach good results, AL does not focus on reaching training data that represents its entire space, but those samples that are the most informative for each class, what can be based on several heuristics, as described in [23].

The AL approach is being adopted in many remote sensing studies, mainly those related to hyperspectral images [26], [27], [28], [29], [30], [31], whereas there are few [32], [33], [34] studies aiming time series classification. In a multi-temporal data study, [32] proposed a framework that uses AL to optimize the training set based on class labels transferred from a domain (source) to n target images, where authors could reach up to 90 percent overall accuracy on image classification.

Débonnaire et al. [33] tested AL performance over traditional supervised learning methods on time series classification by different training set building heuristics. Results showed that the performance of each algorithm and the heuristic adopted depends on dataset's characteristics.

In this paper, we proposed a sugarcane cultivated area detection framework based on time series classification so that it presents scalability and efficiency by restricting human intervention to training set building phase and using algorithims to automate classifying process in order to reduce its operational costs.

2 Methodology

2.1 Overview

This study was divided into steps, as illustrated in Fig. 1. First, time series were organized and divided into two sets representing the same area and period: one submitted through seasonal modeling and another one used to test its effectiveness. Then, both seasonally modeled (SM) and raw (R) time series were submitted through dimensionality reduction to reduce its size and eliminate remaining noise that could harm final classification.

Using the AL approach, we built the training set by sampling regions of interest (ROIs), or pixels, from SM time series images and used it to train classifiers for both SM and R time series. Finally, we compared sugarcane area estimates for every harvest cycle considered in this study with the analogous INPE/CANASAT project [11] and other official estimates.

2.2 Study Area and Data Acquisition

We have chosen the state of São Paulo (Fig. 2) due to its leadership in sugarcane production sector as the national main producer. The study period comprises all sugarcane harvests occurred between 2006 and 2012. In this period, boosted by favorable economic conjunctures, large investments were made in sugarcane production sector, what resulted in a great sugarcane area expansion dynamic [1].

TABLE 1
Input Dataset Information

Parameter	Value
Data product	MOD13Q1 – NDVI; Version 5
Spatial resolution	Width: 250m Height: 250m
Sinusoidal tile grid	From h12, v10 to h14, v11
Spacial origin point	Longitude: -53.1101 degree, Latitude: -19.7797 degree
Image size	Width: 3580 pixels Height: 2214 pixels
Pixel size	0.0025 degree
Temporal resolution	16-day
Number of images	23 images/year; 230 images in total
Data acquisition period	From April, 2004 to March, 2014
Missing images	0
Coordinate system	WGS-84, geographic
Data value type	16-bit integer
Geometric correction	Already corrected
Atmospheric correction	Already corrected

MODIS measures the entire Earth surface every 1 to 2 days, but only the best ground image of each 16-days interval was used in order to minimize noise caused by cloud cover.
In sinusoidal tile grid values, h means horizontal quadrant and v means vertical quadrand.

We analyzed the Normalized Difference Vegetation Index (NDVI), a MODIS [35] product related to vegetation phenological behavior. This dataset was timely organized and synchronized with Brazilian center-south region sugarcane harvest cycle, which starts in April and lasts until March of the next year. Table 1 shows the dataset main chacacteristics.

2.3 Time Series Organization and Pre-Processing

The data pre-processing is a key point in data mining projects, being responsible for data cleaning and transforming in order to improve mining tasks results [36]. Here, we applied two major procedures: (i) noise detection and removal, followed by seasonal extraction on R time series to generate SM time series, and (ii) dimensionality reduction on both R and SM time series.

Initially, NDVI time series were composed of 23 images/year (Table 1), however, Timesat tool [37], used in step (i), works with full seasons, i.e., the center-most seasons. As sugarcane phenological behavior peaks in summer (December to March, in south hemisphere), a single yearly time series wouldn't allow Timesat to correctly detect its seasonality [37]. Due to this, we reorganized time series by joining 4 of then in a windowed form and used the two center-most ones as the target for seasonal extraction (Table 2). The reason for choosing the two center-most seasons is related to sugarcane harvest period extension. As sugarcane cycle lasts at least a year, seasonal extraction must be able to detect seasons that starts both in the beginning, e.g., April, and in the end, e.g., November, of a year. Therefore, for each image pixel, time series length increased from 23 NDVI values to 92. Also, its important to emphasize that cycles are not dependent on each other, since we conducted cycle analysis individually.

TABLE 2
Time Series Reorganization for Seasonal Modelling

Sugarcane cycle	Joined time series	Target seasons
1	2004 2005 2006 2007	2005 2006
2	2005 2006 2007 2008	2006 2007
3	2006 2007 2008 2009	2007 2008
4	2007 2008 2009 2010	2008 2009
5	2008 2009 2010 2011	2009 2010
6	2009 2010 2011 2012	2010 2011
7	2010 2011 2012 2013	2011 2012

Both in columns 2 and 3, each year refers to harvest starting point, e.g., target seasons (column 3) in sugarcane cycle 1, which refers to harvest period occurred between April, 2005 and March, 2006 and between April, 2006 and March, 2007.

TABLE 3
Timesat Processing Parameters

Function	Parameter	Value
STL	Stiffness	3
STL	Spyke method	STL replace
DL	Amplitude	0
DL	Seasonal parameter	1
DL	Envelope iterations	3
DL	Adaptation strength	5
DL	Season start/end	Seasonal amplitude
DL	Season start	10% of amplitude
DL	Season end	10% of amplitude

Noisy data can persist even after choosing the best image of each 16-day interval, and due to that, we applied, also on Timesat, the Seasonal Trend Decomposition (STL) [38] method in order to remove spikes and outliers from time series that could harm seasonal modelling. This method is based on locally weighted regression smoother (LOESS) and produces a reminder (or residual) parameter used to weight time series values and classify then as outliers or not [37]. As expressed in (1), a value $y(t)$ is classified as an outlier in two situations:

$$y\ (t) = \begin{cases} < \ (mean\ (y(t-1),\ y(t+1)) - c) \\ > \ (max\ (y(t-1),\ y(t+1)) + c), \end{cases} \tag{1}$$

where t refers to time; and c is the product of time series standard deviation and a stiffness factor (Table 3).

Yet in step (i), the seasonal extraction was applied to model vegetation phenological phenomena in response to its phases/seasons. The double-logistic (DL) function [37], [39] adopted here provides both data smoothing and seasonal parameters extraction from time series. Its general form is defined in (2):

$$\begin{aligned} g(t; x1, \ldots, x4) = 1/\left(1 + e^{(x1-t)/x2}\right) \\ - 1/\left(1 + e^{(x3-t)/x4}\right), \end{aligned} \tag{2}$$

where, t refers to time; $x1$ is the inflection point in the left side of the resulting curve spike; $x2$ is the changing rate in $x1$; $x3$ is the inflection point in the right side of the resulting curve spike; and $x4$ is the changing rate in $x3$.

We adjusted DL function hyperparameters in order to place its curve at the upper envelope of original NDVI time series curve. This procedure is important because most noise in remote sensed data is negatively biased [37]. Table 3

TABLE 4
Seasonal Parameters Extracted

Seasonal parameter	Description
Beginning	Season start
Ending	Season end
Length	Season lengh
Mid-season	Time to reach maximum NDVI value
Maximum	Maximum NDVI value in season
Base value	Average between minimum NDVI values
Amplitude	Between base and maximum values
Growth rate	Growth rate from season start to mid-season
Decline rate	Decline rate from mid-season to season end

Beginning, ending, length and mid-season parameters are related to time interval, i.e., each image or each 16-day, whereas the other ones are value-related. Base value refers to average between minimum NDVI values in the right and left curve sides.

shows parameters values defined for STL and DL functions application on Timesat.

Several seasonal parameters related to vegetation phenological dynamics are available on Timesat. Table 4 shows extracted parameters from NDVI time series. Other available parameters include integrated values and were not considered for this study.

Finally, in step (ii), we applied the Principal Components Analysis (PCA) [36], [40], [41] on both SM and R time series using ENVI 5.2 system. This method reduces the number of attributes of the original dataset in response to the variance contained in it. The original dataset attributes were then converted into eigenvectors with eigenvalues that represents their variance. Following experiments conducted in [42] and [43], we retained only eigenvectors with eigenvalues ≥ 0.7, what corresponds to approximately 90 percent of the variance contained in original datasets.

2.4 Active Learning – Training Set Building

To build the training set, we applied the AL approach by sampling, for each class, with ENVI 5.2 system, a balanced number of samples (ROIs) from SM time series images and evaluating their separability, i.e., how distinct is one class from the others, repeating this procedure until the optimal result (training set) was achieved (Fig. 3).

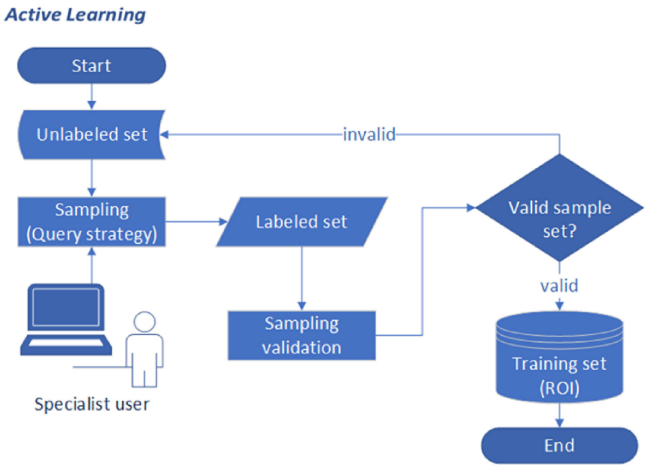

Fig. 3. Active learning workflow application for training set building.

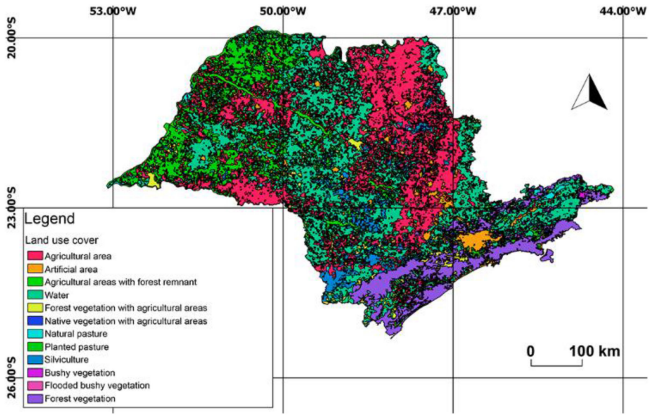

Fig. 4. São Paulo state LU map for 2010. Reference map used to help on training set building.

For each harvest cycle, a training set was created by sampling ROIs from the study area to represent different LU classes. In this step, covering maps (Fig. 4) from Brazilian Institute of Geography and Statistics (acronym IBGE, in Portuguese) [44] were used to subsidize the specialist user (Fig. 3) with information about different LU classes and their location.

Due to similarities between some classes of reference mapping (Fig. 4), we merged then aiming to reduce class redundancy and facilitate the sampling process (Table 5). Moreover, given that reference mapping (Fig. 4) treats both sugarcane and annual crops classes as one (agriculture), we (as the specialist user in Fig. 3) sampled those classes using also references [45] about their cultivated area assuming that there is a higher chance of correctly sample from places that are known to have bigger planted area for one class than for another.

To define the minimum training set sample size, we used the rule presented in [46] and defined in (3):

$$n \approx 30 * m * k, \tag{3}$$

where n is the training set sample size; m is the number of classes; and k is the number of attributes. In [47] this rule is referred to as a commom training set size dimensioning method in remote sensing studies.

As described in [23], the AL goal is to use a query strategy to sample, from the unlabeled set, the most informative data to build a small but robust training set and provide only essential information to classification algorithms. Hence, it's a key step to have a distance measurement of within- and between-class to assess the sampling process and also make obtained results feasible to relate with other studies [47]. So, we conducted the training set sample validation by applying two distance measurement methods: Jeffries-Matusita (JM) [41], which is often employed in remote sense studies using feature selection (and assessment), e.g., [47], [48] and [49], and (ii) Transformed Divergence (DT) [41], which is said to be almost as effective as JM [41]. These methods are asymptotic to 2.0, that is, a distance of 2.0 between two classes implies that a pixel (in our context, time series are represented by pixels) belongs to a class with 100 percent accuracy [41]. Due to that, both methods were applied aiming to ensure sampling quality.

TABLE 5
Merge Scheme of Reference Map Classes

Class number	Reference mapping [44]	Class number	LU cover class adopted (abbreviation)
1	Artificial area	1	Artificial area
2	Water	2	Water
3	Silviculture	3	Native vegetation, forest and silviculture
3	Forest vegetation	4	Agriculture – Annual crops
3	Forest vegetation with agricultural area	5	Agriculture – Sugarcane
3	Native vegetation with agricultural area	6	Pasture with native vegetation (Pasture)
4	Agricultural area (Annual crops)		
5	Agricultural area (Sugarcane)		
6	Natural pasture		
6	Planted pasture		
6	Agricultural area with forest remnant		
6	Bushy vegetation		
6	Flooded bushy vegetation		

Class number in column 1 refers to the correspondent class number in column 3 after merge reference mapping classes.

In [47] and [49], JM distance flew between 0.5 and near 2.0. In [48], JM distance reached 1.33 at maximum. In [50], the optimal point was achieved at 1.98. To be stricter, in this study we considered the sample valid when both JM and DT distances reached 95 percent (1.9) of the saturation point (2.0).

2.5 Classification

We have used the SVM algorithm to classify the SM and R time series because it usually has good performance in remote sensing applications and can generate satisfying results even with a compact training set [51], [52].

Originally, SVM is a binary classifier that defines a hyperplane to separate instances into two given classes, as described in (4):

$$w_0 + w_1 x_1 + w_2 x_2, \tag{4}$$

where each instance is classified by its x_i attributes weighted by w_i factor.

Although (4) defines the hyperplane, it does not give optimal class separation yet. So, to define it, SVM takes some key instances, also called support vectors, and use then to define the class separation optimal margins, as defined in (5) and (6):

$$H_1 : \; w_0 + w_1 x_1 + w_2 x_2 \geq +1 \tag{5}$$

$$H_2 : \; w_0 + w_1 x_1 + w_2 x_2 \leq -1, \tag{6}$$

where, H_i represents the optimal separation margins for classes $+1$ and -1 [36]. This support vector strategy explains SVM's robustness even when the training set is small and justifies its use in remote sensing applications, as can be seen in [53] revision.

Sometimes, classification does not have a linear solution. In this case, SVM uses an internal algorithm, also called kernel, to change data dimension and transform a non-linear separation problem into linear. In this paper, we used the Radial Basis Function (RBF), considered one of the best kernels available in literature [54], defined in (7):

$$K(X_i, X_j) \; = e^{-X_i - X_j{}^2 / 2\sigma^2}, \tag{7}$$

where, x_i and x_j are instance attributes.

We conducted the SVM hyperparameters configuration adopting the default values for most of then. Particularly, for *Classification probability threshold* (Table 6), we empirically defined its value as 0.5, i.e., the minimum probability required to SVM classify a pixel (a time series) is 50 percent. Table 6 shows hyperparameters configuration for SVM used in this study.

TABLE 6
SVM Hyperparameter List

Hyperparameter	Value	Value description
Kernel type	RBF	Presented in [54]
Gamma in kernel func. (γ)	$f\,(x) = 1/x$	$x =$ no. of bands.
Penalty parameter (C)	100.00	Default value
Pyramid levels	0.00	Default value
Classif. prob. threshold	0.5	Empirically defined

Configuration avalaible on ENVI 5.2 system. Gamma in kernel function value is the software default.

2.6 Validation

After classification, all classes, excepting sugarcane, (Table 5) were merged into a single class (no-sugarcane) to validate model's performance. Correspondent sugarcane area estimates from INPE/CANASAT project [11] were used as the test set (ground truth) to validade SVM classification for both SM and R time series. In this step, as

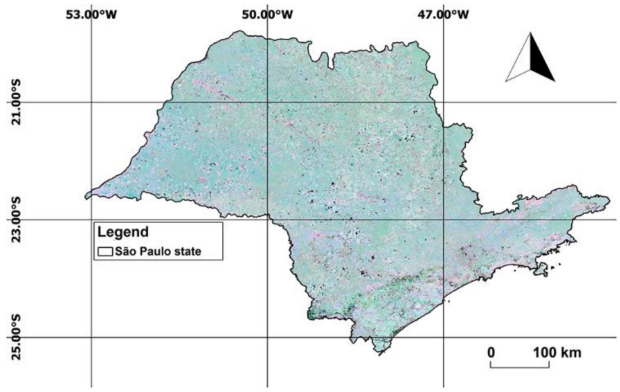

Fig. 5. Colored composition of time-based seasonal parameters: Start (Red), end (Green) and mid-season (Blue).

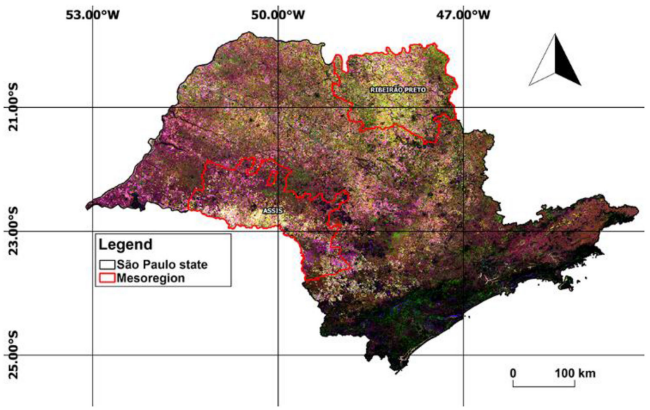

Fig. 6. Colored composition of value-based seasonal parameters: Amplitude (Red), growth rate (Green) and decline rate (Blue).

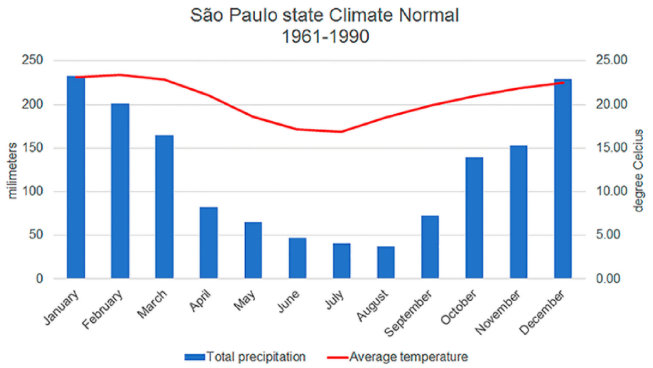

Fig. 7. São Paulo state Climate Normals (1961-1990).

opposed to [50], we performed a totally random sample from the test, set aiming to avoid any bias in model's validation, e.g., intentionally selecting regions with higher or lower sugarcane area concentration. The test set size was dimensioned to have, as usually employed, 20 percent of the training set size.

We also compared both SM and R time series area estimates with official government agencies like National Food Supply Company (acronym CONAB, in Portuguese), IBGE and INPE/CANASAT itself.

3 Results and Discussion

3.1 Seasonal Modeling

Seasonal extraction showed that value-based parameters performed higher vegetation cover contrast than time-based parameters. Temporal parameters colored compositions (Fig. 5) were more homogeneous when compared to value-based ones (Fig. 6), evidencing that these parameters are better to be used in LU studies, as it provides more information about different vegetation types.

Although temporal composition did not show differences between vegetation types, we observed that temporal seasonal parameters agree with 1961-1990 Climate Normals [55] (Fig. 7) observed for São Paulo state. The homogeneity of seasonal compositions using temporal related parameters means that all vegetation types have similar temporal behavior.

3.2 Dimensionality Reduction

PCA analysis of SM and R time series reduced the original dataset dramatically. The 9 original attributes of SM time series were transformed into 2 attributes keeping 87 percent of data variance in every harvest cycle. In case of R time series, PCA reduced from 46 attributes to 2 or even 1 attribute (in case of 2006 and 2009 cycles), keeping an average data variance of 93 percent of the original dataset. The number of retained eigenvector and variance in this study is similar to those observed in [56] and [57], which retained three (with 99 percent of the original variance) and one (with an accuracy of 90 percent in final classification) eigenvector, respectively.

3.3 Active Learning

The training set built resulted in a dataset with 1,200 sampled ROIs (200 sampled for each class) for each harvest cycle. According to (2), the minimum size for each harvest cycle should be 360. We tested SVM classification results with 180 samples for each period, keeping the 20 percent proportion of test over training samples defined before.

The class separability measurement satisfied the validation criteria for all SM time series on both JM and DT methods. However, even using the same areas to assess class separability, JM and DT distances satisfied our criteria in only 81 percent of cases for R time series. Without seasonal modeling, annual crops and sugarcane samples did not show satisfactory separability, presenting, in average, 0.77 for JM distance and 0.65 for DT distance.

In [47], comparing spectral distances between grass and grain crops, JM distances flew between 0.5 and almost 2.0, and although authors considered class separability as good for most class depending on pairwise combination and period of year, they didn't define a fixed distance threshold. In [48], assessing smoothing techiques effect in JM distance measurement for NDVI time series classification, its authors defined 1.41 as the optimal threshold.

TABLE 7
Model Validation of Seasonally Modeled (SM) Time Series for Minority (Sugarcane) Class Classification

Year	False positive rate	False negative rate	Recall	Specificity	Precision
2006	67.16%	48.84%	51.16%	80.00%	32.84%
2007	61.11%	70.21%	29.79%	90.05%	38.89%
2008	61.54%	62.26%	37.74%	85.12%	38.46%
2009	48.57%	71.88%	28.13%	91.67%	51.43%
2010	50.79%	54.41%	45.59%	84.00%	49.21%
2011	51.56%	53.03%	46.97%	83.66%	48.44%
2012	58.67%	56.34%	43.66%	77.66%	41.33%
Average	57.06%	59.57%	40.43%	84.59%	42.94%

TABLE 8
Model Validation of Raw (R) Time Series for Minority (Sugarcane) Class Classification

Year	False positive rate	False negative rate	Recall	Specificity	Precision
2006	80.00%	83.72%	16.28%	87.56%	20.00%
2007	65.57%	55.32%	44.68%	81.90%	34.43%
2008	50.00%	71.70%	28.30%	93.02%	50.00%
2009	67.42%	54.69%	45.31%	70.59%	32.58%
2010	42.86%	70.59%	29.41%	92.50%	57.14%
2011	56.99%	39.39%	60.61%	73.76%	43.01%
2012	61.86%	36.62%	63.38%	62.94%	38.14%
Average	60.67%	58.86%	41.14%	80.32%	39.33%

There, although tested the JM distance in classes similar to those tested in this study (Table 5), they reached a maximum JM distance of 1.3, what is 35 percent below the saturation point. Results for JM distance obtained in [49] shows that this measurement is highly dependent of spectral chacacteristics (in that case, crops' spectral signature and vegetal cycle, translated as NDVI values and time series length, respectvely).

3.4 Validation

The model validation with test samples showed that classification results using SM time series (Table 7) were a few better than those using R time series for minority (sugarcane) class (Table 8). In general, average Kappa index for SM time series classification was 0.25 against 0.21 from R time series, what is still weak, although it is still better than aleatory classification. Also, as mentioned before, we have performed a totally random sampling on test set in our experiments. In [50], the authors conducted a similar study and emphatize that they validated their results based on strategic test areas, what could've masked model's real performance. In our case, what could had happened is that we randomly sampled test ROIs from small sugarcane areas, as in São Paulo state exists large continuous sugarcane areas as well as small spread areas. Also, even been considered a reliable project, remainder noise on INPE/CANASAT [11] data, taken here as the test set, can still be present, as it involved visual image interpretation tasks which is susceptible to errors as well.

SVM applications in remote sensing focus most on hyperspectral classification, so there is a gap in time series classification using it [53]. In recent studies [18], [58], authors combine time series with other data sources in order to generate their results. In this paper, we have used only NDVI time series in our analysis. Carrão, Gonçalves and Caetano [59], classifying vegetation index time series, concluded that spectral diversity is more important than temporal diversity in order to obtain satisfactory results. Despite of these evidences, São Paulo area has a peculiar characteristic. As it is the major national sugarcane producer, it still has a considerable pasture area (Fig. 4), and this type of vegetation has similar spectral behavior to sugarcane [60], what could represent a challenge for LU assessment.

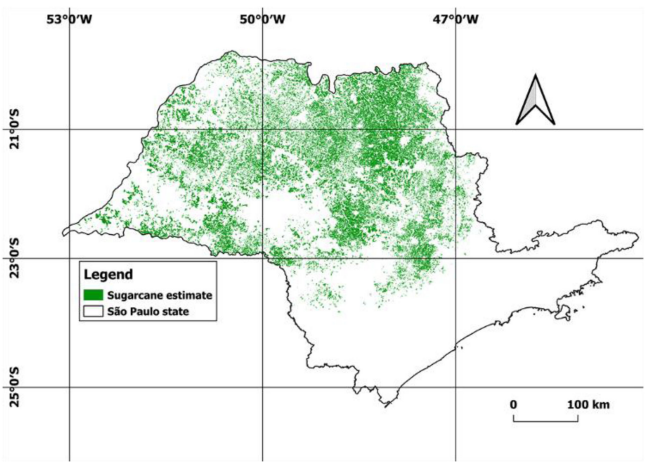

Fig. 8. INPE/CANASAT sugarcane planted area estimate for 2012 cycle.

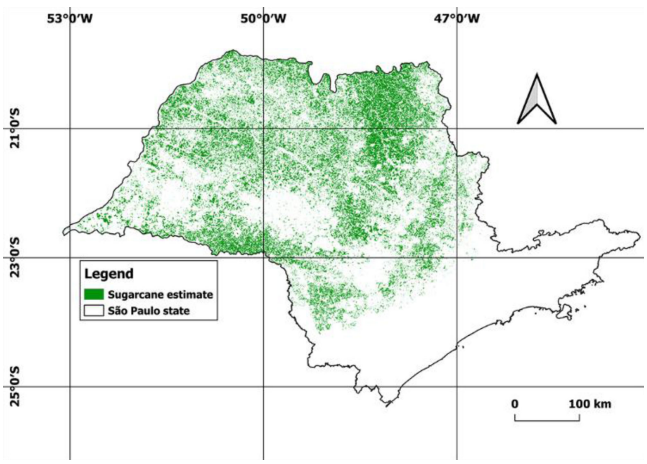

Fig. 9. SM time series sugarcane planted area estimate for 2012 cycle.

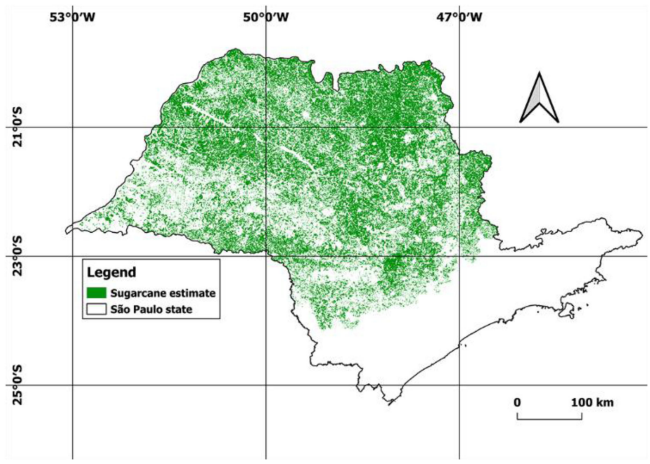

Fig. 10. R time series sugarcane planted area estimate for 2012 cycle.

TABLE 9
Sugarcane Area (in Hectare) Estimates Based on Different Methods

Year	Seasonally modeled (SM) time series	Raw (R) time series	INPE/CANASAT	CONAB	IBGE
2006	4,195,048	2,128,026	3,016,262	3,288,200	3,498,265
2007	2,890,060	4,193,774	3,289,761	3,824,200	3,890,414
2008	4,048,791	2,194,391	3,746,039	3,882,100	4,541,509
2009	2,485,076	6,060,411	4,562,179	4,129,900	4,977,077
2010	4,511,049	2,850,983	4,810,327	4,357,000	5,071,205
2011	4,298,490	6,590,394	4,664,610	4,370,100	5,216,491
2012	4,675,054	8,232,584	4,601,335	4,419,460	5,172,611
Average (1)	3,871,938	4,607,223	4,098,645	4,038,709	4,623,939
Average (2)	4,084,689	4,812,425	4,222,414	4,170,572	4,778,446

Average (1) considers all cycles. Average (2) discards 2006 and 2009 cycles.

Comparing the Pearson (r) correlation for 2006 to 2012 cyles, we assessed that SM time series estimates presented 23% correlation with both INPE/CANASAT and CONAB and 24% with IBGE, whereas R time series estimates presented 59% correlation with INPE/CANASAT, 63% with IBGE and a 67% with CONAB.

3.5 Sugarcane Area

Although results observed in spatial validation results (Tables 7 and 8), when we compared the total area estimates with other estimates and when we visually compared INPE/CANASAT mapping (Fig. 8) with generated maps, we realized similarities between then, more on SM time series (Fig. 9) than R time series (Fig. 10). Table 9 shows total sugarcane area in São Paulo based on both SM and R time series classification, INPE/CANASAT [11], CONAB [61] and IBGE [45] estimates, taken as official by both government and private agencies. For 2006 and 2009 cycles, respectively, SM time series estimates differ, in average, 29 and −45 percent from all official estimates (39 and −46 percent considering only INPE/CANASAT estimates). For these two periods, the dimensionality reduction step coincidentally retained just one eigenvector on R time series PCA application, what suggests that some noise from input data set [35] persisted even after applying seasonality extraction and dimensionality reduction. Discarding those cycles for both SM and R time series results, Pearson (r) correlation for SM time series estimates increased to 90 percent with INPE/CANASAT, 84 percent with CONAB and 95 percent with IBGE, whereas for R time series estimates, maximum correlation found is 67 percent.

4 Conclusion

The success of NDVI time series classification by SVM depends on the type of vegetation that covers the ground even when applied methodologies that assure class separation with confidence. Comparatively, it's the same as concluded in [33], where results vary due to the characteristics of the dataset. SVM was more sensible to training set quality than its size. This characteristic, aligned with a preprocessing step related to input data characteristic, confirms the potential of SVM to be used in remote sensing predictive tasks.

The AL approach made it possible to build a compact but efficient dataset as it represents reduction of training set construction efforts. Common methodologies are costly in training set build because they demand large amounts of data. This way, the use of AL in training set building can improve model's performance and efficiency by reducing its operational costs and adding technical criteria for training instances selection.

The sugarcane area estimates by seasonal extraction and classification generated results that are similar to other methodologies estimates. This kind of product can be used along with those methodologies in order to improve and assess model's performance.

In this paper, we have used only NDVI time series in our analysis, but adopting other vegetation index (e.g., Enhanced Vegetation Index – EVI) could help to improve our results and increase model's accuracy.

Aknowledgments

This work was submitted for review on Jan 31, 2018 and resubmitted after peer review on Sep 07, 2018. It was supported in part by Brazilian Government. The authors want to thanks V. R. Pereira for given advices about techniques and for text reviewing. The MOD13Q1 was retrieved from the online tool Reverb, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. Nevertheless, Reverb tool was discontinued on Jan 01, 2018 and EOSDIS data is now accessible on Earthdata Search portal (https://go.nasa.gov/2oLQiEm).

References

[1] União das Indústrias de Cana-de-Açúcar, "Dados da produção canavieira - Área total por estado," UNICADATA, [Online]. Available: https://bit.ly/2ccgVwx, 2016.
[2] G. M. S. Câmara, and E. A. M. Oliveira, ed.1, *Produção de cana-de-açúcar*, FEALQ, 1993.
[3] J. A. Marengo, ed.2, *Mudanças climáticas globais e seus efeitos sobre a biodiversidade: caracterização do clima atual e definição das alterações climáticas para o território brasileiro ao longo do século XXI*. Ministry of Environment, 2007.
[4] H. S. Pinto and E. D. Assad, "O Papel do Agronegócio da Economia Brasileira," *Aquecimento global e a nova geografia da produção agrícola no Brasil*, J.G. Deconto, ed.1, Campinas: Brazilian Agricultural Research Corporation, pp. 14–17, 2008.
[5] F. R. Marin and D. S. P. Nassif, "Mudanças climáticas e a cana-de-açúcar no Brasil: Fisiologia, conjuntura e cenário futuro," *Revista Brasileira de Engenharia Agrícola e Ambiental*, vol. 17, no. 2, pp. 232–239, [Online]. Available: https://bit.ly/2PZlD2J, Jan. 2013.
[6] S. J. Wand, G. Midgley, M. H. Jones, and P. S. Curtis, "Responses of wild C4 and C3 grass (Poaceae) species to elevated atmospheric CO_2 concentration: A meta-analytic test of current theories and perceptions," *Global Change Biol.*, vol. 5, no. 6, pp. 723–741, [Online]. Available: https://bit.ly/2LRUAmK, Aug. 1999.

[7] A. P. Souza, M. Gaspar, E. A. Silva, E. C. Ulian, A. J. Waclawovsky, R. V. Santos, M. M. Teixeira, G. M. Souza, and S. Buckeridge, "Elevated CO_2 increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane," *Plant Cell Environ.*, vol. 31, no. 8, pp. 1116–1127, [Online]. Available: https://bit.ly/2PBjysR, Aug. 2008.

[8] D. B. Leakey, M. Uribelarrea, E. A. Ainsworth, S. L. Naidu, A. Rogers, D. R. Ort, and S. P. Long, "Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO_2 concentration in the absence of drought," *Plant Physiology*, vol. 140, no. 2, pp. 779–790, [Online]. Available: https://bit.ly/2oB7hJi, Feb. 2006.

[9] G. W. Wall, T. J. Brooks, N. R. Adam, A. B. Cousins, B. A. Kimball, P. J. Pinter Jr., R. L. Lamorte, J. Triggs, M. J. Ottman, S. W. Leavitt, and A. D. Matthias, "Elevated atmospheric CO_2 improved sorghum plant water status by ameliorating the adverse effects of drought," *New Phytologist*, vol. 152, no. 2, pp. 231–248, [Online]. Available: https://bit.ly/2CbPCS2, Nov. 2001.

[10] R. R. V. Gonçalves, P. P. Coltri, A. M. H. Avila, L. A. S. Romani, J. Zullo Jr., and H. S. Pinto, "Análise comparativa do clima atual e futuro para avaliar a expansão da cana-de-açúcar em São Paulo," in *Proc. 17th Congresso Brasileiro de Agrometeorologia*, Jul. 2011, pp. 1–5.

[11] B. F. T. Rudorff, D. A. Aguiar, W. F. Silva, L. M. Sugawara, M. Adami, and M. A. Moreira, "Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data," *Remote Sens.*, vol. 2, no. 4, pp. 1057–1076, [Online]. Available: https://bit.ly/2L4cG4X, Apr. 2010.

[12] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, [Online]. Available: https://bit.ly/2oCDL6a, Nov. 2006.

[13] P. R. Meneses, T. D. Almeida, A. N. D. C. S. Rosa, E. E. Sano, E. B. D. Souza, G. M. D. M. Baptista, and R. S. Brites, ed.1, *Introdução ao processamento de imagens de sensoriamento remoto*. CNPq, pp. 34–41, 2012.

[14] Y. L. Everingham, K. H. Lowe, D. A. Donald, D. H. Coomans, and J. Markley, "Advanced satellite imagery to classify sugarcane crop characteristics," *Agronomy Sustainable Develop.*, vol. 27, no. 2, pp. 111–117, [Online]. Available: https://bit.ly/2wClL04, Jun. 2007.

[15] R. K. Dewi and R. V. H. Ginardi, "Feature extraction for identification of sugarcane rust disease," in *Proc. Int. Conf. Inf. Commun. Technol. Syst.*, Sep. 2014, pp. 99–104, doi: 10.1109/ICTS.2014.7010565.

[16] E. K. Ratnasari, M. Mentari, R. K. Dewi, and R. V. H. Ginardi, "Sugarcane leaf disease detection and severity estimation based on segmented spots image," in *Proc. Int. Conf. Inf. Commun. Technol. Syst.*, Sep. 2014, pp. 93–98, doi: 10.1109/ICTS.2014.7010564.

[17] J. Zhang, L. Feng, and F. Yao, "Improved maize cultivated area estimation over a large scale combining MODIS–EVI time-series data and crop phenological information," *J. Photogrammetry Remote Sens.*, vol. 50, no. 9, pp. 102–113, [Online]. Available: https://bit.ly/2wDh59Z, Aug. 2012.

[18] D. Xie, P. Sun, J. Zhang, X. Zhu, W. Wang, and Z. Yuan," Autumn crop identification using high-spatial-temporal resolution time-series data generated by MODIS and Landsat remote sensing images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 2118–2121, doi: 10.1109/IGARSS.2014.6946884.

[19] J. Li, "Remote sensing image information mining with HPC cluster and DryadLINQ," in *Proc. 49th Annu. Southeast Regional Conf.*, Mar. 2011, pp. 227–232, doi: 10.1145/2016039.2016099.

[20] S. Niazmardi, S. Homayouni, H. Mc Nairn, J. Shang, and A. Safari, "Multiple kernels learning for classification of agricultural time-series data," in *Proc. 3rd Int. Conf. Agro-Geoinformatics*, Aug. 2014, pp. 1–4, doi: 10.1109/Agro-Geoinformatics.2014.6910640.

[21] J. Gruijter, D. J. Brus, M. F. Bierkens, and M. Knotters, *Sampling for Natural Resource Monitoring*. Berlin, Germany: Springer, 2006, pp. 1–334.

[22] W. G. Müller. *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*. Linz, Upper Austria: Springer, 2007, pp. 1–242.

[23] M. M. Crawford, D. Tuia, and H. L. Yang, "Active learning: Any value for classification of remotely sensed data?," in *Proc. IEEE*, vol. 101, no. 3, pp. 593–608, Mar. 2013, [Online]. Available: https://bit.ly/2CdRIB5.

[24] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification," New York, NY, USA: Wiley, 2001, pp. 1–680.

[25] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, [Online]. Available: https://bit.ly/2wES0e5, Jul. 2009.

[26] S. Patra and L. A. Bruzzone, "A fast cluster-assumption based active-learning technique for classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1617–1626, May 2011, [Online]. Available: https://bit.ly/2Cndi6m.

[27] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, [Online]. Available: https://bit.ly/2LZCT4Q, Sep. 2011.

[28] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011, [Online]. Available: https://bit.ly/2Q3aoWT.

[29] J. Li, J. M. Bioucas-dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Jul. 2013, [Online]. Available: https://bit.ly/2LSyXCI.

[30] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217–2223, Apr. 2014, [Online]. Available: https://bit.ly/2oyGaia.

[31] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Oct. 2011, [Online]. Available: https://bit.ly/2wBFAok.

[32] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1014–1031, May 2012, [Online]. Available: https://bit.ly/2LRsnw9.

[33] N. Débonnaire, A. Stumpf, and A. Puissant, "Spatio–temporal clustering and active learning for change classification in satellite image time series," *IEEE J. Sel. Topics Appl. Earth Observation Remote Sens.*, vol. 9, no. 8, pp. 3642–3650, Mar. 2016, [Online]. Available: https://bit.ly/2wJpBUu.

[34] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time-series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 300–312, Jan. 2013, [Online]. Available: https://bit.ly/2NesAhL.

[35] K. Didan, "MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC," doi: https://doi.org/10.5067/MODIS/MOD13Q1.006.

[36] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Waltham, MA, USA: Elsevier, 2011, pp. 1–703.

[37] L. Eklundh and P. Jönsson, *TIMESAT 3.3 with Seasonal Trend Decomposition and Parallel Processing*, vol. 1, Lund, Lund University Publications, 2017.

[38] R. B. Cleveland, W. S. Cleveland, J. E. Mc Rae, and I. Terpenning, "A seasonal-trend decomposition procedure based on loess," *J. Official Statist.*, vol. 6, no. 1, pp. 3–73, Jan. 1990, [Online]. Available: https://bit.ly/2wDzFOB.

[39] P. S. Beck, C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 321–334, Feb. 2006, [Online]. Available: https://bit.ly/2PzwcIS.

[40] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 1986, pp. 115–128.

[41] J. A. Richards, "The principal components transformation," *Remote Sensing. Digital Image Analysis: An Introduction*, J. A. Richards eds., Berlin, Germany: Springer, 2012, pp. 163–186.

[42] I. T. Jolliffe, "Discarding variables in a principal component analysis I: Artificial data," *Appl. Stat.*, vol. 21, no. 2, pp. 160–173, 1972, [Online]. Available: https://bit.ly/2wE1KFn.

[43] I. T Jolliffe, "Discarding variables in a principal component analysis. II: Real data," *Appl. Stat.*, vol. 22, no. 8, pp. 21–31, 1973, [Online]. Available: https://bit.ly/2NL8Cbu.

[44] Instituto Brasileiro de Geografia e Estatística, "Cobertura e Uso da Terra - vetores," [Online]. Available: https://bit.ly/2LWFVGZ, 2017.

[45] Instituto Brasileiro de Geografia e Estatística, "Produção agrícola municipal," *Banco de Dados Sidra*, [Online]. Available: https://bit.ly/2NN714S. 2017.

[46] M. J. Canty, *Image Analysis, Classification and Change Detection in Remote Sensing: with Algorithms for ENVI/IDL*. Boca Raton, FL, USA: CRC Press, Aug. 2006, pp. 1–576.
[47] T. G. Van Niel, T. R. McVicar, and B. Datt, "On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification," *Remote Sens. Environ.*, vol. 98, no. 4, pp. 468–480, 2005, [Online]. Available: https://bit.ly/2oGNyrI.
[48] Y. Shao, R. S. Lunetta, B. Wheeler, J. S. Liames, and J. B. Campbell, "An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data," *Remote Sens. Environ.*, vol. 174, pp 258–265, Mar. 2016, [Online]. Available: https://bit.ly/2M3Y6KC.
[49] P. Hao, Y. Zhan, L. Wang, Z. Niu, and M. Shakir, "Feature selection of time series MODIS data for early crop classification using random forest: A case study in kansas, USA," *Remote Sens.*, vol. 7, no. 5, pp. 5347–5369, [Online]. Available: https://bit.ly/2NOHkks, Apr. 2015.
[50] H. Cai and S. Zhang, "Regional land cover classification from MODIS time-series and geographical data using support vector machine," in *Proc. IEEE Youth Conf. Inf. Comput. Telecommun.*, Nov. 2010, pp. 102–105, doi: 10.1109/YCICT.2010.5713055.
[51] G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification," *Remote Sens. Environ.*, vol. 93, no. 1, pp. 107–117, Oct. 2004, [Online]. Available: https://bit.ly/2LT7QYi.
[52] P. Mantero, G. Moser, and S. B. Serpico, "Partially supervised classification of remote sensing images through SVM-based probability density estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 559–570, Feb. 2005, [Online]. Available: https://bit.ly/2NKVkM0.
[53] G. Mountrakis, J. IM, and C. Ogole, "Support vector machines in remote sensing: A review," *J. Photogrammetry Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011, [Online]. Available: https://bit.ly/2NJBuB0.
[54] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," in *Int. J. Appl. Earth Observation Geoinformation*, vol. 11, no. 5, pp. 352–359, Oct. 2009, [Online]. Available: https://bit.ly/2PC09bh.
[55] Instituto Nacional de Meteorologia, "Dados meteorológicos," *Normais Climatológicas*, [Online]. Available: https://bit.ly/2wGUKsa. 2017.
[56] T. I. R. Almeida, N. C. Penatti, L. G. Ferreira, A. E. Arantes, and C. H. Amaral, "Principal component analysis applied to a time-series of MODIS images: the spatio-temporal variability of the Pantanal wetland, Brazil," *Wetlands Ecology Manage.*, vol. 23, no. 4, pp. 737–748, Feb. 2015, [Online]. Available: https://bit.ly/2oBKgGb.
[57] S. Réjichi and F. Chaabane, "Feature extraction using PCA for VHR satellite image time-series spatio-temporal classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 485–488, doi: 10.1109/IGARSS.2015.7325806.
[58] Z. Xue, P. Du, and L. Feng, "Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series," *IEEE J. Sel. Topics. Appl. Earth Observations Remote Sens.*, vol. 7, no. 4, pp. 1142–1156, Jan. 2014, [Online]. Available: https://bit.ly/2Cgsvpu.
[59] H. Carrão, P. Gonçalves, and M. Caetano, "Contribution of multispectral and multitemporal information from MODIS images to land cover classification," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 986–997, Mar. 2008, [Online]. Available: https://bit.ly/2LTHOnC.
[60] A. C. Xavier, B. F. T. Rudorff, L. M. S. Berka, and M. A. Moreira, "Multi-temporal analysis of MODIS data to classify sugarcane crop," *Int. J. Remote Sens.*, vol. 27, no. 4, pp. 755–768, Feb. 2006, [Online]. Available: https://bit.ly/2MLLNb5.
[61] Companhia Nacional do Abastecimento, "Séries históricas," *Cana-de-Açúcar—Área Total*, [Online]. Available: https://bit.ly/2Y6t4IF. 2017.

João Paulo da Silva received the MSc degree in agricultural engineering from the University of Campinas, in 2017 and the graduate degree in agronomy from the Federal University of São Carlos, in 2015. He is currently working toward the PhD degree in the University of Campinas. He's interested in issues related to agricultural system improvement and monitoring using remote sensing techniques, agrometorology, zoning and data mining.

Jurandir Zullo, Jr. received the graduate degree in applied mathematics, in 1985 and agricultural engineering, in 1987. He received the master's degree in operational research from the Department of applied mathematics, in 1990, and the PhD degree in computer engineering and automation from the Faculty of Electric Engineering, in 1994, from the University of Campinas (UNICAMP), Brazil. He has been a researcher at the Center for Meteorological and Climatic Researches in Agriculture (CEPAGRI/UNICAMP) since 1987, and was its director from August 2002 to August 2008. He has experience in agricultural engineering, focusing on agrometeorology, remote sensing, image processing, agricultural zoning and climate change.

Luciana Alvim S. Romani received the BSc degree in computer science from the Federal University of São Carlos, São Carlos, Brazil, in 1993, the MSc degree in computer science from the State University of Campinas, Campinas, Brazil, in 2000, and, the PhD degree in computer science from the University of São Paulo, São Carlos, Brazil, in 2010. She has been a researcher with the Brazilian Agricultural Research Corporation since 1994. Her research interests include remote sensing applied to agriculture, data mining, information visualization, and human–computer interaction.