High-resolution map of sugarcane cultivation in Brazil using a phenology-based method

Yi Zheng[1], Ana Cláudia dos Santos Luciano[2,3], Jie Dong[4], Wenping Yuan[1*]

[1]School of Atmospheric Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Zhuhai 519082, Guangdong, China
[2]Department of Biosystems Engineering, College of Agriculture "Luiz de Queiroz" (ESALQ), University of Sao Paulo, Sao Paulo, P.O. Box 9 Av. Padua Dias 11, 13418-900 Piracicaba-SP, Brazil
[3]Interdisciplinary Center on Energy Planning (NIPE), UNICAMP, 13083-896 Campinas, Brazil
[4]College of Geomatics & Municipal Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, Zhejiang, China

Correspondence to: Wenping Yuan (yuanwpcn@126.com).

Abstract. Sugarcane is the most important source of sugar, and its cultivation area has undergone rapid expansion, replacing other crops, pastures, and forests. Brazil is the world's largest sugarcane producer and contributed to approximately 38.6% of the world's total production in 2019. Sugarcane in Brazil can be harvested from April to December in south-central area and from September to April in northeast area. The flexible phenology and harvest conditions of sugarcane in Brazil make it difficult to identify the harvest area at state to country scales. In this study, we developed a phenology-based method to identify the harvest area of sugarcane in Brazil by incorporating the multiple phenology conditions into a time-weighted dynamic time warping method (TWDTW). Then, we produced annual 30-m spatial resolution sugarcane harvest maps (2016-2019) for 14 states in Brazil (over 98% of the harvest area) based on the proposed method by using Landsat-7/8 and Sentinel-2 optical data. The proposed method performed well in identifying sugarcane harvest area with limited training sample data. Validations for the 2018 harvest year displayed high accuracy, with user's, producer's, and overall accuracies of 94.35%, 87.04%, and 91.47% in Brazil, respectively. In addition, the identified harvest area of sugarcane exhibited good correlations with the agricultural statistical data provided by the Brazilian Institute of Geography and Statistics (IBGE) at the municipality, microregion, and mesoregion levels. The 30-m Brazil sugarcane harvest maps can be obtained at https://doi.org/10.6084/m9.figshare.14213909 (Zheng et al., 2021).

1 Introduction

Sugarcane (*Saccharum officinarum*) is a semi-perennial crop that can be cut multiple times throughout several years in tropical and subtropical areas (Abdel-Rahman and Ahmed, 2008; Sindhu et al., 2016; Mulyono et al., 2017). Sugarcane is an important sugar and energy crop. Over 70% of world sugar comes from sugarcane (Brar et al., 2015; Iqbal et al., 2015). Sugarcane ethanol is an alternative energy source that can reduce CO_2 emission resulting from fossil fuel use (Bordonal et al., 2015; Jaiswal et al., 2017). In addition, sugarcane leaves can be used as fodder and be manufactured into paper and bioenergy (Leal et al., 2013). Sugarcane accounts for approximately 20% of global crop production over the period from 2000–2018, and this value is almost twice the share of maize, the second most produced crop worldwide (FAO, 2020). Sugarcane has now become a crop with high socio-economic importance for Brazil and other producers, such as India, China, Thailand, Pakistan, the United States, and Australia (Monteiro et al., 2018). Recently, sugarcane has undergone rapid expansion by occupying the land of other crops, pastures, and forests (Adami et al., 2012; Ferreira et al., 2015; Wachholz de Souza et al., 2017; Defante et al., 2018). The expansion of sugarcane production areas can influence regional land cover use, water use, greenhouse gas emission, soil carbon balance, and climate change (Loarie et al., 2011; Mello et al., 2014; Zhang et al., 2015; Jaiswal et al., 2017). On the one hand, the replacement of other crops with sugarcane may directly affect on agriculture and food security (Mello et al., 2014; Jaiswal et al., 2017). On the other hand, sugarcane expansion may influence the local climate by altering surface albedo and evapotranspiration (Loarie et al., 2011). Additionally, sugarcane has a large amount of water requirement and is often planted in the areas where water is limiting, therefore, sugarcane expansion may cause concerns about water security (Zhang et al., 2015). Timely and accurate estimates of the distribution, harvest area and growing conditions of sugarcane are crucial for sustainable sugarcane production and national food security.

Recently, single-date or time series moderate to high resolution remote sensing optical data (e.g., Landsat, Sentinel-2, SPOT, and CBERS) and synthetic aperture radar (SAR) data (e.g., PALSAR and Sentinel-1) have been used for crop mapping based on unsupervised and supervised classification methods (Lin et al., 2009; Johnson et al., 2014; Li et al., 2015; Belgiu et al., 2018). The most current and popular classification method is machine learning, such as random forest (Zhou et al., 2015; Luciano et al., 2018, 2019; Wang et al., 2019), support vector machine (Johnson et al., 2014; Zheng et al., 2015), and neural networks (Cai et al., 2018). For sugarcane identification, Junior et al. (2017) mapped the area of sugarcane in Paraná state in Brazil based on supervised maximum likelihood classification using Landsat/TM/OLI and IRS/LISS-3 images. Depending on the sugarcane map from 2009-2014 generated by the Canasat Project (Rudorff et al., 2010) as ground truth sample data, a random forest classification model was calibrated at 10 sites located across São Paulo state and employed to the entire state in 2015, showing the ability to create spatial and time generalization models (Luciano et al., 2019). Jiang et al. (2019) identified sugarcane in Zhanjiang city in China using machine learning methods and Sentinel-1A/2 time series satellite data. However, these methods strongly depend on extensive training samples, which are time-consuming and labor-intensive to obtain at the state and country scales (Dong et al., 2020a; Wang et al., 2020). For example, the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) produced the 30-m resolution Cropland Data Layer (CDL) product using a

decision tree classification method, and this approach mainly relied on the large volume of USDA Common Land Unit (CLU) data to establish training samples that are updated annually (Boryan et al., 2011). Only in Nebraska did the CLD product need more than 250,000 CLU records for training the classification method.

Phenology-based algorithms are also commonly used in regional classification (Wardlow et al., 2007; Zhong et al., 2014; Massey et al., 2017; Dong et al., 2020b). These methods have been proposed and developed based on the crop calendar, e.g., the germination, tillering, grand growth, and ripening phases. Li et al. (2015) identified sugarcane on the Leizhou Peninsula in China by comparing the polarization features (such as scattering angle and polarization entropy) of sugarcane with those of other land use types in the early, middle, and late tillering periods using TerraSAR-X images. The result indicated that the tillering period is a suitable growing phase that can be used for sugarcane cultivation area mapping. Mulyono et al. (2017) identified sugarcane plantations in the Magetan district of East Java Province in Indonesia using support vector machine and the crop phenology profile of enhanced vegetation index (EVI) time series derived from Landsat 8 images. Phenology-based methods would be a potential alternative approach to identify crop cultivation areas at the country scale with a low volume of training samples (Dong et al., 2020a). However, these studies were developed based on several phenological thresholds (e.g., green-up date, senescence date, and length of growing season) of crops in different growing stages, which need to be calibrated when extended to other regions or years. Therefore, traditional phenology-based methods are still insufficient to perform ideal mapping and may be limited by multiple thresholds when applied to large scales, such as the country to continental scales.

Dynamic time warping (DTW) is an effective phenology-based method in crop classification at large scales that compares the differences in seasonal variations in the vegetation index of a target crop with those of other crop types and natural vegetation. The original DTW method was developed for speech recognition and then employed for phenology-based classification using time series satellite images (Sakoe and Chiba, 1978; Petitjean et al., 2012; Petitjean and Weber, 2014). However, the DTW neglects the temporal ranges when searching the best alignment between two time series. Maus et al. (2016) proposed a time-weighted version of the DTW method (TWDTW) by adding a temporal weight that accounts for the seasonality of crops into the DTW method to balance shape matching and phenological changes. The TWDTW method performed well in identifying winter wheat, crop, and forest with limited training data (Maus et al., 2016; Belgiu et al., 2018; Manabe et al., 2018; Dong et al., 2020a).

Brazil is the world's largest sugarcane production country and contributed to 37.6% of the world's total harvest area in 2019, followed by India (18.9%), Thailand (6.9%), China (5.3%), Pakistan (3.9%), and Mexico (3.9%) (FAO, 2020). Sugarcane in Brazil is mainly located in south-central and northeast areas. Sugarcane is a semi-perennial crop. Its life cycle begins with the planting of a stem cutting and grows for approximately 12 months or 12-18 months depending on the season, variety, and region of planting (Rudorff et al., 2010). Generally, in the south-central area, sugarcane can be harvested from April to December with a harvesting season spanning 9 months. In the northeast areas, sugarcane can be harvested from September to April in the next year with a harvesting season spanning 8 months. The flexible phenology and harvest conditions of sugarcane in Brazil make it difficult to identify the harvest area at large scales, particularly at the state to country scales. In this study, based on Landsat and Sentinel-2 satellite data, we proposed a phenology-based method by incorporating multiple phenological

conditions of sugarcane into the TWDTW method to automatically identify the harvest area of sugarcane in Brazil at a spatial resolution of 30-m from 2016 to 2019.

2 Data and methods

2.1 Study area

In Brazil, sugarcane is mainly located in the south-central and northeast regions, particularly in the state of São Paulo. In our study, we mapped the harvest area of sugarcane in 14 states in Brazil, including São Paulo (SP, accounting for 55.1% of the harvest area in Brazil), Goiás (GO, 9.5%), Minas Gerais (MG, 9.2%), Mato Grosso do Sul (MS, 6.8%), Paraná (PR, 6.1%), Mato Grosso (MT, 2.8%), Bahia (BA, 0.7%), Rio de Janeiro (RJ, 0.5%), and Espírito Santo (ES, 0.4%) in south-central Brazil; and Alagoas (AL, 2.8%), Pernambuco (PE, 2.4%), Paraíba (PB, 1.0%), Rio Grande do Norte (RN, 0.6%), and Sergipe (SE, 0.4%) in northeast Brazil (Fig. 1). The mapped areas constitute over 98% of the Brazilian sugarcane harvest area.

<<Figure 1>>

2.2 Datasets

2.2.1 Satellite data

In this study, all the available Landsat-7/8 and Sentinel-2 reflectance data from July 2015 to February 2020 were used to generate 16-day composite normalized difference vegetation index (NDVI) series at the Google Earth Engine (GEE) platform. Landsat-7 and Landsat-8 data had a 30-m spatial resolution and a 16-day temporal resolution. The quality band BQA was used to remove pixels contaminated by clouds. The Sentinel-2 data had a 10-m spatial resolution and a 5-day temporal resolution. The quality band QA60 was used to remove pixels contaminated by clouds. NDVI was computed using the reflectance data from the near infrared band (ρNIR) and red band (ρRED):

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

The cloud-free frequencies of the 16-day composite NDVI in a year for each pixel are shown in Fig. 2. For 2018 and 2019, most of the pixels had more than 20 times of good observations. For 2016 and 2017, the pixels with more than 20 times of good observations were less, while most of the pixels had more than 18 times of good observations (Fig. 2). To reconstruct the near real-time temporal profile of the 16-day NDVI, we filled the gaps in NDVI time series using a linear interpolation method and then filtered the NDVI curves.

<<Figure 2>>

2.2.2 Sample data

The sample data used for calibration and validation in this study were mainly obtained based on the high-resolution images from Google Earth. We selected the samples for sugarcane and non-sugarcane according to the following rules. First, we

selected the samples for sugarcane or non-sugarcane using visual interpretations according to color and textures of the images from Google Earth. As shown in Fig. 3a, sugarcane exhibits unique color and textures on the high-resolution images from Google Earth. Sugarcane has coarser surface than most of the crops and smoother surface than forest or trees in growing season, which can be used to separate sugarcane from these types. Second, we cross checked and confirmed each of these samples using its NDVI timeseries. Sugarcane has a long life cycle ranging from 12 to 18 months (Rudorff et al., 2010), which is longer than most of the crops (Fig. 3b). And the sharply decreased NDVI in sugarcane harvest period can separate sugarcane from forest and pasture (Fig. 3b). We only use the samples which can satisfy the above two criteria. Finally, we collected totally 2909 samples with 1393 for sugarcane and 1516 for non-sugarcane in the year 2018 (Fig. 1).

<<Figure 3>>

2.2.3 Agricultural statistical data

The agricultural statistical data for the sugarcane harvest area were derived from the Municipal Agricultural Production (PAM) provided by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE; www.ibge.gov.br). The PAM was carried out yearly across the entire country with the statistical data at the Brazil, major region, federation unit, mesoregion, microregion and municipality levels. It provides information related to the plant area and harvest area, production and average yield, and average price in the reference year for 64 agricultural products. In our study, we used the harvest area from 14 major sugarcane planted states at the federation unit, municipality, microregion, and mesoregion levels from 2016 to 2018 (the boundary lines of these regions are shown in Fig. 4).

<<Figure 4>>

2.3 Methods

In this study, we employed a phenology-based method, namely the TWDTW method, to identify the harvest area of sugarcane in Brazil. The workflow was as follows: (1) pre-processing of the satellite dataset (i.e., Landsat 7/8 and Sentinel-2) to obtain the 16-day composite NDVI series, including low quality data removing (e.g., the elimination of clouds, cloud shadows, and Landsat 7 ETM+ Scan Line Corrector (SLC-off) gaps), NDVI compositing, NDVI gap filling and data filtering; (2) extraction of the standard NDVI curves of sugarcane in Brazil; (3) model development and sugarcane harvest area identification in Brazil based on the TWDTW method; and (4) assessment of the mapping accuracy of the Brazil sugarcane harvest area.

2.3.1 Time-weighted dynamic time warping (TWDTW) method

TWDTW is a time-weighted version of the DTW method for land use and land cover classification (Belgiu et al., 2018; Dong et al., 2020a). The DTW works by comparing the similarity between two sequences, such as the unknown sequence (Y) and target sequence (X), through warping the unknown sequence (Y) to search for their optimal path and obtain their minimum distance, namely the similarity (dissimilarity) (Petitjean et al., 2012; Petitjean and Weber, 2014). The TWDTW method adds a temporal cost accounting for the phase difference between the two time series to the minimum distance (Maus et al., 2016).

The TWDTW algorithm includes three steps in land use classification and identification: (1) generate the standard curve of a selected index (e.g., NDVI) for several crops or a single crop (e.g., sugarcane) based on time series images and field sample training data; (2) find the best alignment, and generate the dissimilarities (TWDTW distances) between unknown curves of NDVI time series (i.e., the NDVI curves of the unknown pixels) with the standard NDVI curve of sugarcane; (3) identify the unknown pixels based on the TWDTW distance; in this process pixels with low distance indicate high similarity and a high probability of being associated with the specified class (i.e., sugarcane).

2.3.2 Employing the TWDTW method for sugarcane mapping

<<Figure 5>>

The workflow by employing the TWDTW method for sugarcane harvest area mapping are shown in Fig.5. The growing period of sugarcane can be separated into four phases: the germination, tillering, grand growth, and ripening, and the NDVI values vary in different growing phases (Fig. 6). (1) Germination phase. Sugarcane begins to germinate approximately 15-30 days after planting. The NDVI starts to increase in this period. (2) Tillering phase. Tillering phase starts after approximately 2 months of germination, and tillers emerge from the base of the mother shoot to form 5-10 stalks. The NDVI increases quickly in this period. (3) Grand growth phase. This phase spans a period of approximately 4-10 months after planting. The NDVI peaks during this period. (4) Ripening and harvesting phase. In this phase, the NDVI starts to decrease, and the moisture content in sugarcane drastically drops.

From planting until the first cut, the crop is called planted sugarcane, and the growth cycle lasts between 12 and 18 months depending on the season, variety, and region of planting. After the first harvest, ratoon sugarcane is harvested yearly with a normal cycle of 12 months for a period of approximately 5 to 7 years or more (Rudorff et al., 2010). Although planted sugarcane and ratoon sugarcane have different length in growing cycles, they shared similar NDVI curves from the grand growth to harvest phases (red symbols in Fig. 6) which can be used as the standard seasonal curve for harvest area mapping. In Fig. 6, the standard NDVI curves for sugarcane were generated by randomly selecting 50 sugarcane samples across Brazil from the field data in 2018 (Section 2.2.2) and calculating their averaged NDVI values in the same growing period. The standard NDVI curves were produced as following: (1) randomly selecting 50 sugarcane samples across Brazil from the samples; (2) extracting their NDVI curves from the 16-day composited NDVI time series in 2018/2019 crop year; (3) adjusting the timeline of each NDVI curves by moving forward or back to ensure all the 50 NDVI curves generally in the same growing period; (4) calculating their averaged NDVI values to obtain the standard NDVI curves (Fig. 6).

<<Figure 6>>

Sugarcane in Brazil covers an extensive harvesting period (Rudorff et al. 2010). In the south-central area (including São Paulo, Goiás, Minas Gerais, Mato Grosso do Sul, Paraná, Mato Grosso, Bahia, Rio de Janeiro, and Espírito Santo), sugarcane is often harvested from April to December, with a harvesting season spanning 9 months. In the northeast area (including Alagoas, Pernambuco, Paraíba, Rio Grande do Norte, and Sergipe), sugarcane is harvested from September to April in the next year, with a harvesting season spanning 8 months. According to the phenology of sugarcane in the south-central and northeast areas,

Fig.7 shows the possible standard NDVI curves (these NDVI curves were repeated from the NDVI standard curve for sugarcane, as denoted by the red symbols and line in Fig. 6) for sugarcane in south-central and northeast Brazil. In this study, we incorporate the flexible phenological and harvest conditions of sugarcane in Brazil into the TWDTW method as follows.

<<Figure 7>>

(1) Calculate the dissimilarities (TWDTW distances). For each year from 2016-2019, we calculated all the "distance" values for each unknown pixel by comparing its NDVI timeseries with the standard NDVI curves of sugarcane (Fig. 7) based on the TWDTW method. In this process, we can obtain 13 "distance" values corresponding to the 13 standard NDVI curves in Fig. 7 for each pixel, and we selected the minimum value of the "distance" as the final distance value for each pixel. (2) Remove the influence of other vegetation with similar NDVI changes to sugarcane harvesting period, such as the Brazilian Cerrado biomes and pasture. In the Brazilian Cerrado, some vegetations (such as grassland, shrubland, woodland, and deciduous forest) exhibited low NDVI values from the end of the drought season to the beginning of the rainy season (August to October), which is similar to the sugarcane harvesting practice. However, the "difference" in NDVI between the growing season and non-growing season for these vegetations were lower than those for sugarcane (Ferreira et al. 2004; Mueller et al. 2015). Therefore, we used the NDVI "difference" as another supplement criterion to further separate sugarcane from the aforementioned vegetation types across the states related to the Brazilian Cerrado (i.e., São Paulo, Goiás, Minas Gerais, Mato Grosso do Sul, Mato Grosso, and Bahia). The "difference" between the maximum NDVI value in sugarcane growing season ($NDVI_{max}$: mean value of the two maximum NDVI in the growing season) and the minimum NDVI value in sugarcane non-growing season ($NDVI_{min}$: mean value of the two minimum NDVI in the non-growing season) was calculated for each pixel in these states (Fig. 7). From the statistics of 50 randomly selected sugarcane samples, we found the "difference" for sugarcane is mostly greater than 0.31. Therefore, we set the pixels with "difference" less than 0.31 as non-sugarcane. Additionally, pasture, which has similar temporal-spectral behaviour to sugarcane (Xavier et al., 2006), was further removed using the pasture maps (overall accuracy of 87%) produced by Parente et al. (2017). Across all the 14 studied states, pixels consecutively labelled as pasture on the pasture maps from 2016 to 2019 was set as non-sugarcane. (3) Identify to produce the sugarcane maps. We used the agricultural statistical harvest area for sugarcane at the state level to determine the "distance" threshold. A pixel with "distance" value lower than the "distance" threshold was considered a "sugarcane" pixel, and the total area of all sugarcane pixels should be equal to the statistical harvest area of sugarcane in the investigated state. In our study, municipalities with small areas of planted sugarcane (less than 1000 ha or less than 1% of the total sugarcane area in the investigated state) were identified separately to improve the identification accuracy of the entire investigated state.

2.3.3 Accuracy assessment

In this study, we first assessed the identification accuracy using the selected sugarcane and non-sugarcane samples based on the high-resolution images from Google Earth (Section 2.2.2). The producer's accuracy (PA), user's accuracy (UA), and overall accuracy (OA) were used for validation. The producer's accuracy (PA) is the percentage of surveyed reference samples correctly identified as the target class; the user's accuracy (UA) is the percentage of surveyed reference samples identified as

the target class on the classification map actually confirmed by field surveys; and overall accuracy (OA) is the ratio of correctly classified samples to all the samples. Additionally, we calculated the sugarcane harvest area on the map in different administrative regions and compared them with agricultural statistical data at the municipality, microregion, and mesoregion levels. The coefficient of determination (R^2) and RMAE (relative mean absolute error) between the statistical harvest area and the estimated harvest area were adopted to assess the map accuracy. The MAE (mean absolute error) can be expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|S_i - \hat{S}_i\right| \quad (2)$$

where S_i and $\hat{S}_i$ are the statistical area and identified area for the i th administrative region, respectively. n is the number of the administrative regions with valid statistical data. RMAE is the value of MAE relative to the mean value of the statistical area for all the n administrative regions:

$$RMAE = \frac{MAE}{\sum_{i=1}^{n} S_i/n} \quad (3)$$

3 Results

Annual cultivation maps of sugarcane in the 14 states of Brazil from 2016 to 2019 were produced using the TWDTW method (taking 2018 as an example in Fig. 8). The validation demonstrated good performance of the proposed method in identifying sugarcane harvest area in 2018. To show the detailed information of the sugarcane maps produced in our study. We selected four typical areas in different states to zoom-in and compared the sugarcane maps with high-resolution images from Google Earth (Fig. 9). In general, the sugarcane maps well captured the delineation of fields even though with some noises. What's more, the sugarcane maps can exhibit detailed information, such as small parcels, roads, and ridges between the fields. Based on the 2859 samples derived from Google Earth in 2018, the user's, producer's, and overall accuracies were 94.35%, 87.07%, and 91.47% in Brazil, respectively. The performance of the method varied by state and region. For all the 14 studied states, the overall accuracy (OA) varied from 83.51% to 94.46%, with the user's accuracy (UA) ranging from 86.96% to 98.25% and producer's accuracy (PA) ranging from 78.18% to 91.52% for sugarcane; and the user's accuracy (UA) ranging from 76.00% to 94.42% and producer's accuracy (PA) ranging from 86.05% to 98.15% for non-sugarcane (Table 1). São Paulo, the state with the largest planted area of sugarcane (accounting for over 50% of the sugarcane in Brazil), displayed high user's, producer's, and overall accuracies of 94.53%, 91.52%, and 94.46%, respectively. Goiás, the state with the second-largest planted area of sugarcane (accounting for approximately 10% of the sugarcane in Brazil), had high user's, producer's, and overall accuracies of 93.96%, 90.48%, and 93.38%, respectively. All the investigated states exhibited overall accuracy (OA) over 83% (Table 1).

<<Figure 8>>

<<Figure 9>>

<<Table 1>>

Additionally, the proposed method can accurately estimate the sugarcane harvest area compared with the agricultural statistical data at different administrative levels. The estimated harvest area of sugarcane in 2018 exhibited good correlations with the agricultural statistical area data derived from PAM at the municipality, microregion, and mesoregion levels. The coefficients of determination (R^2) were 0.88 (N=3637), 0.96 (N=369), and 0.99 (N=87) at the municipality, microregion, and mesoregion levels, respectively, and the respective RMAEs were 34.0% (MAE=0.09×10^4 ha), 20.5% (MAE=0.55×10^4 ha), and 13.7% (MAE=1.55×10^4 ha) (Fig. 10). The performance was better when the validated regions were aggregated to larger areas, namely the accuracy from low to high were at the municipality, microregion, and mesoregion levels.

<<Figure 10>>

The correlations between the agricultural statistical and the estimated harvest areas varied by state and region. At the municipality level, the coefficient of determination (R^2) between the agricultural statistical and the estimated harvest areas ranged from 0.62 to 0.98 in south-central Brazil and from 0.85 to 0.89 in northeast Brazil (Figs. 11-12); the RMAE ranged from 18.8% to 64.5% in south-central Brazil and from 38.8% to 53.2% in northeast Brazil (Fig. 11; Fig. 13). At the microregion level, the R^2 between the agricultural statistical and the estimated harvest areas ranged from 0.67 to 1 in south-central Brazil and from 0.70 to 1 in northeast Brazil (Fig. 12); the RMAE ranged from 14.2% to 60.5% in south-central Brazil and from 7.7% to 46.6% in northeast Brazil (Fig. 13). At the mesoregion level, the R^2 between the agricultural statistical and the estimated harvest areas ranged from 0.43 to 1 in south-central Brazil and from 0.99 to 1 in northeast Brazil (Fig. 12); the RMAE ranged from 1.7% to 58.1% in south-central Brazil and from 3.1% to 22.6% in northeast Brazil (Fig. 13). Validation at all three levels showed high performance, with high R^2 and slope close to 1. In general, Mato Grosso and Bahia in the northern part of south-central Brazil and Sergipe in the southern part of northeastern Brazil displayed lower R^2 and higher RMAE values (Figs. 11-13). The performance for São Paulo and Goiás were higher than those for most other states, except Rio de Janeiro (Figs. 11-13). Finally, we assessed the capability of the method and standard seasonal changes in NDVI (i.e., standard NDVI curves) acquired from a single year (2018) to apply them to other years (2016 and 2017). Results indicated the R^2 and RMAE values for the period of 2016–2018 changed little in most states (Figs. 12-13).

<<Figures 11-13>>

4 Discussion

As the largest global producer of sugarcane, Brazil contributed to approximately 38.6% of the world's sugarcane production in 2019 and played an important role in retaining the global demand for sugarcane (FAO, 2020). While the cultivation area of sugarcane in Brazil has increased by approximately 11% over the past 10 years according to census data (FAO, 2020), which indicated substantial land cover changes and important feedback to regional climate systems (Loarie et al., 2011; Mello et al., 2014; Jaiswal et al., 2017). The first harvest area map of sugarcane in Brazil at a 30-m spatial resolution was generated only for São Paulo state for the 2003/2004 crop year based on an automatic image classification method (Rudorff et al., 2005). Subsequently, the cultivation map of sugarcane was updated using visual/manual image interpretation through 2013, and the

coverage of the map extended from São Paulo state to a total of 8 states in the south-central region of Brazil, accounting for 88% of the sugarcane cultivation area in Brazil, as reported by the Canasat Project (Rudorff et al., 2010). Souza et al. (2020) reconstructed annual land use and land cover information at a 30-m spatial resolution from 1985-2017 for Brazil based on random forest method and Landsat data trained by plenty of samples selected from existing land cover maps, which provides the cultivation map of sugarcane belonging to a subclass of agriculture (the producer's and user's accuracies for agriculture were 83.3% and of 81.3%, respectively). However, these prevailing methods strongly require large volume of training samples, which makes it difficult to update annually at large scales.

In this study, we proposed a phenology-based sugarcane classification method by incorporating multiple phenological conditions of sugarcane into the TWDTW method. Then, we identified the harvest area of sugarcane with a spatial resolution of 30-m in Brazil from 2016 to 2019 using 16-day composite NDVI series derived from Landsat-7/8 and Sentinel-2 data. Our proposed method can automatically identify the sugarcane areas with limited training data and needs only one standard NDVI curve (Fig. 6) for all the 14 sugarcane planted states in Brazil. Validations against field sample data and agricultural statistical area data showed that the generated sugarcane harvest maps were in high accuracy. For example, validations for 2018 displayed the user's, producer's, and overall accuracies of 94.35%, 87.07%, and 91.47% in Brazil, respectively.

Although our method can effectively and accurately identify the sugarcane harvest area from the regional to national or continental scales, there are still several potential uncertainties in the identification process. First, because of the quality of the processed NDVI series, the speckle "salt-and-pepper" effects/noises exist in some areas of the harvest map. According to the map statistics in 2018, sugarcane patches with only one pixel account for 0.7% (São Paulo) to 9.1% (Espírito Santo) of the sugarcane area (Fig. 14). In the future, the object-based identification by segmenting images into homogeneous objects, instead of the pixel-based method used in our study, may alleviate the "salt-and-pepper" effects and improve the identification performance (Belgiu et al., 2018).

<<Figure 14>>

Second, the identification accuracy for Bahia state was lower than that for other states. Because of the different harvest seasons in south-central and northeast Brazil, we identified sugarcane in the south-central and northeast areas using different phenological characteristics and standard curve combinations (Fig. 7). Bahia is a transition state between south-central and northeast Brazil. Namely, sugarcane in southern Bahia has a harvest season similar to that in south-central Brazil, and sugarcane in northern Bahia has a harvest season similar to that in northeast Brazil. In our study, we treated Bahia as a state in south-central Brazil with a harvest season from April to December (Fig. 7), which may introduce errors to the identification in the northern part of Bahia.

Third, Mato Grosso state exhibited a lower R^2 and higher RMAE than other states when comparing the identified sugarcane harvest area with the agricultural statistical sugarcane harvest area. Two major biomes are located in Mato Grosso, including the humid tropical forests of the Amazon in the north and the heterogeneous Cerrado area (a tropical savanna) in the south-central part of the state (Kastens et al. 2017). In Mato Grosso, sugarcane may be misclassified with some kinds of grassland, grazing areas, or seasonal forest, which exhibit phenological changes similar to those of sugarcane (Mueller et al. 2015; Bendini

et al. 2019). The NDVI values of these vegetation types decrease between August and October (from the end of the drought season to the beginning of the rainy season) and increased thereafter (Ferreira et al. 2004), which is quite similar to that in the harvesting stage of sugarcane, resulting in misclassification with the sugarcane harvested from August to October. In our study, we used the "difference" between the maximum NDVI value in the growing season and the minimum NDVI value in the non-growing season (see method in Section 2.3.2) to alleviate the misclassification at some extent because the "difference" for the abovementioned vegetation types is generally lower than that for sugarcane (Ferreira et al. 2004). In the future, incorporating more complex spectral-temporal variability metrics, such as the combination of more spectral information instead of NDVI and a longer time window for several harvest seasons instead of one harvest season may help improve model performance (Mueller et al. 2015).

5 Data availability

The 30-m Brazil sugarcane harvest area dataset from 2016-2019 is available at https://doi.org/10.6084/m9.figshare.14213909 (Zheng et al., 2021). The dataset is provided in .tif format with pixel values of 1 for sugarcane and 0 for non-sugarcane.

6 Conclusion

Brazil is the world's largest sugarcane producer and contributes to approximately 38.6% of total global sugarcane production. Based on the available Landsat-7/8 and Sentinel-2 optical images, we produced sugarcane harvest maps with a 30-m spatial resolution (2016-2019) by incorporating multiple phenological conditions of sugarcane in Brazil into the TWDTW method. The proposed method can automatically identify sugarcane harvest area with limited training sample data. Based on 2859 samples derived from Google Earth, the validation experiment reflected high accuracy across the 14 sugarcane planted states in Brazil in 2018, with the user's, producer's, and overall accuracies of 94.35%, 87.07%, and 91.47%, respectively. Additionally, the identified harvest area of sugarcane exhibited a good correlation with the agricultural statistical area data derived from PAM at the municipality, microregion, and mesoregion levels. The maps can be used to monitor the harvest area and yield of sugarcane, and evaluate the feedback to regional climate.

Author contributions. Wenping Yuan and Yi Zheng designed the research, performed the analysis, and wrote the paper; Ana Cláudia dos Santos Luciano and Jie Dong edited and revised the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements

This study was supported by the National Science Fund for Distinguished Young Scholars (41925001), National Natural Science Foundation of China (41971018 and 31930072), and Fundamental Research Funds for the Central Universities (19lgjc02).

References

Abdel-Rahman, E. M. and Ahmed, F. B.: The application of remote sensing techniques to sugarcane (Saccharum spp. hybrid) production: a review of the literature, Int. J. Remote Sens., 29, 3753-3767, https://doi.org/10.1080/01431160701874603, 2008.

Adami, M., Theodor Rudorff, B. F., Freitas, R. M., Aguiar, D. A., Sugawara, L. M., and Mello, M. P.: Remote Sensing Time Series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil, Sustainability, 4, 574-585, https://doi.org/10.3390/su4040574, 2012.

Belgiu, M. and Csillik, O.: Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, Remote Sens. Environ., 204, 509-523, https://doi.org/10.1016/j.rse.2017.10.005, 2018.

Bendini, H. d. N., Fonseca, L. M. G., Schwieder, M., Korting, T. S., Rufin, P., Sanches, I. D. A., Leitao, P. J., and Hostert, P.: Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series, Int. J. Appl. Earth Observ. Geoinfo., 82, https://doi.org/10.1016/j.jag.2019.05.005, 2019.

Bordonal, R. d. O., Lal, R., Aguiar, D. A., de Figueiredo, E. B., Perillo, L. I., Adami, M., Theodor Rudorff, B. F., and La Scala, N.: Greenhouse gas balance from cultivation and direct land use change of recently established sugarcane (Saccharum officinarum) plantation in south-central Brazil, Renew. Sust. Energ. Rev., 52, 547-556, https://doi.org/10.1016/j.rser.2015.07.137, 2015.

Boryan, C., Yang, Z., Mueller, R. and Craig, M.: Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program, Geocarto Int., 26(5), 341–358, doi:10.1080/10106049.2011.562309, 2011.

Brar N S, Dhillon B S, Saini K S, et al. Agronomy of sugarbeet cultivation-A review. Agricultural Reviews, 36(3): 184-197, 2015.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. and Li, Z.: A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach, Remote Sens. Environ., 210, 35–47, doi:10.1016/j.rse.2018.02.045, 2018.

Cechim Junior, C., Johann, J. A., and Antunes, J. F. G.: Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images, Revista Brasileira de Engenharia Agrícola e Ambiental, 21, 427-432, https://doi.org/10.1590/1807-1929/agriambi.v21n6p427-432, 2017.

Dong, J., Fu, Y., Wang, J., Tian, H., Fu, S., Niu, Z., Han, W., Zheng, Y., Huang, J., and Yuan, W.: Early-season mapping of winter wheat in China based on Landsat and Sentinel images, Earth Syst. Sci. Data, 12, 3081-3095, https://doi.org/10.5194/essd-12-3081-2020, 2020a.

de Almeida, C. A., Coutinho, A. C., Dalla Mora Esquerdo, J. C., Adami, M., Venturieri, A., Diniz, C. G., Dessay, N., Durieux, L., and Gomes, A. R.: High spatial resolution land use and land cover mapping wf the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data, Acta Amazonica, 46, 291-302, https://doi.org/10.1590/1809-4392201505504, 2016.

Dong, J., Liu, W., Han, W., Xiang, K., Lei, T., and Yuan, W.: A phenology-based method for identifying the planting fraction of winter wheat using moderate-resolution satellite data, Int. J. Remote Sens., 41, 6892-6913, https://doi.org/10.1080/01431161.2020.1755738, 2020b.

dos Santos Luciano, A. C., Araujo Picoli, M. C., Rocha, J. V., Junqueira Franco, H. C., Sanches, G. M., Lima Verde Leal, M. R., and le Maire, G.: Generalized space-time classifiers for monitoring sugarcane areas in Brazil, Remote Sens. Environ., 215, 438-451, https://doi.org/10.1016/j.rse.2018.06.017, 2018.

dos Santos Luciano, A. C., Araujo Picoli, M. C., Rocha, J. V., Duft, D. G., Camargo Lamparelli, R. A., Lima Verde Leal, M. R., and Le Maire, G.: A generalized space-time OBIA classification scheme to map sugarcane areas at regional scale, using Landsat images time-series and the random forest algorithm, Int. J. Appl. Earth Observ. Geoinfo., 80, 127-136, https://doi.org/10.1016/j.jag.2019.04.013, 2019.

Defante, L. R., Vilpoux, O. F., and Sauer, L.: Rapid expansion of sugarcane crop for biofuels and influence on food production in the first producing region of Brazil, Food Policy, 79, 121-131, https://doi.org/10.1016/j.foodpol.2018.06.005, 2018.

FAOSTAT: Food and Agriculture Organization of the United Nations (FAO): FAO Statistical Databases., 2020.

Ferreira, L. G. and Huete, A. R.: Assessing the seasonal dynamics of the Brazilian Cerrado vegetation through the use of spectral vegetation indices, Int. J. Remote Sens., 25, 1837-1860, https://doi.org/10.1080/0143116031000101530, 2004.

Ferreira, M. P., Alves, D. S., and Shimabukuro, Y. E.: Forest dynamics and land-use transitions in the Brazilian Atlantic Forest: the case of sugarcane expansion, Reg. Environ. Change, 15, 365-377, https://doi.org/10.1007/s10113-014-0652-6, 2015.

Iqbal M A, Saleem A M. Sugar beet potential to beat sugarcane as a sugar crop in Pakistan[J]. Am. Eurasian. J. Agric. Environ. Sci., 15(1): 36-44, 2015.

Jaiswal, D., De Souza, A. P., Larsen, S., LeBauer, D. S., Miguez, F. E., Sparovek, G., Bollero, G., Buckeridge, M. S., and Long, S. P.: Brazilian sugarcane ethanol as an expandable green alternative to crude oil use, Nat. Clim. Change, 7, 788-+, https://doi.org/10.1038/nclimate3410, 2017.

Jiang, H., Li, D., Jing, W., Xu, J., Huang, J., Yang, J., and Chen, S.: Early Season Mapping of Sugarcane by Applying Machine Learning Algorithms to Sentinel-1A/2 Time Series Data: A Case Study in Zhanjiang City, China, Remote Sens., 11, https://doi.org/10.3390/rs11070861, 2019.

Johnson, B. A., Scheyvens, H., and Shivakoti, B. R.: An ensemble pansharpening approach for finer-scale mapping of sugarcane with Landsat 8 imagery, Int. J. Appl. Earth Observ. Geoinfo., 33, 218-225, https://doi.org/10.1016/j.jag.2014.06.003, 2014.

Kastens, J. H., Brown, J. C., Coutinho, A. C., Bishop, C. R., and Esquerdo, J. C. D. M.: Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil, PloS One, 12, https://doi.org/10.1371/journal.pone.0176168, 2017.

Leal, M. R. L. V., Galdos, M. V., Scarpare, F. V., Seabra, J. E. A., Walter, A., and Oliveira, C. O. F.: Sugarcane straw availability, quality, recovery and energy use: A literature review, Biomass Bioenergy, 53, 11-19, https://doi.org/10.1016/j.biombioe.2013.03.007, 2013.

Li, H., Chen, J., Liang, S., and Li, Q.: Sugarcane Mapping in Tillering Period by Quad-Polarization TerraSAR-X Data, IEEE Geosci. Remote Sens. Lett., 12, 993-997, https://doi.org/10.1109/lgrs.2014.2372037, 2015.

Lin, H., Chen, J., Pei, Z., Zhang, S., and Hu, X.: Monitoring Sugarcane Growth Using ENVISAT ASAR Data, IEEE Trans. Geosci. Remote Sens., 47, 2572-2580, https://doi.org/10.1109/tgrs.2009.2015769, 2009.

Loarie, S. R., Lobell, D. B., Asner, G. P., Mu, Q., and Field, C. B.: Direct impacts on local climate of sugar-cane expansion in Brazil, Nat. Clim. Change, 1, 105-109, https://doi.org/10.1038/nclimate1067, 2011.

Manabe, V. D., Melo, M. R. S., and Rocha, J. V.: Framework for Mapping Integrated Crop-Livestock Systems in Mato Grosso, Brazil, Remote Sens., 10, https://doi.org/10.3390/rs10091322, 2018.

Massey, R., Sankey, T. T., Congalton, R. G., Yadav, K., Thenkabail, P. S., Ozdogan, M., and Meador, A. J. S.: MODIS phenology-derived, multi-year distribution of conterminous US crop types, Remote Sens. Environ., 198, 490-503, https://doi.org/10.1016/j.rse.2017.06.033, 2017.

Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., and de Queiroz, G. R.: A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping, IEEE J. Sel. Top. Appl., 9, 3729-3739, https://doi.org/10.1109/jstars.2016.2517118, 2016.

Mello, F. F. C., Cerri, C. E. P., Davies, C. A., Holbrook, N. M., Paustian, K., Maia, S. M. F., Galdos, M. V., Bernoux, M., and Cerri, C. C.: Payback time for soil carbon and sugar-cane ethanol, Nat. Clim. Change, 4, 605-609, https://doi.org/10.1038/nclimate2239, 2014.

Monteiro, L. A., Sentelhas, P. C., and Pedra, G. U.: Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation, Int. J. Clim., 38, 1571-1581, https://doi.org/10.1002/joc.5282, 2018.

Mueller, H., Rufin, P., Griffiths, P., Barros Siqueira, A. J., and Hostert, P.: Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape, Remote Sens. Environ., 156, 490-499, https://doi.org/10.1016/j.rse.2014.10.014, 2015.

Mulyono, S., Nadirah, and Iop: Identifying Sugarcane Plantation using LANDSAT-8 Images with Support Vector Machines. In: 2nd International Conference of Indonesian Society for Remote Sensing, IOP Conference Series-Earth and Environmental Science, 2017.

Parente, L., Ferreira, L., Faria, A., Nogueira, S., Araujo, F., Teixeira, L., and Hagen, S.: Monitoring the brazilian pasturelands: A new mapping approach based on the landsat 8 spectral and temporal domains, Int. J. Appl. Earth Observ. Geoinfo., 62, 135-143, https://doi.org/10.1016/j.jag.2017.06.003, 2017.Petitjean, F., Inglada, J., and Gancarski, P.: Satellite Image Time Series Analysis Under Time Warping, IEEE Trans. Geosci. Remote Sens., 50, 3081-3095, https://doi.org/10.1109/tgrs.2011.2179050, 2012.

Petitjean, F., Weber, J.: Efficient satellite image time series analysis under time warping. IEEE Geosci. Remote Sens. Lett., 11, 1143–1147, 2014.

Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition, IEEE T. Acoust. Speech, 26, 43–49, https://doi.org/10.1109/TASSP.1978.1163055, 1978.

Sindhu, R., Gnansounou, E., Binod, P., and Pandey, A.: Bioconversion of sugarcane crop residue for value added products - An overview, Renew. Energy, 98, 203-215, https://doi.org/10.1016/j.renene.2016.02.057, 2016.

Souza Jr, C. M., Shimbo, J. Z., Rosa, M. R., Parente, L. L., Alencar, A. A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., Ferreira, L. G., Souza-Filho, P. W. M., de Oliveira, S. W., Rocha, W. F., Fonseca, A. V., Marques, C. B., Diniz, C. G., Costa, D., Monteiro, D., Rosa, E. R., Velez-Martin, E., Weber, E. J., Lenti, F. E. B., Paternost, F. F., Pareyn, F. G. C., Siqueira, J. V., Viera, J. L., Ferreira Neto, L. C., Saraiva, M. M., Sales, M. H., Salgado, M. P. G., Vasconcelos, R., Galano, S., Mesquita, V. V., and Azevedo, T.: Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine, Remote Sens., 12, https://doi.org/10.3390/rs12172735, 2020.

Theodor Rudorff, B. F., Berka, L.M.S., Moreira, M.A., Duarte, V., Xavier, A.C., Rosa, V.G.C., Shimabukuro, Y.E.: Imagens de satélite no mapeamento e estimativa de área de cana-de-açúcar em São Paulo: ano-safra 2003/04. Agricultura em São Paulo, 52, 21-39, 2005

Theodor Rudorff, B. F., de Aguiar, D. A., da Silva, W. F., Sugawara, L. M., Adami, M., and Moreira, M. A.: Studies on the Rapid Expansion of Sugarcane for Ethanol Production in Sao Paulo State (Brazil) Using Landsat Data, Remote Sens., 2, 1057-1076, https://doi.org/10.3390/rs2041057, 2010.

Wachholz de Souza, C. H., Cervi, W. R., Brown, J. C., Rocha, J. V., and Camargo Lamparelli, R. A.: Mapping and evaluating sugarcane expansion in Brazil's savanna using MODIS and intensity analysis: a case-study from the state of Tocantins, J. Land Use Sci., 12, 457-476, https://doi.org/10.1080/1747423x.2017.1404647, 2017.

Wang, S., Azzari, G. and Lobell, D. B.: Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques, Remote Sens. Environ., 222, 303–317, doi:10.1016/j.rse.2018.12.026, 2019.

Wang, J., Xiao, X., Liu, L., Wu, X., Qin, Y., Steiner, J. L., and Dong, J.: Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images, Remote Sens. Environ., 247, https://doi.org/10.1016/j.rse.2020.111951, 2020.

Wardlow, B., Egbert, S. and Kastens, J.: Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains, Remote Sens. Environ., 108(3), 290–310, doi:10.1016/j.rse.2006.11.021, 2007.

Xavier, A. C., Rudorff, B. F. T., Berka, L. M. S., and Moreira, M. A.: Multi-temporal analysis of MODIS data to classify sugarcane crop, International Journal of Remote Sensing, 27, 755-768, https://doi.org/10.1080/01431160500296735, 2006.

Zhang, H., Anderson, R. G., and Wang, D.: Satellite-based crop coefficient and regional water use estimates for Hawaiian sugarcane, Field Crop. Res., 180, 143-154, https://doi.org/10.1016/j.fcr.2015.05.023, 2015.

Zheng, B., Myint, S. W., Thenkabail, P. S. and Aggarwal, R. M.: A support vector machine to identify irrigated crop types using time-series Landsat NDVI data, Int. J. Appl. Earth Observ. Geoinfo., 34, 103–112, doi:10.1016/j.jag.2014.07.002, 2015.

Zheng, Y., dos Santos Luciano, A. C., Dong, J., Yuan, W.: 30-m sugarcane harvest maps in Brazil, figshare, Dataset, https://doi.org/10.6084/m9.figshare.14213909, 2021.

Zhong, L., Gong, P., Biging, G.S.: Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery, Remote Sens. Environ., 140, 1–13, doi.org/10.1016/j.rse.2013.08.023, 2014.

Zhou, Z., Huang, J., Wang, J., Zhang, K., Kuang, Z., Zhong, S., and Song, X.: Object-Oriented Classification of Sugarcane Using Time-Series Middle-Resolution Remote Sensing Data Based on AdaBoost, PloS One, 10, https://doi.org/10.1371/journal.pone.0142069, 2015.

Tables

Table 1: Confusion matrix of the sugarcane harvest map for the 14 states in Brazil in 2018.

	State	Reference	Map		Producer's accuracy	User's accuracy	Overall accuracy
			Sugarcane	Non-sugarcane			
South-central	São Paulo (SP)	Sugarcane	259	24	91.52%	94.53%	94.46%
		Non-sugarcane	15	406	96.44%	94.42%	
	Goiás (GO)	Sugarcane	171	18	90.48%	93.96%	93.38%
		Non-sugarcane	11	238	95.58%	92.97%	
	Minas Gerais (MG)	Sugarcane	92	17	84.40%	92.00%	87.37%
		Non-sugarcane	8	81	91.01%	82.65%	
	Mato Grosso do Sul (MS)	Sugarcane	81	9	90.00%	97.59%	94.44%
		Non-sugarcane	2	106	98.15%	92.17%	
	Paraná (PR)	Sugarcane	77	16	82.80%	90.59%	91.84%
		Non-sugarcane	8	193	96.02%	92.34%	
	Mato Grosso (MT)	Sugarcane	79	19	80.61%	94.05%	87.88%
		Non-sugarcane	5	95	95.00%	83.33%	
	Bahia (BA)	Sugarcane	56	11	83.58%	98.25%	89.29%
		Non-sugarcane	1	44	97.78%	80.00%	
	Rio de Janeiro (RJ)	Sugarcane	76	8	90.48%	97.44%	92.65%
		Non-sugarcane	2	50	96.15%	86.21%	
	Espírito Santo (ES)	Sugarcane	61	11	84.72%	96.83%	88.50%
		Non-sugarcane	2	39	95.12%	78.00%	
Northeast	Alagoas (AL)	Sugarcane	53	8	86.89%	96.36%	90.29%
		Non-sugarcane	2	40	95.24%	83.33%	
	Pernambuco (PE)	Sugarcane	52	9	85.25%	94.55%	89.09%
		Non-sugarcane	3	46	93.88%	83.64%	
	Paraíba (PB)	Sugarcane	43	12	78.18%	91.49%	83.51%
		Non-sugarcane	4	38	90.48%	76.00%	
	Rio Grande do Norte (RN)	Sugarcane	40	5	88.89%	86.96%	87.50%
		Non-sugarcane	6	37	86.05%	88.10%	
	Sergipe (SE)	Sugarcane	29	7	80.56%	96.67%	88.57%
		Non-sugarcane	1	33	97.06%	82.50%	

Figures

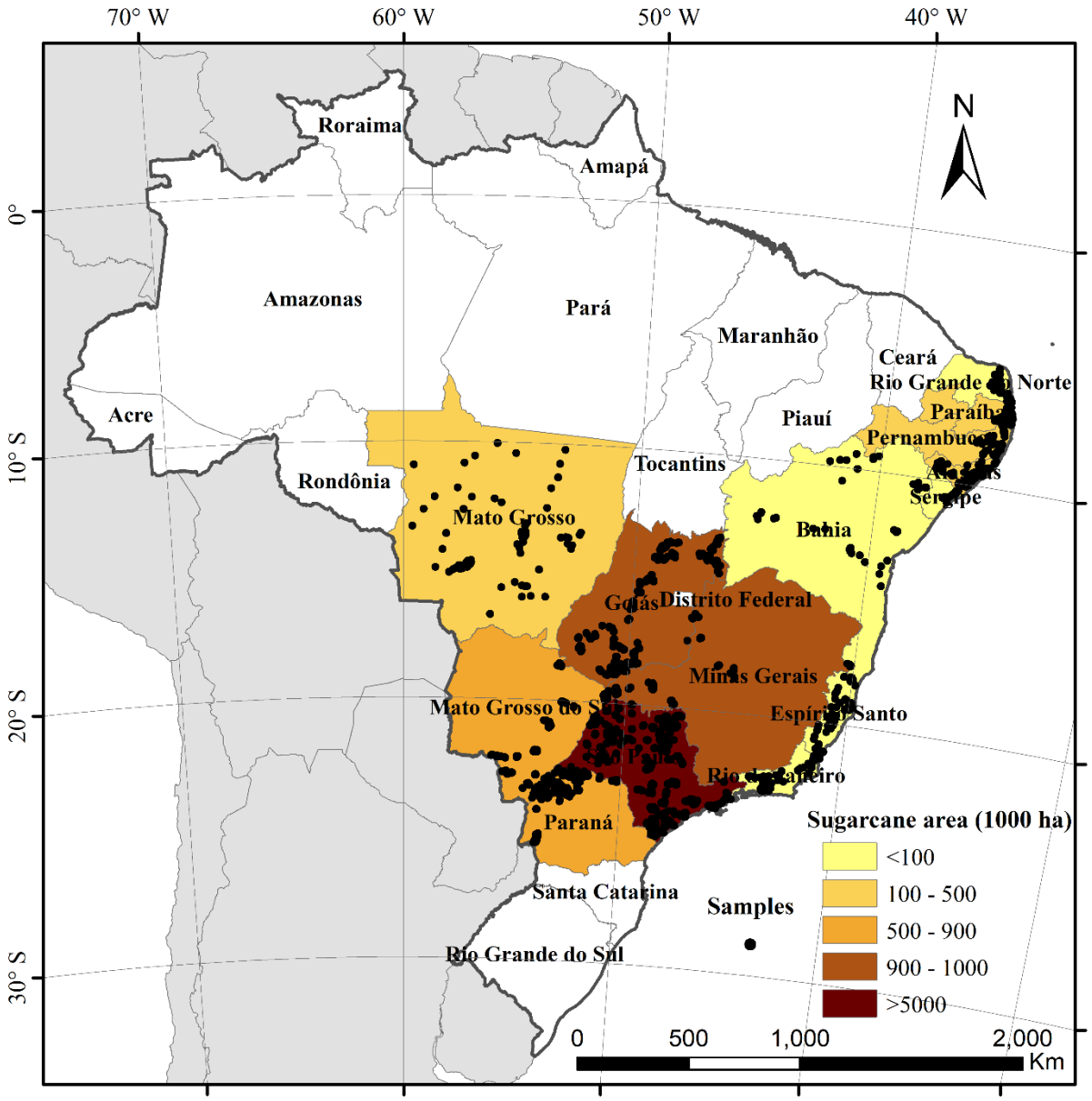


Figure 1. Study areas in Brazil for sugarcane harvest area identification, including 9 states in south-central Brazil (i.e., São Paulo, Goiás, Minas Gerais, Mato Grosso do Sul, Paraná, Mato Grosso, Bahia, Rio de Janeiro, and Espírito Santo) and 5 states in northeast Brazil (i.e., Alagoas, Pernambuco, Paraíba, Rio Grande do Norte, and Sergipe), which account for over 98% of the sugarcane harvest area in Brazil. The dots represent the samples used for validation. The administrative boundary data were derived from the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE; www.ibge.gov.br).

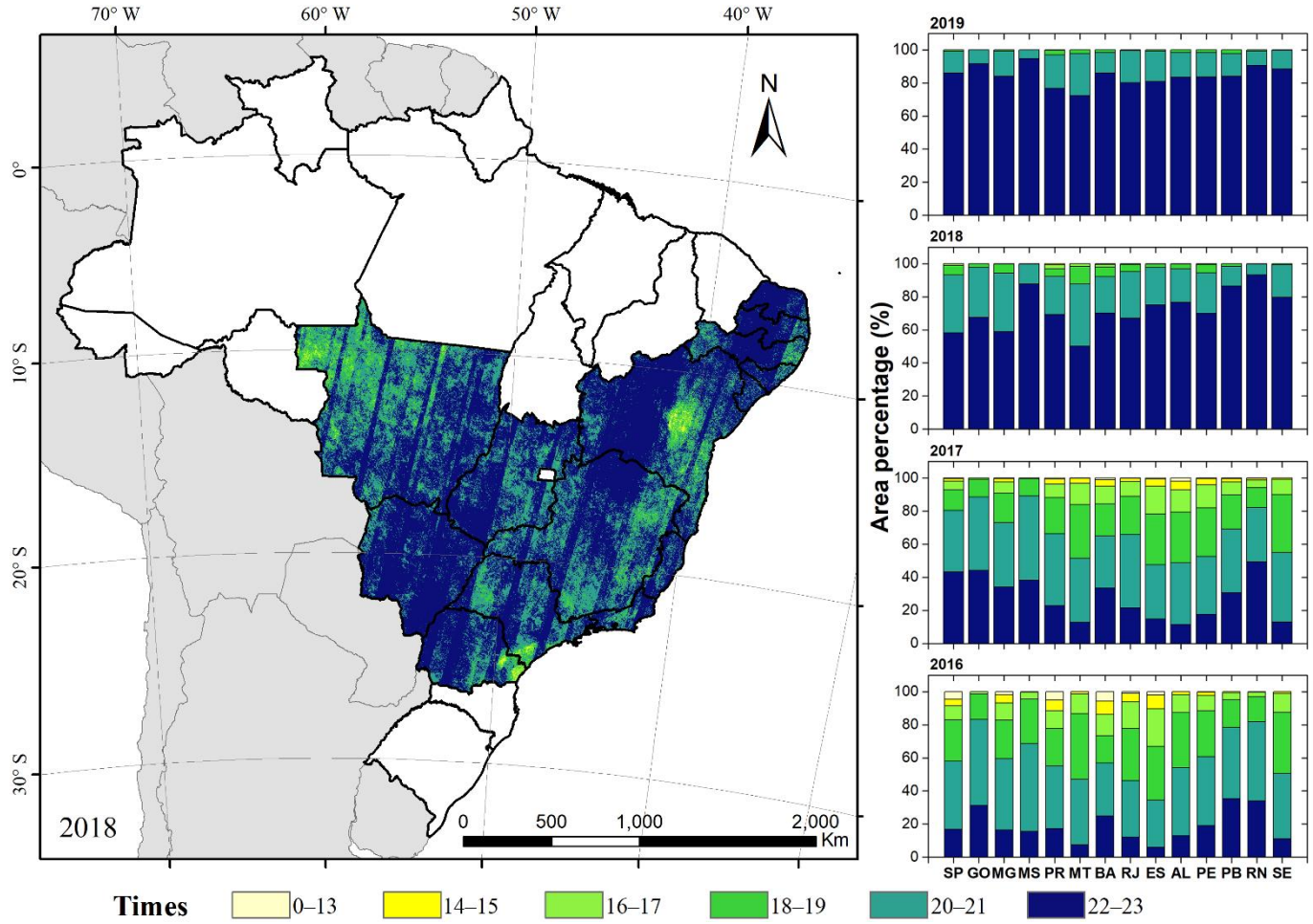


Figure 2. Times of good observations for the 16-day composite satellite data over a year. (left) Times of good observations in 2018; (right) Area percentages of the times of good observations for each state from 2016-2019. The administrative boundary data were obtained from the IBGE.

500

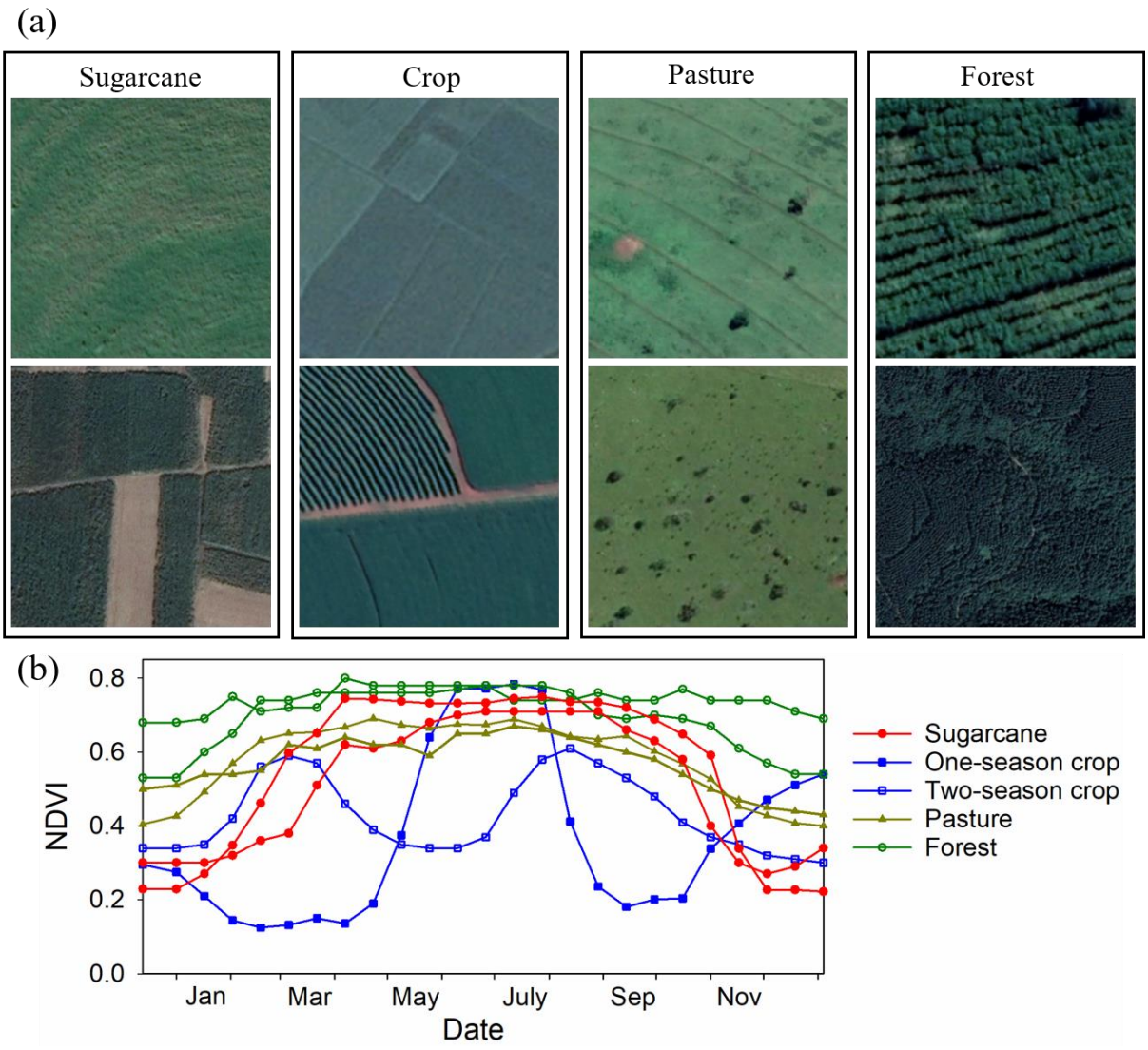


Figure 3. Examples of the (a) color and textures on the high resolution images from Google Earth © Google Earth, and (b) timeseries of NDVI for different vegetations.

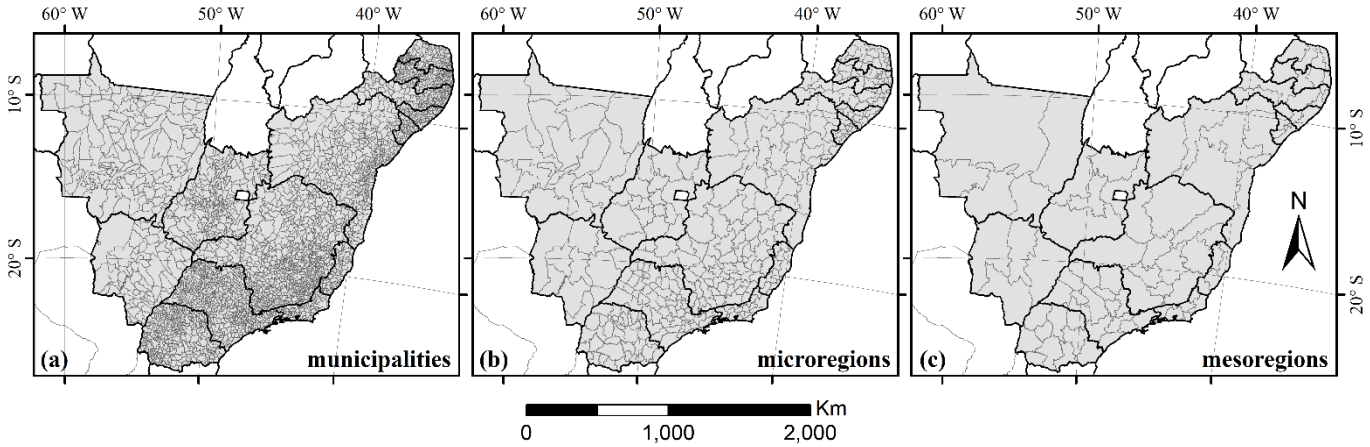


Figure 4. The administrative boundaries for (a) municipalities, (b) microregions, and (c) mesoregions. The boundary lines at the municipality and state levels were downloaded directly from the IBGE, and we aggregated the municipalities into microregions and mesoregions according to the regions in PAM denoted by the IBGE.

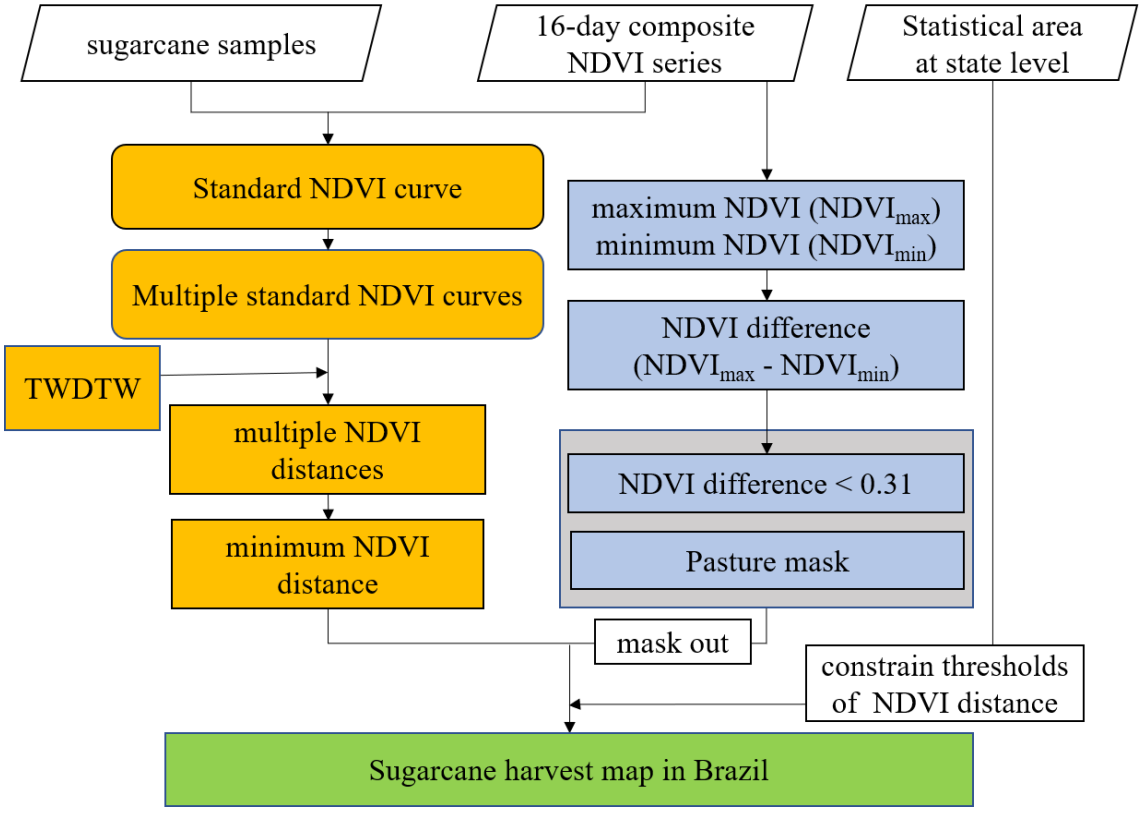


Figure 5. The workflow by employing the TWDTW method for sugarcane harvest area mapping.

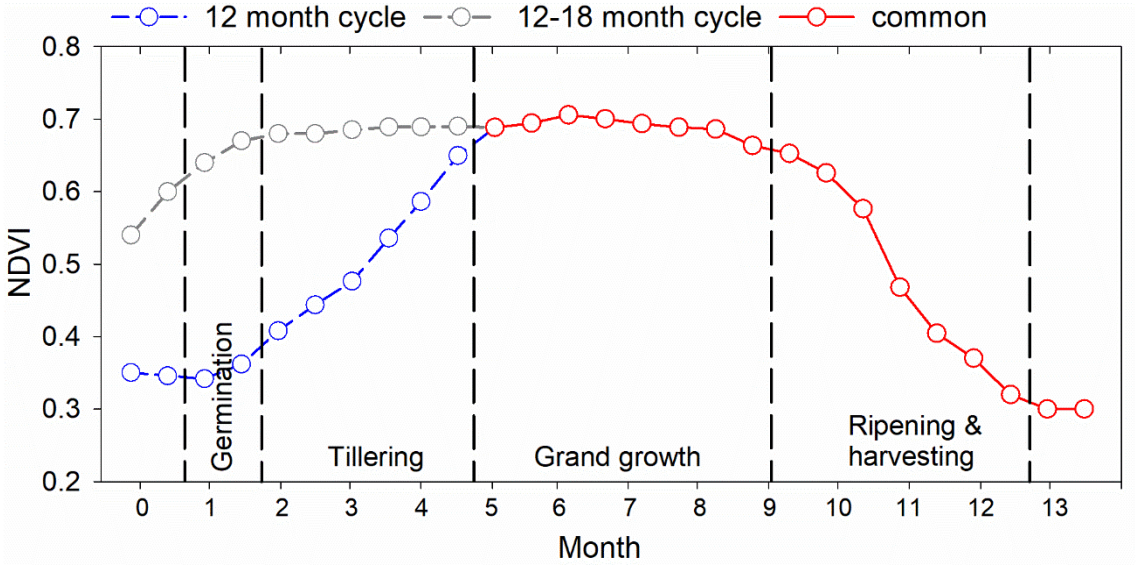


Figure 6. Growing stages of sugarcane and the respective NDVI changes. The gray and red symbols represent the NDVI curve for the planted sugarcane with a 12–18 month cycle, and the green and red symbols represent the NDVI curve for the ratoon sugarcane with a 12 month cycle. The growing stages of the 12 month cycle sugarcane (germination, tillering, grand growth, and ripening) are labeled in the figure.

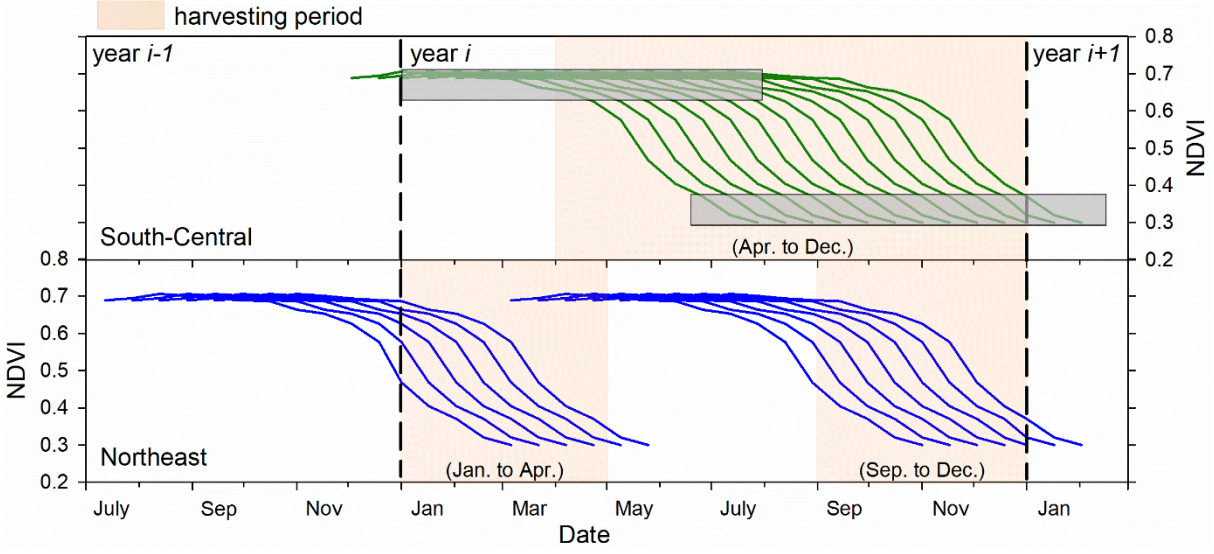
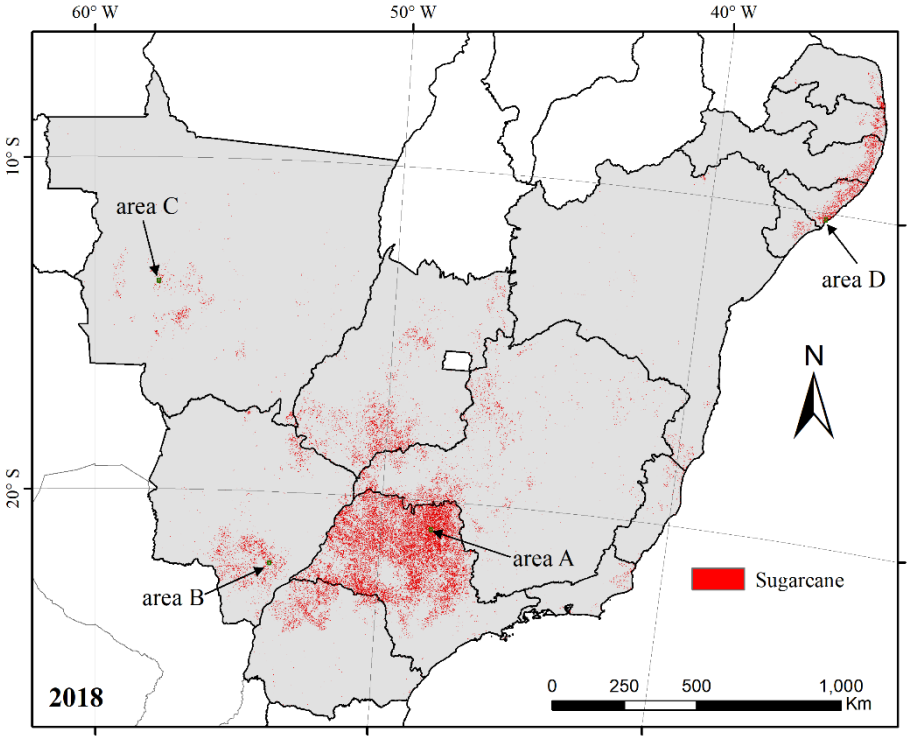


Figure 7. Seasonal changes in the NDVI series for sugarcane in the south-central and northeast Brazil. The gray areas are the periods used to calculate the "difference" between the maximum NDVI value in growing season ($NDVI_{max}$) and the minimum NDVI value in non-growing season ($NDVI_{min}$).



525 **Figure 8. Sugarcane harvest map for the 14 studied states in Brazil in 2018. The administrative boundary data were obtained from the IBGE.**

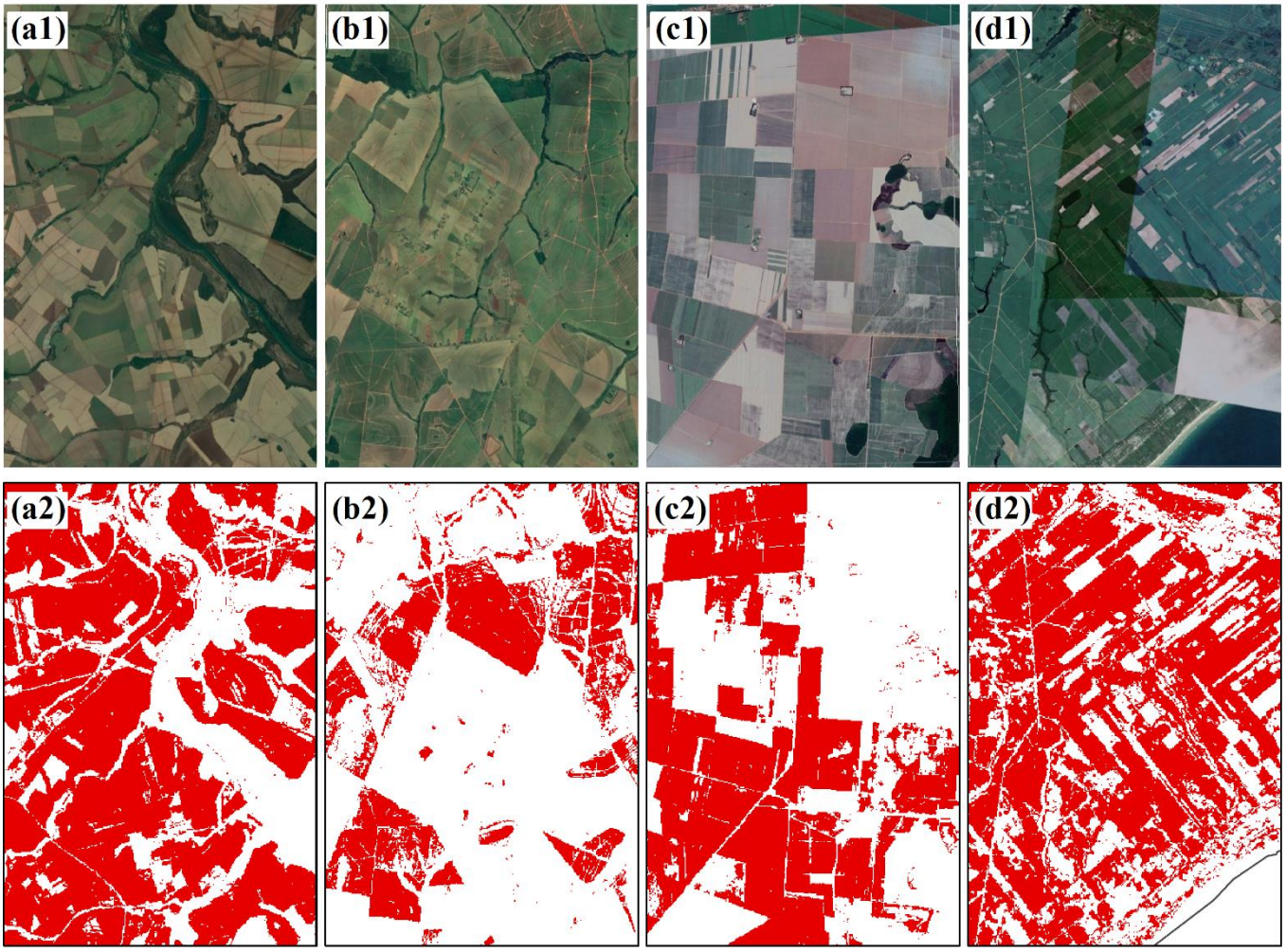


Figure 9. Zoomed-in of (a1)-(d1) high-resolution images from Google Earth © Google Earth, (a2)-(d2) presence (red) and absence (white) of sugarcane on the harvest maps in 2018 for the typical area A-D in Fig. 8.

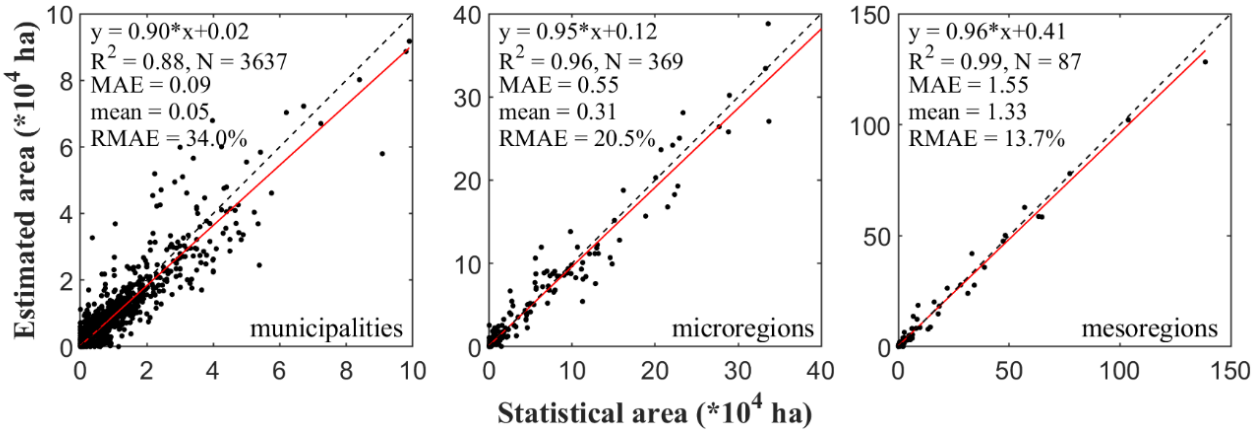


Figure 10. Comparisons between the agricultural statistical harvest area and the estimated harvest area of sugarcane at the municipality level in Brazil in 2018. "N" and "mean" represents the number and mean value of all the valid statistical data in each figure, respectively.

535

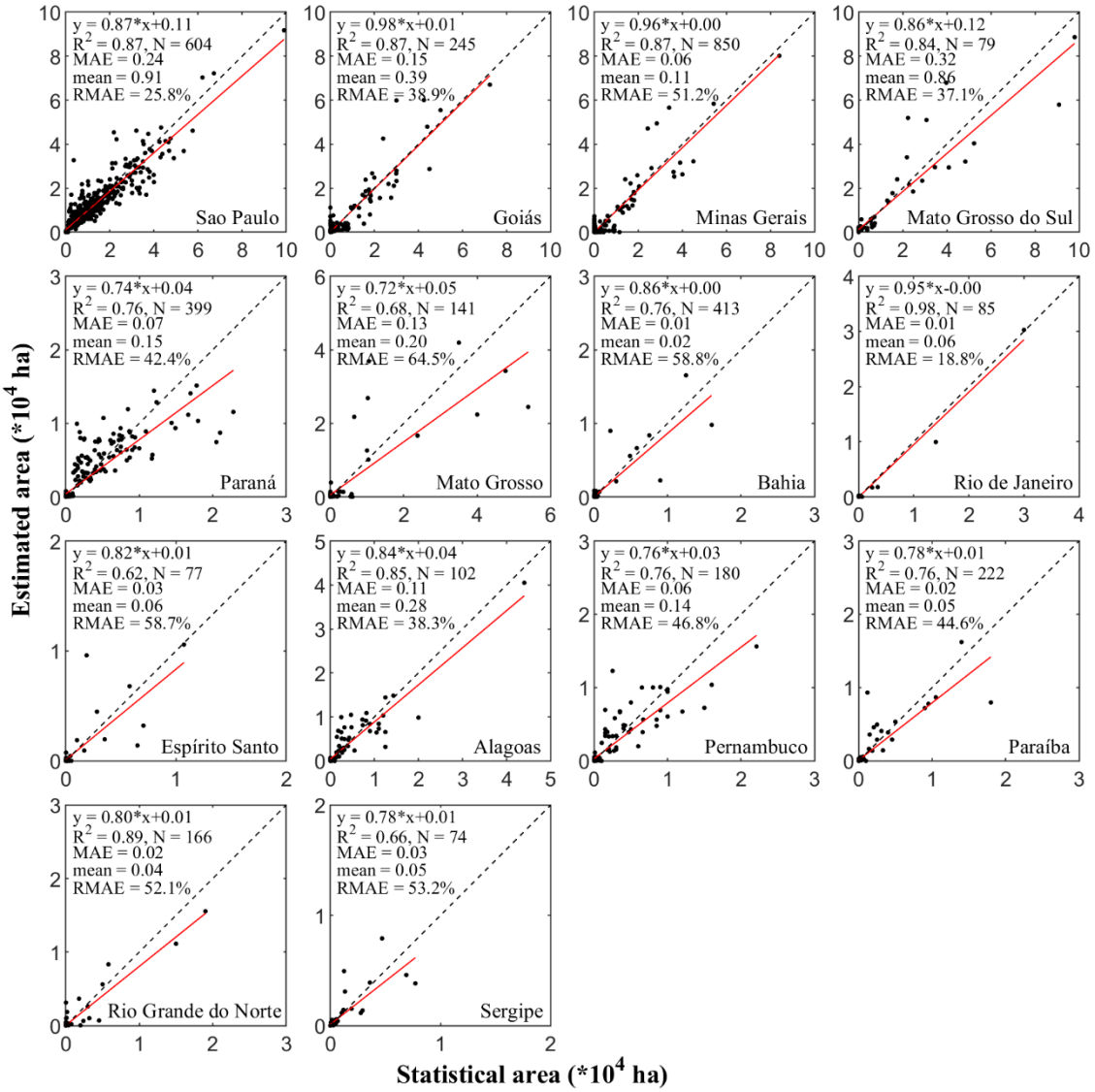


Figure 11. Comparisons between the agricultural statistical harvest area and the estimated harvest area of sugarcane at the municipality level in the 14 states in 2018. "N" and "mean" represents the number and mean value of all the valid statistical data in each figure, respectively.

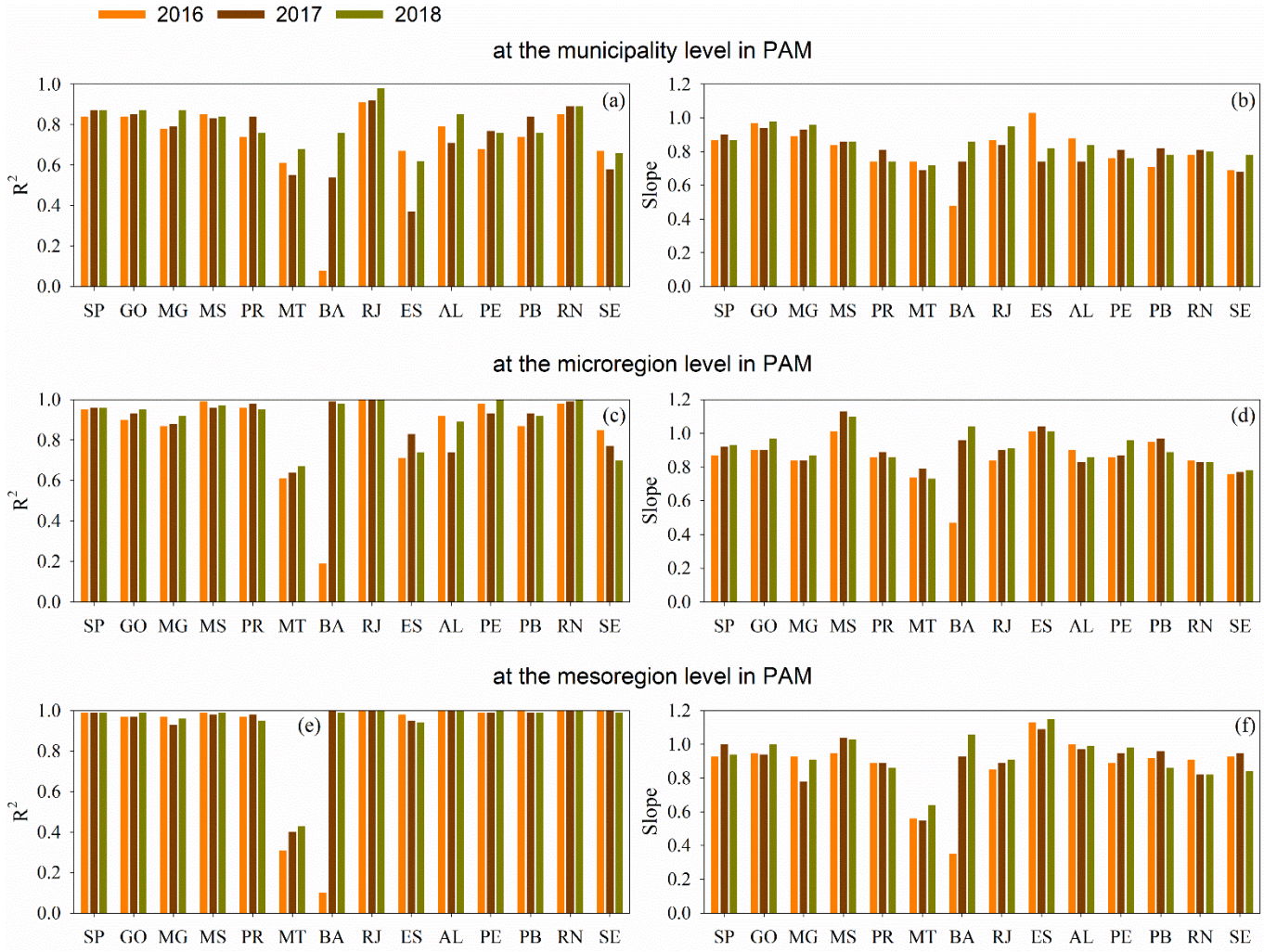


Figure 12. Comparisons between the agricultural statistical harvest area and the estimated harvest area of sugarcane (R^2 and slope) at the (a-b) municipality, (c-d) microregion, and (e-f) mesoregion levels from 2016-2018.

545

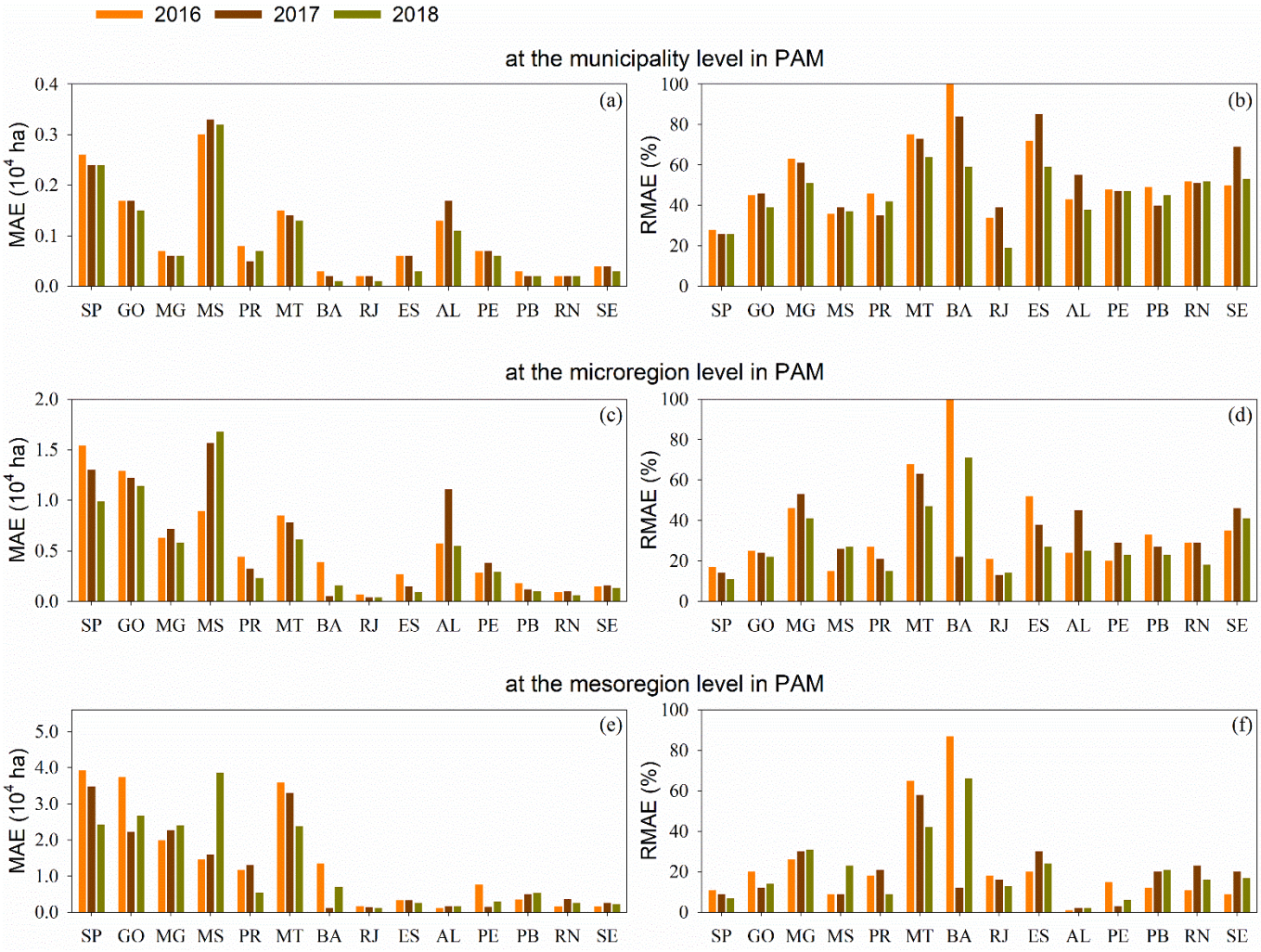


Figure 13. Comparisons between the agricultural statistical harvest area and the estimated harvest area of sugarcane (MAE and RMAE) at the (a-b) municipality, (c-d) microregion, and (e-f) mesoregion levels from 2016-2018.

550

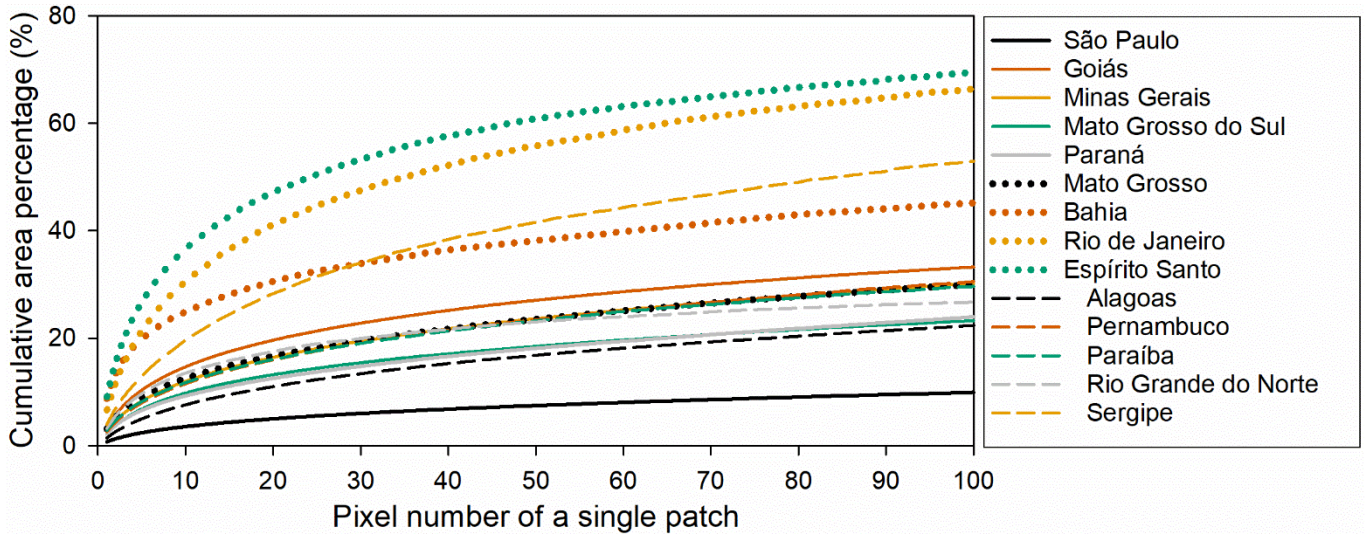


Figure 14. Statistics for patches with different pixel numbers in the sugarcane harvest map for the 14 states in Brazil in 2018.